UNIVERSITY COLLEGE LONDON

DOCTORAL THESIS

Preservation management modelling in archival and library collections

Author:
Cristina DURAN-CASABLANCAS

Supervisors:
Dr. Josep GRAU-BOVÉ
Prof. Matija STRLIČ

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

UCL Faculty of the Built Environment
Institute for Sustainable Heritage

December 14, 2022

Declaration of Authorship

I, Cristina DURAN-CASABLANCAS, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

"Forever—is composed of Nows."

Emily Dickinson

Acknowledgements

Doing a PhD is anything but a lonely undertaking. The amount of interactions with supervisors, researchers, practitioners, fellow PhD students and many others leaves me with the feeling that writing a PhD is essentially a collaborative effort, for without the thoughts, feedback and suggestions of so many people, I could not have aspired to end this dissertation at the level where it is today.

I would be remiss not to mention first and foremost Josep Grau-Bové, who has spent many hours in meetings, both live (prior to the Covid era) and through Teams, to discuss, challenge and support my research activities and the writing of this final document that is being submitted. There can be no doubt that, without Josep, this dissertation would not have come about. *Moltíssimes gràcies* Josep!

It has been a privilege to have had the chance to work closely with Matija Strlič and to benefit from his extensive knowledge on conservation science. His presence at UCL and willingness to support my PhD were important reasons for my decision to go ahead with this undertaking in 2017.

Two partners, the Nationaal Archief and Helicon Conservation Services in the Netherlands, supported this research project, and I would be remiss not to thank specifically Gabriëlle Beentjes. I am looking forward to continuing collaborating on new ideas and projects. Many thanks also to Stadsarchief Amsterdam, Noord Hollands Archief, Utrechtsarchief and NIOD, who generously provided me with help and access to their collections and data sets for my case studies. I am very grateful for the financial support of Metamorfoze, the Netherlands' national programme for the preservation of paper heritage.

Writing a PhD part time, whilst also being employed, is only possible if the employer is willing to be flexible and forthcoming. I am so lucky to be working for the Amsterdam City Archives and, since last year, for the University of Amsterdam. I am even more lucky to have such fine colleagues at these institutions, Jochem, Mike, Sjaak, Jean-Marieke, Emmy, Leila, and the collection care team, who went beyond what anyone could have expected of them in picking up my work whenever I was focused on my research.

To my fellow PhD students, Scott, Yun, Natalie, Betty, Morena, it was such a pleasure to share with you meetings, lectures and activities at UCL and learn from each other these past couple of years.

And last but not least, a special thanks to my family and friends, who were incredibly supportive and interested in what I was doing. It's finally over. Let's get back to doing friends and family stuff!

Abstract

As cultural heritage institutions, libraries and archives are responsible for managing collections in order to ensure access for present and future generations, and sustainable preservation. In pursuing these two goals, institutions face the challenge of determining to what extent preservation actions are beneficial in the context of their own collections. This project contributes to the complex decision-making processes of collections management by developing a mathematical model that shows, quantitatively, the effects of different preservation decisions during a collection's lifetime. The novelty of this research lies in its approach to preservation management not as single, independent measures, but as a process that is part of a complex system: preservation management is not seen in isolation, but in relation to the other archival and library functions in the broader context of collections management. To meet this aim, complex systems modelling and simulation paradigms, such as system dynamics (SD) and agent-based modelling (ABM), are applied. Applying simulation to model preservation management decisions has the potential to develop into an integrated approach for evaluating and comparing the potential benefits of different preservation measures, which, so far, is lacking. This model will support collection keepers in the complex decision-making process of collection management by comparing different strategies, and therefore finding potential synergies as well as counter-intuitive decision outcomes which otherwise might not have been identified.

Impact Statement

In this project, a mathematical model has been developed to support informed decision making in the field of preservation management in archives and libraries. The model has been developed and tested in several archival institutions, and proved that new insights that emerged from the model's development and application are directly applicable to decision-making processes within institutions, thereby impacting upon their collection management strategies. Although far from claiming that the challenges have been overcome, this research has provided awareness and understanding into how to solve the dichotomy between ensuring the use of the collections, and the need to preserve them. In addition, the results of this research have the potential to eventually make institutions' work processes more efficient and effective by identifying objects and collections in need more easily by, for example, maximising the use of the management systems already in place. Furthermore, as institutions are confronted with limited budgets, the model can help to allocate available human and economic resources to preservation needs, and, at the same time, support the development of sustainable preservation programmes, contributing to better collections access for present and future generations.

This research belongs to a body of work where heritage collections are viewed as populations of objects. At the same time, this research enhances new knowledge and scientific advancement by investigating simulation modelling methods in heritage collections. This is the first time that simulation was applied to model demand levels in archival reading rooms, or to model the life expectancy of archival and library collections depending on the preservation measures taken. Whereas the simulation of complex systems is an increasing area of research in other fields, this is a new topic in heritage. This project has resulted in new exiting collaborations with other research groups, creating new synergies within the modelling community.

Engagement with the professional heritage community has been an intrinsic part of the project. The model has been developed in collaboration with end users to ensure that it provides support on those questions that are relevant to practitioners caring for heritage collections. The results were communicated to the field through presentations in conferences and publications in academic and professional journals. In addition, public engagement activities, organised to enhance information on how the collections are used by readers, turned out to be a good opportunity to bring heritage science research to the general public.

Contents

Declaration of Authorship 1

Acknowledgements 3

Abstract 5

Impact Statement 6

List of Figures 14

List of Tables 21

List of Abbreviations 24

1 Scope of work 25

1.1 Introduction . . . 25
1.2 Research questions, aims and objectives . . . 28
1.3 Outline research methods . . . 29
1.4 Thesis outline . . . 30
1.5 A technical note . . . 31

2 Frameworks, models and tools in preservation management 34

2.1 Description of available approaches . . . 34
2.1.1 Risk management approaches . . . 35
2.1.2 Damage functions . . . 36
2.1.3 Indoor climate evaluation models . . . 38
2.2 Input data . . . 39
2.3 Time horizon . . . 43
2.4 Dealing with uncertainty . . . 44
2.5 Presenting the results . . . 45
2.6 Combining approaches . . . 47
2.7 Adapting the existing models . . . 48
2.8 Conclusions . . . 49

3 Qualitative modelling **55**
3.1 Causal loop diagram 56
3.1.1 Method 56
3.1.2 Results and discussion 58
3.1.2.1 Preservation management 59
3.1.2.2 Access of the collections 60
3.1.3 General observations and conclusion 61
3.2 Defining the boundaries of the APM model 63
3.2.1 Problem definition 63
3.2.2 Key variables 64
3.2.2.1 Characteristics of the collections 64
3.2.2.2 Preservation and conservation options 65
3.2.2.3 Access 65
3.2.2.4 Costs 66
3.2.3 Time horizon 67
3.3 Conclusions 67

4 Simulation modelling **71**
4.1 (Complex) systems, models and simulation 71
4.2 Simulation paradigms 76
4.2.1 Paradigms description 77
4.2.2 Heterogeneity: Aggregates, individual entities or agents . . 82
4.2.3 Randomness: Deterministic vs stochastic elements 87
4.2.4 Handling of time: Continuous or discrete? 89
4.2.5 How does the dynamic behaviour emerge? 89
4.2.6 Hybrid models 91
4.2.7 What type of data is required? 93
4.3 Conclusions 94
4.4 Model building methodology 95

5 Access **106**
5.1 Digitisation strategies: An introduction 107
5.1.1 Drivers for digitisation 108
5.1.2 Costs of digitisation 109
5.1.3 Criteria of selection 110
5.1.4 User studies 111
5.1.5 Data driven decision making 113
5.2 Analysis of actual collection data 116
5.2.1 Introduction 116

5.2.2 Methodology 117
5.2.3 Results and discussion 117
5.2.3.1 Instances of use in the reading room 119
5.2.3.2 Impact of digitisation on the use of physical collections 121
5.2.3.3 Digitisation match 122
5.2.3.4 Scan on Demand versus block digitisation 124
5.2.3.5 Readers' behaviour 126
5.2.4 Conclusions 131
5.3 Survey of readers' preferences 133
5.3.1 Introduction 133
5.3.2 Methodology 133
5.3.3 Results and discussion 134
5.3.4 Conclusions 137
5.4 Modelling access with system dynamics 138
5.4.1 Model boundaries 139
5.4.2 A brief reflection on the system dynamics application 139
5.4.3 Stocks and flows 141
5.4.4 Mathematical model 142
5.4.5 Experiments 146
5.4.5.1 Experiments description 146
5.4.5.2 Results and discussion 147
5.4.5.3 Limitations of the system dynamics model 149
5.5 Modelling access with agent based modelling 151
5.5.1 Model description 151
5.5.2 Experiments 153
5.5.2.1 Experiments description 153
5.5.2.2 Results and discussion 155
5.5.2.3 Limitations of the model 161
5.6 Verifying and validating the simulation models 163
5.6.1 Noord-hollands Archief data set 165
5.6.1.1 Data analysis 165
5.6.1.2 Experiments 167
5.6.2 Utrechtsarchief data set 170
5.6.2.1 Data analysis 170
5.6.2.2 Experiments 173
5.7 Conclusions 174

6 Preservation **184**

6.1 Modelling chemical degradation in paper collections 185

6.1.1 The kinetics of cellulose degradation 186

6.1.2 Studies on the comparison of preservation measures 191

6.1.3 Pollutants 195

6.1.4 pH of paper and deacidification 196

6.1.4.1 Mass-deacidification 198

6.2 Modelling mechanical degradation in paper collections 199

6.2.1 Wear and tear function 199

6.2.2 Survey data 201

6.3 APM-Preservation model boundaries 203

6.4 APM-Preservation model description 206

6.4.1 Chemical condition 206

6.4.2 Mechanical condition 207

6.4.3 Layout of the APM-preservation model 208

6.5 Model data set generation 210

6.5.1 Reference data sets 210

6.5.2 Model data sets 212

6.5.2.1 Degree of polymerisation values generation 212

6.5.2.2 pH values generation 214

6.5.2.3 Generated model data set 214

6.6 Exploring uncertainty 216

6.6.1 Number of agents and stochasticity 217

6.6.2 Parametric uncertainty 219

6.6.2.1 Sensitivity dose–response function to pH 219

6.6.2.2 Distribution of pH after deacidification 221

6.7 Exploring preservation management scenarios 222

6.7.1 Effect of selection criteria for deacidification treatment . . . 222

6.7.2 Effect of deacidification vs storage conditions 224

6.7.3 Effect of delaying deacidification 229

6.8 Conclusions 230

7 Cost **239**

7.1 Cost model examples 240

7.1.1 Collection care 240

7.1.2 Cost models for digital collections 241

7.1.3 Energy cost models 244

7.1.4 Key findings for the APM-cost model 248

7.2 APM-Cost model description . . . 249
7.2.1 Purpose and boundaries of the model . . . 249
7.2.2 Main structure of the model . . . 250
7.3 Experiments . . . 251
7.3.1 Digital access vs reading room access . . . 251
7.3.2 Deacidification and environmental storage strategies . . . 260
7.3.2.1 Energy costs . . . 261
7.3.2.2 Deacidification costs . . . 263
7.4 Conclusions . . . 267

8 Model application 273
8.1 Description of the APM model . . . 273
8.1.1 Inputs . . . 274
8.1.2 Outputs . . . 275
8.2 Case studies . . . 277
8.2.1 NIOD Institute for War, Holocaust and Genocide Studies . . . 278
8.2.1.1 Data collection . . . 279
8.2.1.2 Data analysis . . . 283
8.2.1.3 Simulation . . . 290
8.2.2 Utrechtsarchief . . . 292
8.2.2.1 Data collection . . . 292
8.2.2.2 Data analysis . . . 294
8.2.2.3 Simulation . . . 298
8.2.3 Conclusions . . . 301

9 Conclusions 308
9.1 Research question 1 . . . 309
9.2 Research question 2 . . . 311
9.3 Research question 3 . . . 313
9.4 Strengths and limitations of the proposed model . . . 316
9.5 Future work . . . 320
9.6 Final (personal) reflection . . . 322

A Causal loop diagram 323

B Questionnaire visitors reading room 326

C Calculator for energy use 329

D APM-Cost model 332

E APM-Preservation model **333**

F APM-Access model in system dynamics **361**

G APM model **363**

Bibliography **419**

List of Figures

2.1 Example of a stacked horizontal bar graph ranked according to the total magnitude of risk . . . 45
2.2 Output of the eClimateNotebook. . . . 45
2.3 Demographic plots developed for the collection of the Amsterdam City Archives (Duran-Casablancas et al.; 2017). . . . 46
2.4 Comparison of preservation scenarios at the Danish Royal Library collections from 1850–1985 given by Vinther Hansen and Vest (2008). 47

3.1 Group model building session . . . 56
3.2 Causal loop diagram with the causal relationships between preservation management, access and costs. . . . 58

4.1 Scheme describing the three major paradigms in simulation given by Djanatliev (2013). . . . 78
4.2 Results of a search in Google Scholar (conducted on 26-11-2021), showing the percentage of articles since 2020 that include the term 'Covid' and the name of one of the simulation methods in combination with one of the search terms. . . . 82
4.3 Hybrid model architectures according to Borshchev (2013) . . . 92
4.4 The proposed APM model composed of three interconnected submodels. . . . 96

5.1 Number of access requests per year in the reading room of the SAA between 1999 and 2018 and number of archival records digitised per year since 2004. . . . 118
5.2 Frequency distribution of records that had been requested in the reading room according to the number of access requests in 1999, 2013 and 2018, and cumulative percentage of the number of records that had been requested in the reading room between 1 and 113 times for 20 years at the SAA. . . . 118
5.3 Number of records requested 1–3 times per year. The number of annual requests made between 1999 and 2018 is shown as a reference at the SAA. . . . 120

5.4 Number of records requested in the reading room between 1 and 3 times per year, indicated as percentage of the total number of annual requests at the SAA. . . . 120

5.5 Number of records classified according to the total number of access requests during the 20-year-period at the SAA. . . . 121

5.6 Number of records digitised per year; number of records requested in the reading room that have been digitised; and sum of requests history in this group (accessed and digitised), referred to as 'digitisation match' at the SAA. . . . 123

5.7 'Digitisation match', sum of requests history of requested *and* digitised records at the SAA. . . . 123

5.8 Percentage of records requested in the reading room *and* digitised according to the two digitisation programs at the SAA. . . . 124

5.9 Number of records requested in the reading room that have been digitised and sum of requests history in this group, according to the two digitisation programs at the SAA. . . . 125

5.10 Cumulative percentage of number of records digitised within the two digitisation programmes, grouped by accumulated instances of access at the SAA. . . . 125

5.11 (*Top*) Number of individual readers between 1998 and 2018, with fitting line. (*Bottom*) Average number of requests per reader at the SAA. . . . 126

5.12 Annual change in the number of readers and requests, and 'digitisation match', compared to the reference year (2006) at the SAA. . . . 127

5.13 Number of readers grouped by the number of access requests per reader in 1999, 2006, 2013 and 2018 at the SAA. . . . 128

5.14 Number of access requests in 1999, 2006, 2013 and 2018, grouped by reader type, based on the annual number of access requests per reader at the SAA. . . . 128

5.15 Number of readers and requests in 2011 and 2012, grouped by reader type, based on the annual number of access requests per reader at the SAA. . . . 129

5.16 Percentage of records requested by the top five readers between 1999 and 2018 at the SAA. . . . 129

5.17 (*Top*) Number of individual readers and requests at the SAA. (*Bottom*) Number of individual readers and requests, not including the group of readers with more than 100 requests. . . . 130

5.18 (*Left*) Percentage of research studies that do not use indexes, conducted by the respondents. (*Right*) Number of records requested by the respondents on the day they participated in the survey. . . . 135

5.19 The actual use of SoD (*left*) and respondents' willingness to use the SoD (*right*), grouped by the number of times the respondents visited the reading room. 136

5.20 Visualisation of the access part of the APM model. 141

5.21 Comparison of the decrease in access requests according to the SAA data set and the output of four experiments conducted according to the SD approach. 148

5.22 Annual 'digitisation match', sum of requests history of requested *and* digitised records, given as percentage, in the SAA data set. . . . 149

5.23 Comparison of the ABM output (number of access requests in the reading room (R_r) at the end of the run) depending on the number of agents. 156

5.24 Comparison of the decrease in access requests according to the SAA data set and the output output of 10 repeated runs of the ABM simulation with 4,000 agents. 157

5.25 Annual number of requests in the reading room according to ABM simulation showing the mean value of the repeated experiment shown in Figure 5.24 and the SAA data as a reference. 157

5.26 Comparison of the ABM simulation output (Fig. 5.24) and the SAA data set, regarding the number of records/agents requested (*top*) and the total number of requests in both cases (*bottom*). 158

5.27 Perfect fit of the decrease in requests in an ABM simulation with the same percentage of M_r at the end of the run as in the SAA data set. 159

5.28 (*Left*) Comparison of the cumulative distribution (proportion) of the numbers of records requested in the SAA reading room and the experiment with exponential distribution for the number of requests per year. (*Right*) Comparison of the ABM experiment and the SAA data showing the number of requests in the reading room. 160

5.29 Annual number of requests in the reading room according to experiments using different distributions for the number of requests per year. 160

5.30 Comparison of the conversion of number to records to agents (*top*) and the number of requests in both cases (*bottom*). 163

5.31 Number of requests in the NHA reading room between 2008 and 2019 and number of records that have been digitised annually since 2015. 166
5.32 (*Top*) Cumulative distribution of the numbers of records that have been requested in the NHA reading room during a period of 11 years. (*Bottom*) Number of records requested in the reading room that have been digitised , and sum of requests history of this group (accessed and digitised), according to the two digitisation programs. 167
5.33 Number of records digitised per year; number of records requested in the reading room that have been digitised; and sum of requests history of this group at the NHA. 168
5.34 (*Top*) 'Digitisation match' given as a cumulative percentage (orange line) and percentage decrease in requests compared to 2016 at the NHA. (*Bottom*) Comparison of SD and ABM experiments and the NHA data showing the number of requests in the reading room. . . 169
5.35 Number of requests in the UA reading room between 2007 and 2019 and number of records that have been digitised annually since 2009. 170
5.36 Number of records digitised per year, number of records requested in the reading room that have been digitised, and sum of request history of this group at the UA. 171
5.37 (*Top*) Cumulative percentage of the numbers of records that have been requested in the UA reading room during a period of 12 years. (*Bottom*) Number of records requested in the reading room that have been digitised, and sum of request history of this group, according to the two digitisation programs at the UA. 172
5.38 (*Top*) 'Digitisation match' given as a cumulative percentage (orange line) and percentage decrease in requests compared to 2009 at the UA. (*Bottom*) Comparison of SD and ABM experiment and the UA data showing the number of requests in the reading room. 174

6.1 Diagram of the preservation part of the APM model. 185
6.2 Changes in DP of Whatman filter paper sample with DP_0 of 2900, aged at 90 an 65%, and the same plot transformed according to the Ekenstam equation (Strlič and Kolar; 2005, p. 34). 188
6.3 Accumulation of tears per challenge (equivalent to 10 instances of handling) depending on the DP of paper, per 100 sheets. 200
6.4 Number of tears per handling per 100 sheets depending on the DP of paper, according to a logarithmic or linear equation. 201

6.5 Boxplots of the percentage of failed pages and the number of requests for the two groups in the case–control study conducted on an archive containing acidic paper. 203
6.6 Causal loop diagram showing the variables included in the preservation part of the APM model. 204
6.7 Screenshot of dashboard of the (chemical) preservation part of the APM model. 209
6.8 Screenshoot of one of the agents in the APM model. 209
6.9 pH and DP of SurveNIR acidic papers and pH and generated DP of groundwood papers of the survey data set with regression line. . 211
6.10 DP values of the Dutch archives reference data set with distribution used to generate the DP values of the model data set representing acidic collections. 213
6.11 DP values of the Dutch archives reference data set with distribution used to generate the DP values of the model data set representing modern collections. 213
6.12 Generated data sets for acidic, modern and mixed collections. . . . 215
6.13 Generated data sets for three acidic collections with different DP frequency distributions of a population of 2.000 agents. 216
6.14 Run time of an experiment with a time horizon of 100 years depending on the number of agents in the model. 218
6.15 Output of the number of metres of the SAA collections in critical, poor and fair condition after a time horizon of 100 years generated by models with different number of agents. 218
6.16 Percentage of the collection in critical condition (DP below 300) depending on pH input values to explore pH uncertainty. 220
6.17 Expected DP loss in 100 years depending on the pH value of the record, assuming storage conditions of 18 °C and 50% relative humidity, according to the dose–response function developed for acidic catalysed degradation of cellulose (initial DP 1000) developed by Strlič et al. (2015b). 221
6.18 Frequency distribution of the DP values after 500 years, as result of the application of a uniform (*top*) customised distribution of pH values (*bottom*) 222
6.19 Results of the four preservation strategies on acidic, modern and mixed collection, in the time horizon of 500 years. 225
6.20 A comparison of strategies for deacidification with the same effectiveness as storage conditions at 16 °C and 40% RH. 229

6.21 Effect of the strategy of deacidifying part of the collection in 50 years (delay 0 years) and the effect of the same strategy if this implementation is delayed in the steps of 50 years. 230

7.1 Model structure of LIFE^3 (Life Cycle Information for E-Literature) (Ayris et al.; 2010) 242
7.2 Lifecycle collection management proposed by Shenton (2003). . . . 243
7.3 Scheme of the energy balance model to calculate energy costs in museums proposed by Padfield (2010). 245
7.4 Calculation of energy loads according to Padfield's model (2010), using different AER inputs in three climate scenarios: mean external temperature of 12.6 °C in 2009 to 15.4 °C in 2050 and 17.7 °C in 2080. 246
7.5 Simplified scheme of the APM cost model 251
7.6 (*Left*) Annual cost of preserving and providing access of the collections at the SAA. (*Right*) Costs per activity related to digital collections during a period of ten years. Online storage price is €2.85/GB per year. 254
7.7 Annual cost of preserving and providing access of the collections at the SAA; (*Right*) Costs per activity related to digital collections during a period of ten years. Price of online storage is €1.78/GB per year. 254
7.8 Prediction of the number of requested records in the reading according to four digitisation strategies as described in Table 7.3. . . . 256
7.9 Scheme of the APM cost model per metre of archive. In red, the new dependencies that arise between variables when costs are calculated for one metre archive rather than for the whole collection (Fig. 7.5). 257
7.10 Annual costs per metre of archive during a period of ten years. . . 258
7.11 Cumulative cost of providing digital access versus access in the reading room per metre of archive, according to run 1 in Table 7.3. . 260
7.12 Correlation between thickness and the number of sheets of a inventory number, grouped by date of the records. 264

8.1 Causal loop diagram showing the variables included in the APM model, and the variables omitted. 273
8.2 Mock-up of the dashboard including several inputs of the APM model as example. 274

8.3 NIOD building located in the city centre of Amsterdam, The Netherlands (credit: NIOD). . . . 278
8.4 Repository facilities at the NIOD . . . 279
8.5 Histogram of the DP values for paper surveyed at the NIOD. . . . 284
8.6 Scatterplot showing the DP and pH and tensile strength values for the surveyed paper at the NIOD. . . . 286
8.7 Loadings of the PCA, showing correlations between pH, DP, tensile strength, rosin and lignin content. . . . 287
8.8 Scatterplot showing the DP values of the surveyed papers at the NIOD, dating from 1910 to 2010, grouped by paper types. . . . 287
8.9 Storage conditions (temperature and RH) in the NIOD repository, from July 2019 to June 2020. . . . 288
8.10 Number of requests in the reading room and cumulative number of metres of archive (partially) digital available at the NIOD. . . . 289
8.11 Cumulative distribution (percentage) of the numbers of records that have been requested in the reading room between 1 and 70 times during a period of 10 years at the NIOD. . . . 290
8.12 Number of requests of four NIOD collections; two collections fully digitally available (*left*), and two collections partially digitally available (*right*). . . . 291
8.13 UA buildings. . . . 293
8.14 Histogram of DP values for the surveyed paper at the UA. . . . 294
8.15 Scatterplot showing the DP values of the surveyed papers dating from 1600 to 2000, grouped by paper type. *Left*, values of the surveyed paper at the UA; *right*, at the SAA. . . . 295
8.16 Scatterplot showing DP and pH for paper from 1600–2000, grouped by paper type. *Left*, values of the surveyed paper at the UA; *right*, at the SAA. . . . 296
8.17 Comparison of papers at the SAA and UA, including bleached pulp and groundwood paper after 1850. . . . 296
8.18 Scatterplot showing the pH (*top*) and tensile strength values (*bottom*) for paper at the SAA and UA, dating from 1600–2000. . . . 297
8.19 Monthly average temperature and relative humidity in one of the repositories of the UA over a period of two years. . . . 298

9.1 pH values of records that are expected to reach the critical DP of 300 within the time horizon of 500 years, depending on the DP_0 value and according to two storage climate scenarios. . . . 314

List of Tables

4.1 Key characteristics of the three major modelling approaches. 80

5.1 Type of reader given by respondents of the questionnaire ($n = 42$). 135
5.2 Increase in the percentage of respondents willing to consider using the SoD based on the scan delivery times. 136
5.3 Responses to the survey items regarding the factors associated with the decision to use the SoD instead of visiting the reading room. . . 137
5.4 Increase in the percentage of respondents willing to consider using the SoD based on the opening hours of the reading room. 137
5.5 Nomenclature access model 143
5.6 Input variables of the system dynamics model. See Appendix F for SAA data set. 147
5.7 List of variables assigned to the agents in the ABM model. 152
5.8 Conversion of number of records to agents in the ABM model. . . . 162
5.9 Comparison of control outputs between the NHA data set and the outputs of the ABM model. 168
5.10 Statistics on the use of the collections at the NHA and the UA. . . . 171

6.1 Number of years to reach the critical DP of 300, depending on the initial pH and DP at 18 °C and 50% RH, according to the dose–response function for acid-catalysed degradation. 192
6.2 Predicted extension of useful time of paper (factor) under different storage conditions according to Balazic (2007), Vinther Hansen (2008), and Tétreault (2019) compared to the dose–response function for acid-catalysed degradation (initial DP 1500, pH 6) developed by Strlič et al. (2015b) 194
6.3 Predicted extension of the useful life of the collections at the Royal Library Denmark, according to different preservation scenarios, suggested by Vinther Hansen and Vest (2008), and the dose–response function for acid-catalysed degradation developed by Strlič et al. (2015b). 195

6.4 Input values of the normal distribution to generate pH values depending on DP value. 214
6.5 Comparison of the effect of a deacidification treatment on the percentage of the collection unfit for purpose after 500 years when different selection criteria for deacidification were applied. 223
6.6 Comparison of strategies for three acidic collections after a time horizon of 500 years. 225
6.7 The preservation strategies presented in Figure 6.19 are compared at the end of the time horizon of 500 years. 226
6.8 Number of metres that need to be deacidified per year during a period of 100 years to achieve similar results on the preservation of the collection of 50.000 metres compared to storage conditions of 16°C and 40% RH. 227
6.9 Comparison of environmental strategies versus deacidification for acidic collections and modern and mixed collections after a time horizon of 500 years. 228

7.1 Quality of Envelope (QoE) specifications described by Kramer et al. (2015). 247
7.2 Activities included in the APM-cost model. 250
7.3 Four runs to predict the effect of digitisation on the number of requests in the reading room at the SAA. 255
7.4 Input values to calculate the energy use in a repository. 261
7.5 Calculation of the energy use in a repository according to the inputs given in Table 7.4 and the exterior conditions (mean temperature and RH exterior between 1970–2018, de Bilt, The Netherlands) (KNMI; n.d.). 262
7.6 Calculation of the energy use and cost in a repository as given in Table 7.5 263
7.7 Cost of deacidification for different types of collections as described in Table 6.9, to obtain similar preservation levels as environmental storage conditions of 16 °C and 40% RH. 266

8.1 List of inputs of the APM model. 276
8.2 List of inputs/outputs of the agents in the APM model. 276
8.3 List of outputs of the APM model. 277
8.4 Comparison of strategies for the NIOD collections after a time horizon of 500 years. 291

8.5 Comparison of strategies for the UA collections after a time horizon of 500 years. 299
8.6 Comparison of strategies for the UA collections on wear and tear after a time horizon of 100 years. 300

List of Abbreviations

ABM	**A**gent **B**ased **M**odelling
AER	**A**ir **E**xchange **R**ate
AHU	**A**ir **H**andling **U**nit
APM	**A**rchival **P**reservation **M**anagement model
CLD	**C**ausal **L**oop **D**iagram
DES	**D**iscrete-**E**vent **S**imulation
DP	**D**egree of **P**olymerisation
HVAC	**H**eating **V**entilation and **A**ir **C**onditioning
IFLA	**I**nternational **F**ederation of **L**ibrary **A**ssociations and **I**nstitutions
IPI	**I**mage **P**ermanence **I**nstitute
JISC	**J**oint **I**nformation **S**ystems **C**ommittee
MW	**M**olecular **W**eight
NHA	**N**oord **H**ollands **A**rchief
NIOD	Institute for War, Holocaust and Genocide Studies
NIR	**N**ear **I**nfra **R**ed
PCA	**P**rincipal **C**omponent **A**nalysis
RH	**R**elative **H**umidity
SAA	**S**tads**A**rchief **A**msterdam / Amsterdam City Archives
SD	**S**ystem **D**ynamics
SoD	**S**can **o**n **D**emand
TS	**T**ensile **S**trength
UA	**U**trechts**A**rchief

Chapter 1

Scope of work

'For me, and many others too, to engage in creative pursuits and scholarly undertakings is to feel part of a long, rich, and ongoing human dialogue.'

Greene 2021, p. 320

1.1 Introduction

The objective of preservation is to ensure that archive and library holdings will be available to the users when needed (Feather; 1996). This basic objective confronts memory institutions with a challenging dichotomy: how can documentary heritage be handed on to future generations while being made available to current users (Feather; 2004). To meet this challenge, it is not sufficient to follow the best knowledge to prevent or slow degradation. As the American Institute for Conservation states, systematic mitigation of the risks is needed, which means that 'the benefit of mitigating any conceivable risk is considered relative to the costs and benefits of dealing with that risk and, most importantly, the effect on the expected usefulness of the collection over time' (American Institute for Conservation; n.d.). Hence, preservation has often been defined as 'all managerial, technical and financial considerations applied to retard deterioration and extend the useful life of (collection) materials to ensure their continued availability' (Eden et al.; 1998, p. 4). From this definition, the following questions emerge to be addressed by institutions: what is to be preserved, for how long and by what means (Feather; 1996).

Standards have been developed to guide institutions in decision making, especially regarding environmental control. However, standards have progressively moved from being a prescriptive norm to a management strategy tool, such as the PAS 198:2012 and ASHARE. Image Permanence Institute (IPI) director J. Reilly (2013, p. 26) has remarked, '[r]ather than telling us what to do, standards now tell us what not to do. [. . .] You are choosing from a continuum of preservation quality/degree of risk and energy savings.' A clear example of this new approach to preservation management is the isoperms developed for paper (Strlič et al.; 2015). These plots show the risks involved with the vulnerability of the materials and the effect of different environmental conditions. Therefore, they are open to interpretation when applied in the case of each institution.

Another aspect that has become clear in the last decade is that sustainability and the energy economy need to be considered within preservation management. During the Summit on the Museum Preservation Environment at the Smithsonian Museum in 2013 (Smithsonian Institution; 2013), different speakers emphasised that, in addition to the most significant vulnerabilities of collection materials, other aspects need to be introduced, such as local climate, the capabilities of the mechanical system, and the limitations imposed by the building's construction.

The British Specification for Managing Environmental Conditions for Cultural Heritage, PAS 198:2012 (British Standards Institution; 2012), uses a diagram of concentric rings with the cultural collection in the inner circle to summarise the considerations that should be included in the environmental management strategy for collections. The diagram includes environmental considerations (temperature, relative humidity, light, pollutants) as well as collection considerations that include material sensitivity and expected lifetime, energy economy and the usage and display of collections. Herewith another aspect, namely the intended use of collections, is added to the long list of aspects that need to be considered when deciding what the right balance could be in the case of each institution. As Ashley-Smith has expressed, balance means 'the need to optimise four competing and interconnected outcomes—stability, cost, sustainability, and accessibility. In such a complex system, the optimal solution will probably be sub-optimal to at least one of the components' (Ashley-Smith; 2013, p. 30). This PhD study contributes to helping collection keepers with the complex decision-making process of collections management by developing a model to understand the effect of different preservation decisions during the lifetime of the collections.

'Research on life-cycle prediction, the measurements of the natural ageing of materials, the ageing and behaviour of materials late in the life-cycle' was identified as one of the priorities for the applied conservation research in the meeting at the British Library in September 2004 supported by the Andrew W. Mellon Foundation (British Library; 2004, p. 18). However, life-cycle prediction also includes 'quantitative modelling of the effects of existing conservation and preservation options on the life-cycle, and cost-benefit analysis of different preservation strategies, leading to reliable predictive modelling' (British Library; 2004, p. 18). These topics closely align with the topics this PhD project aims to address. The goal behind this research might not be new, but making it reality was a difficult task. Thanks to the major contributions made in the last decade to assist archives and libraries with preservation planning, we can now embark on further modelling the impact of preservation measures on the life cycle of the paper collections.

In this project, preservation management is approached as a complex system, where the key variables and the dynamic complexity of the system are modelled to quantify the effects of different preservation strategies. Therefore, this PhD study uses modelling and simulation paradigms of complex systems, such as system dynamics (SD) and agent-based modelling (ABM).

This PhD contributes to the complex decision-making process of collections management by developing a mathematical model that quantitatively shows, to the extent possible, the effect of different preservation decisions during the lifetime of paper collections. The model will be referred to as the Archival Preservation Management (APM) model, since most of the examples given in this thesis are related to archival collections, but not exclusively. Despite the name, the focus of the model is on paper collections hold in institutions responsible for the preservation as well as providing access to the collections, such as archives and libraries. In this model, preservation management is not seen in isolation but in relation to the other archive and library functions in the broader context of collections management. As it has been stressed by Foot (2013, p. 4), 'the purpose of a collection will strongly influence the actions taken to preserve it, but other library and archive functions that are closely linked to preservation must also be considered, such as acquisition, retention, access and use' .

Modelling and simulation paradigms of complex systems have proven validity in a broad number of areas, for example, medical care, ecology or archaeology. Simulation has been defined as 'the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behaviour of the system and evaluating various strategies for the operation of the system' (Shannon; 1998, p. 7). Simulation is a powerful tool that allows decision makers to use the model as a learning laboratory to explore 'what-if' scenarios (Sterman; 2000). Simulation modelling is particularly concerned with the dynamic behaviour of the system over time, which is one of the sources of complexity. The application of simulation to model preservation management decisions has the potential to develop into an integrated approach for the evaluation and comparison of the potential benefits of different preservation measures, which so far is lacking. This model will support collection keepers with the complex decision-making process of collection management by comparing different strategies and, therefore, finding the potential synergies as well as counter-intuitive results of decisions that otherwise might have not have been identified. The results are directly applicable to inform preservation policies and decision making. Systems modelling has not yet been applied in the heritage field. If its validity is proven, this approach has the potential to be used within the heritage field to a great diversity of collections and preservation problems.

1.2 Research questions, aims and objectives

The following research questions will be addressed in this project:

(i) Can the preservation management of archive and library collections be modelled as a complex system?

(ii) How can data collected in actual archives and libraries be used for the mathematical formulation of the model and to validate the model?

(iii) To what extent can the simulation modelling provide quantitative data to enable the comparison of different preservation actions, such as climate control, deacidification or digitisation?

The main aim of the research is to evaluate the impact of preservation measures during the lifetime of collections using simulation modelling. The overall objective can be broken down into different essential steps:

- To determine which simulation modelling paradigm, discrete event simulation (DES), SD, ABM or hybrid model is the most appropriate to model preservation management in heritage collections.

- To determine which variables related to preservation and other archival and library functions need to be included in the model.
- To develop a simulation model to evaluate the impact of preservation measures during the lifetime of archival and library collections in terms of preservation, access and costs.
- To collect data in archival and libraries for those parameters in the model that so far have not been quantified.
- To conduct verification, validation and sensitivity analyses of the developed model.
- To evaluate the impact of preservation measures and which collections will benefit the most from certain preservation measures by comparing what-if scenarios.
- To conduct case studies to determine what input data are needed to use the developed model as a tool to support decision making.

1.3 Outline research methods

The following methods are present throughout the thesis:

1. *Literature review*: To obtain a good understanding of the research field, a literature review is conducted on systems modelling and simulation and the approaches applicable in the context of this research project. Systems modelling and simulation have been applied in many fields. The literature review will mainly focus on those fields that have some similarities with the heritage field, for example, health care, environment and archaeology.
   Additionally, where required, a literature review will be conducted at the beginning of the chapter to put the topic into context, for instance, digitisation, chemical preservation and cost models.
2. *Qualitative development of the model*: The first step of the modelling process consists of identifying the key variables, their causal relationships and their behaviour over time. During participatory sessions, professionals working in preservation and other archival functions such as acquisition, ordering or access developed a causal loop diagram. The causal loop diagram reflects the mental models of the professionals.

3. *Development of the mathematical model*: The causal loop diagram is used as the starting point for the formulation of the mathematical model. The simulation modelling is composed of:

    - statistical relationships and experimental descriptions obtained in experimental settings (such as damage functions, isochrones);
    - demographic data obtained from surveys: statistical analysis of the observed degradation, using NIR-spectrometry to characterise the condition of paper;
    - numbers on operational costs, energy consumption or times a year that objects/collections are accessed by visitors (physically against digitally).

    The model uses values and mathematical formulations reported in the literature. Every parameter and its relationships needs to be quantified or formulated mathematically. Where needed, and if time allowed, the available data are complemented by data collected in real case studies. Uncertainty in inputs is evaluated using sensitivity analyses. In the case of archival and library collections, it is probable that historical evaluation data are not available to conduct validation analysis. However, predictive validation is not strictly required as the aim of the model is not to be a predictive tool but to gain insight into the dynamical behaviour of the system.

4. *Development of the model as a management tool*: To gain confidence in the model, case studies are conducted to explore how end users could use the model as a management tool. The objective of the case studies is to understand what input data needs to be collected, the type of experiments that can be conducted and how the results can be interpreted.

1.4 Thesis outline

This thesis is divided into nine chapters. In the first chapter, the research topic is introduced, and an overview of the thesis is given. Chapters 2 and 3 focus on problem articulation and summarise the scope and boundary of the model. Chapter 2 reviews the tools that have been developed to support decision making in the context of archives and libraries. The potential and the limitations of each tool are discussed to highlight the elements that could be of use in the model. Chapter 3 reports on how a causal loop diagram was developed in collaboration with professionals working in archives and libraries to discuss the causal structure and the dynamic complexity of the problem. This chapter defines the purpose of the model and identifies the key variables of the model and the causal structure.

Chapter 4 starts with a brief discussion on whether preservation management in libraries and archives can be understood as a complex system. Then, a literature review is given in the three major simulation paradigms for complex systems: DES, SD and ABM. A brief description of these three paradigms is given to explore which simulation paradigms meet the characteristics of the model as identified in the two prior chapters.

The mathematical model is developed in Chapters 5 to 7, according to the three parts of the model identified by the causal loop diagram: access (Chapter 5), preservation (Chapter 6) and cost (Chapter 7). Each chapter follows a similar structure. First, the input data needed to formulate the model are discussed, and available data sources on the subject are reviewed. If gaps have been identified, then the methodology and the results of collecting the missing data are presented. The chapter follows with a description of the architecture of that particular part of the model and the mathematical formulation of the variables. Experiments are conducted to test the model. .

In Chapter 8, an overview of the model is given, followed by a description of how end users can use the model, including the input data needed to run the model and how to interpret the results. Two case studies are presented. In Chapter 9, the main conclusions of the research are summarised, and further work in this subject is also suggested.

1.5 A technical note

The authorial 'we' has been chosen for this thesis to increase readability when describing activities. 'We' refers to the individual work of this author, including the implementation of input and feedback from the supervisory team. In case that parts of the research have been conducted in collaboration, this will be stated in a footnote. In addition, parts of this thesis have been published as co-authored research articles during the course of the PhD. The articles have been written by this author and revised following feedback from the co-authors. This will be further clarified in a footnote on those parts of the thesis, where applicable.

References

American Institute for Conservation (n.d.). Collection Care Network.
URL: *https://www.culturalheritage.org/membership/groups-and-networks/collection-care*

Ashley-Smith, J. (2013). British Publicly Available Specification (PAS) 198, "Specification for managing environmental conditions for cultural collections", *Summit on the Museum Preservation Management*, Smithsonian Institution, pp. 29–32.

British Library (2004). Future life of collections. Report on a meeting on Applied Conservation Research, held at The British Library on 27-28 September 2004, *Technical report*, British Library, London.

British Standards Institution (2012). *PAS 198:2012 Specification for managing environmental conditions for cultural collections*, BSI Standards Limited.

Eden, P., Dungworth, N., Bell, N. and Matthews, G. (1998). *A model for assessing preservation needs in libraries*, British Library Research and Innovation Centre, Boston Spa.

Feather, J. (1996). *Preservation and the management of library collections*, second edn, Library Association Publishing, London.

Feather, J. (2004). Introduction. principles and policies, *in* J. Feather (ed.), *Managing preservation for libraries and archives*, Ashgate, Aldershot, chapter 1.

Foot, M. M. (2013). *Building a preservation policy*, Preservation Advisory Centre, British Library.

Greene, B. (2021). *Until the end of time*, Penguin Books.

Reilly, J. (2013). Choosing standards and best practices for environmental design and operation, *Summit on the Museum Preservation Management*, Smithsonian Institute, pp. 25–28.

Shannon, R. E. (1998). Introduction to the art and science of simulation, *Proceedings of the 1998 Winter Simulation Conference*, pp. 7–14.
URL: *http://cecs.wright.edu/~fciarall/ISE195/Readings/ShannonSimulationART.pdf*

Smithsonian Institution (2013). Summit on the Museum Preservation Environment, *Technical report*, Smithsonian Institution.
URL: *https://repository.si.edu/bitstream/handle/10088/34611/13.03.EnvironPreservationSummit.Final.pdf?sequence=1&isAllowed=y*

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*, Mcgraw-Hill Education, Boston.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015). Damage function for historic paper. Part III: Isochrones and demography of collections, *Heritage Science* **3**(1): 40.

Chapter 2

Frameworks, models and tools in preservation management

'One of the most enduring problems of the human race is that each person must continually take actions that require knowledge of the future.'

Meadows and Robinson 1985, p. 1

In the last decades, deliberative decision making has been embraced in collection management and preservation practice, supported by the development of decision-making frameworks (e.g. risk assessments), mathematical models (e.g. dose–response functions) and tools (e.g. indoor climate evaluation models) (Henderson and Waller; 2016; Bell et al.; 2018; Bülow et al.; 2018). In this chapter, we review these approaches by comparing them on the following aspects: input data, time horizon, uncertainty and presentation of the outputs.

One central theme will run through this review: how strongly the reviewed approaches are based on the use of data (mathematical models) or existing knowledge and experience (mental models). For instance, whereas risk assessments are based partly on experience and partly on data, models that calculate the degradation rate are based on collection data, but there is no room to include experience, and evaluation models use environmental data but not collection data. From each approach, we will identify main concepts and elements that will be used later to build the Archival Preservation Management (APM) model.

2.1 Description of available approaches

The aim of preservation and risk assessment is to gather evidence to assist with collection management by answering the following questions: 'What to do first?

What are the priorities of the heritage asset in its specific context? How to optimise the use of available resources to maximise the benefits of the cultural heritage over time?' (Michalski and Pedersoli; 2016, p. 8). Although preservation frameworks refer to preservation and risk assessments (see the list provided by Bülow et al. (2018)), the focus in this chapter will be on risk assessment methods.

After describing risk assessments models, we discuss the mathematical models. First, we introduce the use of dose–response function models as predictive tools. Then, we look at indoor climate evaluation models, built as web applications to support management decisions regarding storage environments in heritage collections.

2.1.1 Risk management approaches

R. Waller developed the first clear methodology for conducting risk assessments in heritage collections in 1995, known as the Cultural Property Risk Assessment Method (CPRAM) (Waller; 1994, 1995, 2003). More recently, Michalsky (2016) developed the ABC method of risk assessment , which has been widely taught and further adapted by other research groups (Brokerhof et al.; 2017).

Risk assessments are very different from preservation assessments. Whereas in preservation assessments, an actual assessment of the collections is conducted, risk assessment is a theoretical framework and, therefore, it is not strictly necessary to assess the actual condition of the collections. The goal of preservation assessments is to determine the state and preservation needs of the collections by assessing historic damage against established qualitative criteria, mainly related to environment and storage conditions. In risk assessments, the goal is to evaluate potential risks from identified hazards to collections, where risk is defined as 'the possibility of a loss of value to the heritage asset' (Michalski and Pedersoli; 2016, p. 17).

In the ABC method of risk assessment, for each of ten possible causes of deterioration (ten agents of deterioration) (Waller; 1994), the magnitude of risk (MR) is defined as the sum of three logarithmic scores on a 15-point logarithmic scale:

- A score = frequency of the damaging event or the rate of occurrence of a process;
- B score = loss of value for each affected object;

- C score = the fraction of the total collection value that is affected.

A five-step logarithmic scale is used. For example, a B-score of 5 represents 100% loss of value, a score of 4 represents 10% loss of value, and so forth.

In a risk assessment, risks are first identified, and then scenarios related to the risk are developed, analysed and scored. Risks can be then compared and ranked, and priorities can be set based on the calculated MR. Risk assessment is followed by risk management, which refers to the generation, selection, and implementation of options to accept, reduce, or change risks or preservation need.

Risk assessments can be conducted to evaluate the magnitude of single risks or to compare two options assisting risk-based decision making. Examples of how to apply risk assessment to particular preservation issues are given by Bülow (2010) and Garside et al. (2018). However, a more extended assessment is needed to inform risk management of a heritage asset; the risks scenarios affecting the whole collection must be analysed.

Risk assessments and related approaches, such as QuiskScan, a quick scan to identify value and hazards in a collection (Brokerhof and Bülow; 2016), provide a good overview of the main hazards to the collections and how to mitigate specific risks. Related to risk assessments, a more holistic approach has also been proposed by Bülow and Brokerhof (2011) who introduce the concept of 'quality adjusted life years' (QALYs) in collection management, a term originally developed to measure cost effectiveness in health care. An interesting aspect of this approach is that 'collection quality' takes both the loss of value into account and a range of other aspects (accessibility, development, use, life expectancy and costs). Whether risk assessments are directly related to the (physical) preservation of the collections, the QALYs approach takes a more holistic approach by including other processes within institutions. This holistic approach is (still) unusual in preservation management approaches.

2.1.2 Damage functions

Damage functions have been defined as 'functions of unacceptable change to heritage [assets] dependent on agents of change' (Strlič et al.; 2013, p. 80). 'Damage' is a value-dependent term in this definition, where a value function (threshold) is added to the 'dose–response functions' (change functions). Examples of these models are given by Strlič et al. (2013, p. 81).

In these mathematical models, the main parameters to model a certain degradation process are included on the basis of experimental evidence. For example, based on the Ekenstam function, the dose–response function models the loss of degree of polymerisation (DP) with time that can be expressed in the generic equation:

$$k \cdot t = \frac{1}{\mathrm{DP}} - \frac{1}{\mathrm{DP}_0}$$

where DP_0 and DP are the degree of polymerisation of cellulose at time 0 and t and k is the rate constant [year^{-1}]. In the model proposed by Strlič et al. (2015b), the degradation of cellulose is a function of the pH value of the paper and environmental parameters (temperature and RH). In addition to these main parameters, other studies have included other parameters, such as the level and type of gaseous pollutants present in the air (Menart et al.; 2014; Ligterink and Di Pietro; 2018) or the form in which paper decays (stacked or in single sheet) (Tétreault et al.; 2019). But, as noted by Strlič et al. (2013, p. 81), 'the dose–response function cannot be regarded simply as a sum of dose–response functions dependent on individual parameter'.

Dose–response functions are deterministic models: once the input is known, then the output of any point in time can be calculated. The damage function can be applied to a single object and to collections. In that case, the actual chemical characteristics of collections (i.e. pH and DP of the paper items) are introduced to explore different scenarios to support environmental management decisions (Duran-Casablancas et al.; 2017; Coppola et al.; 2020). The annual average of the temperature and RH is used as input for the indoor climate parameters.

To sum, the dose–response functions have a strong predictive character to explore the lifetime of collections depending on the chemical properties of the collections (DP and pH) and on the storage conditions where they are kept. This information can be used as a predictive tool for the life expectancy of collections, and, more importantly, it can be used to evaluate different (future) strategies for the preservation of the collections.

2.1.3 Indoor climate evaluation models

Other mathematical models have also been developed to support decision making, mostly regarding the effect of storage conditions on the chemical degradation of the collections (Cosaert and Beltran; 2021), such as the eClimateNotebook (Image Permanence Institute (IPI); 2018), developed by Image Permanence Institute (IPI), and Physics of Monuments (Smulders and Martens; 2014). These two models have been integrated into a web application that allows end users to enter the temperature and RH during a monitoring period in actual repositories into the model.

Like the damage function, the aim of these models is to explore the rate of deterioration (understood as life expectancy), depending on the environmental conditions where the objects are kept. However, there are two relevant differences between damage functions and indoor climate evaluation models:

1. The indoor climate evaluation models are developed as 'general' models for organic objects. Whereas in the damage functions, the individual characteristics of the objects are an essential input, in the evaluation models, short-lived materials (i.e. with a life expectancy of ca. 50 years at room conditions) are taken as 'average' objects.

2. The level of detail in the indoor climate evaluation models is not found in the collection data but in the environmental data. A crucial aspect of these models is that monitored indoor climate is not simplified to an annual average of the temperature and RH, but the variations in monthly intervals are taken into account to calculate the rate of chemical decay, as 'the nature of chemical deterioration is such that the collection has only "one life to live"; in other words, deterioration is irreversible' (Reilly and Zin; 1995, p. 8). Therefore, the reciprocals of life expectancy are used for the calculations, rather than the life expectancy values themselves (see Reilly and Zin (1995, p. 8) for how reciprocals are calculated).

In the evaluation models, the life expectancy is calculated based on the Arrhenius equation, where the time to deterioration depends on the temperature. The slope of the line on the Arrhenius plot (degradation rate) is represented by the activation energy of the reactions of deterioration. Each material has an activation energy. In the eClimateNotebook, a 'lower-middle' range is chosen as 'average' activation energy to determine the degradation rate. In the Physics of Monuments model, the activation energy for the degradation of cellulose (in the range

80 to 120 kJ mol^{-1} (Zervos; 2010)) is used. A series of Arrhenius plots have been developed at differing humidity values to include the effect of RH.

The eClimateNotebook preservation metric is the time weighted preservation index (TWPI), understood as the amount of time (years) that an organic material would last at any combination of temperature and RH. The preservation metric used by Physics of Monuments is 'lifetime multiplier' (LM). LM is defined as the number of time spans that an object remains usable when compared to a condition of 20°C and 50% RH.

As the input values in the model are the actual conditions, these models are primarily used to evaluate the effect of the actual conditions on the preservation of the collections and are used less as a tool to compare preservation strategies (as seen in the dose–response function section). However, because these models use detailed, monitored environmental conditions as input, these models report on other outputs that would not have been possible using the annual average T and RH, namely the risk of mould, mechanical damage for objects highly sensitive to RH fluctuations (e.g. wooden panels) and metal corrosion.

eClimateNotebook and Physics of Monuments are examples of static, time-invariant models. One generic output is given for the whole of the collection.

2.2 Input data

Risk assessments are classified within the analytic-deliberative mental process involved in decision making because the decision processes 'must be structured, evidence-based, and seek to be comprehensive' (Henderson and Waller; 2016, p. 308). Risks scenarios are systematically considered and evaluated. The ABC scores in risk assessment are quantified based on the data available: from statistics to estimate the frequency and intensity of rare events (e.g. earthquakes) to scientific and technical knowledge (e.g. to estimate the sensitivity of heritage assets to cumulative processes) (Michalski and Pedersoli; 2016). An important point is that, even though data collected in the past are used, confusion between past damage and future risks must be avoided.

If data are not available, a third source of information in risk assessments is local and common knowledge. The quantification of the risks and their effect (e.g. portion of collections affected by certain risk) is then estimated using a heuristic approach. Heuristics is an intuitive process for arriving at decisions, which 'resemble the subjective assessment of physical quantities such as distance or size' (Tversky and Kahneman; 1974, p. 1124). In heuristics approaches, information is ignored to make decisions faster, frugally and even more accurately than more complex methods (Gigerenzer and Gaissmaier; 2011). However, under uncertainty, judgements based on heuristics can lead to severe and systematic errors (Tversky and Kahneman; 1974).

This heuristic approach is even more present in some of the adaptations of the formal approaches to risk assessment, such as the QuiskScan. The aim of the QuickScan is to identify the risky areas within the collection by combining value with vulnerability and exposure to each of the ten agents of deterioration but without developing risk scenarios as the ABC method does. As the assessment is indicative of the main risks, the scores are strongly based on the experience of the assessors. The heuristic approach is further stressed when the value of collection units and vulnerability to each agent of deterioration are graded on a qualitative scale of high, medium or low. Although the use of heuristics is well accepted when conducting risk assessments, two main weaknesses have been identified regarding this approach: the great level of aggregation when risks are quantified and the tendency towards a biased input resulting in a biased output because the input is based on existing knowledge and experience and not on quantitative evidence (Brokerhof and Bülow; 2016).

To sum, the values used in the risk assessment equations are based on collected data (e.g. frequency of earthquakes) or heuristics (e.g. the portion of collections affected by a certain risk). Other values result from a deliberate, analytical approach, where mental (simulation) models are created to estimate, for instance, the effect of risk mitigation measures. We are making predictions and comparing different scenarios in risk assessments. Instead of formal mathematical models, we are using mental models that are difficult to communicate and, therefore, cannot be examined.

Mental models are seen as a powerful tool to make predictions derived from own experience (Klein; 1999). However, a limitation of mental models or simulations is that they rarely include more than three factors and six transitions states because '[...] we have trouble constructing mental simulations when the pieces of the puzzle get too complicated – there are too many parts, and these parts interact with each other' (Klein; 1999, p. 72). The main advantages of mathematical models over the best mental models have been clearly listed by Meadows and Robinson (1985, p. 6):

1. Rigor: the assumptions made in mathematical models are made explicitly and precisely. Mathematical models help to clarify and organise mental models because every single variable needs to be formulated.

2. Comprehensiveness: computers can compute more variables and their interrelationships, allowing the combination to be more comprehensive.

3. Logic: if programmed correctly, once a model has been developed, information can be processed with error-free calculations.

4. Accessibility: because assumptions are explicit, they can be communicated more easily. Critics can examine, assess and alter the model.

5. Flexibility: values of parameters and variables can be easily be adjusted to conduct experiments (what-if scenarios), which is less costly and time consuming than testing in the real system.

The mathematical models reviewed in the previous sections are based on experimental or empirical data. Once the model is developed, then input data collected in actual collections (e.g. climate data, chemical characteristics of the records) can be used to compare the output of the baseline (e.g. actual conditions) to other scenarios. Their limitation is that they include a limited number of parameters related to one single risk (e.g. chemical degradation due to environmental conditions).

As briefly discussed in the previous section, damage functions and indoor climate evaluation models differ in the level of detail in the input data. Whereas in the evaluation models, the focus is on the effect of variations in temperature and RH on an average type of object, damage functions explore the effect of individual characteristics of the objects (e.g. DP and pH) in (annual) average environmental conditions.

This difference in input data is related to the aim of each model. Whereas the damage function is interested in the effect of the individual difference of the objects forming a collection, evaluation models focus on the general preservation effect of changing the environmental conditions. Therefore, if we are interested in collections, understood as groups of objects with their own characteristics, then damage function models are the most suitable for this goal.

Evaluation models stress the importance of including the seasonal variations in the environmental conditions because, for instance, if an object is kept half the year in cold storage and half the year in room conditions, the preservation effect of these two storage conditions is not equivalent to their average. Therefore, the eClimate Notebook works with two metrics: the preservation index (PI) with fixed temperature and RH, and the time weight preservation index (TWPI), which take the variations of temperature and RH into account. From the examples given by Reilly and Zin (1995), it is clear that the difference between PI and TWPI become more clear the more extreme the climate variations are. This difference is not exclusive to the calculation of the PI and TWPI; it applies to any model that calculates the life expectancy of objects. We will observe the same pattern in the dose–response function for cellulose degradation on paper as an example. According to the function, paper (pH value of 6) kept at an average temperature of 18°C at 50% RH is expected to have an annual DP loss of 1.28 and 0.38 in the case of 12°C at 40% RH. The average of these two environmental conditions is 0.82, which is close to the DP loss of 0.702 if we had taken the average conditions of 15°C at 45% RH. However, the difference of working with an average becomes more notable when the conditions greatly differ. For instance, the DP loss of paper kept half the time in cold storage and half the time in room conditions is 0.69. Note this is not equivalent to the expected DP loss (0.39) if the average of 11°C at 45% RH is taken. From these examples, we can conclude that we need to consider large seasonal variations, if expected in the environmental data, to obtain a more realistic result. However, in archival institutions where the temperature and RH are controlled by HVAC systems, it is expected that working with an average temperature and RH would be a good approximation, even in those scenarios when some seasonal variations between winter and summer are allowed.

Interestingly, the currently in-development HERIe platform (Jerzy Haber Institute of Catalysis and Surface Chemistry; 2020) is being built to assist the user in the assessment of risks and deterioration processes. This platform seems to aim to bridge both approaches, the dose–response function of paper and the detailed indoor approach of IPI's eClimateBook. The detailed environmental data are considered as time weight values as seen in the TWPI. The focus in this model is on the actual conditions; it is less of a strategic tool to explore the effect of different preservation conditions on the lifetime of collections.

2.3 Time horizon

One requisite for risk assessments is to determine the time horizon. This is directly related to sustainability: '[the exploitation of] heritage resources in a way that the benefits extracted through use (which can lead to resource loss, e.g. by increased handling or display) do not exhaust the resource and reduce future access.' (Bell et al.; 2018, p. 510). Three planning horizons have been identified for different contexts of preventive conservation decision making (Bell et al.; 2018):

1. Normative (what should be done; planning for inter-generational equity): a period of 100 years

2. Strategic (what could be done; planning large estates and infrastructure projects): a period of 30 years

3. Operational (what can be done in short-term project planning): a period of 5 years.

Risk management operates within the normative planning horizon of 100 to 300 years. Time horizon is important because risks that lead to a loss of value over the long term score a lower urgency. Therefore, the time horizon needs to be long enough to capture the events with a low likelihood but with a high impact and the loss of value resulting from degradation processes that occur at a slow rate and will result in a higher accumulative impact (Michalski and Pedersoli; 2016).

Using the damage function, demographic plots can be built that are a deterministic model of linear, deterministic and continuous relationships showing the deterioration rate of a collection of individual objects over time. Deterministic models can be solved analytically, and once the (causal) interactions are formulated in equations and the initial conditions are fixed, the one possible outcome is calculated. Any time horizon can be chosen and calculated to infinity. However, if long time horizons are chosen (e.g. hundreds of years), uncertainty increases with the time horizon, and the statements and urgency become meaningless. The study by Strlič et al. (2015a) investigated when in the future it is deemed acceptable for an item to become damaged or unfit for use. A long-term planning horizon of 500 years was identified as acceptable by $\sim$ 90% of users defined as visitors or readers.

Very differently, the output of the eClimateNotebook and Physics of Monuments is static; only one output is presented, the expected end of the useful life of objects.

2.4 Dealing with uncertainty

Uncertainty is inherent in risk assessments because decisions are taken under three constraints: limited information available, the limited capacity of human minds to evaluate and process the available information and the limited amount of time to make a decision. There is always the wish to collect more data to make more informed decisions. At a certain point, one must accept the limitations ('bounded rationality') (Michalski and Pedersoli; 2016, p. 29).

In Waller's opinion (2003), uncertainty justifies the need to conduct a formal risk analytical approach because it enables rational decision making under uncertainty. Waller has also provided a taxonomic classification of uncertainties. He distinguishes five sources of uncertainty:

1. Ambiguity: omission of essential information, approximation of quantities or ideas, or linguistic uncertainty.
2. Disagreement: results from experimental research are not totally in agreement with one another.
3. Volitional uncertainty: it is always uncertain whether a person or organization will do what has been agreed.
4. Epistemic uncertainty: lack of knowledge
5. Variability: a certain variability that is present in a collection or the stochastic characteristic of the events.

Of these uncertainties, mostly those that can be captured in values, such as variability, are explicitly included in risk assessments. For example, variability is mostly captured during the risk analysis, when both a score for the most likely scenario or value and the highest and lowest possibility (high and low estimate) are included. More elaborated sensitivity tests using Monte Carlo simulations to test parameter uncertainty can also be conducted (Waller; 2003).

Dose–response functions are useful due to their predictive ability. However, predictions are inevitably associated with uncertainty because they are based on extrapolation (Strlič et al.; 2013). Menart et al. (2014) conducted uncertainty propagation analysis to study the uncertainty related to the Arrhenius slopes when predicting the expected lifetimes of different types of paper. Although the results showed broad uncertainty intervals, the authors concluded, 'on the basis of the observed trends and differences between investigated environmental conditions, guidance can still be given to inform decision makers, similar to other domains where despite uncertainties long-term decisions need to be made, e.g. climate change' (Menart et al.; 2014, p. 3711).

2.5 Presenting the results

The output in risk assessments is the numerical representation of the possibility of loss of value or 'likelihood of a hazard causing damage or loss' (Brokerhof et al.; 2017, p. 12). All inputs are converted to one single outcome, the magnitude of risk. Although risk assessment is seen as a quantitative method due to the numerical outcome, the interpretation is qualitative: inputs are approximated values, given in a step scale (e.g. five-step logarithmic scale or a low, medium, and high scale).

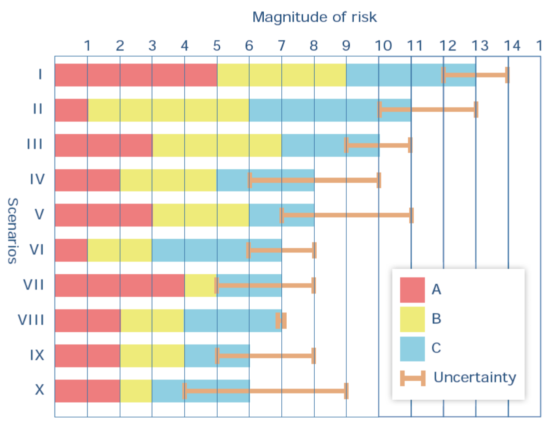


FIGURE 2.1: Example of a stacked horizontal bar graph ranked according to the total magnitude of risk and showing the individual contributions of the A-, B- and C-scores and the uncertainty (Brokerhof et al.; 2017, p. 38)

Location Dataset	Date Range	Natural Aging	Mechanical Damage	Metal Corrosion	Mold Risk	T°F	%RH	DP°F	TWPI	%DC Max	%EMC Min	%EMC Max	MRF
Alabaster, AL	2002-11-12 to 2012-03-30	RISK	RISK	RISK	RISK	62.9	72.7	52.6	22	2.22	11.4	19.3	27.54
History of Medicine	2008-03-04 to 2008-06-10	OK	RISK	OK	GOOD	70.7	28	35.6	70	0.78	4.5	7.2	0
Law	2008-03-10 to 2008-10-01	RISK	RISK	RISK	GOOD	71.5	46	47.9	36	2.04	4.6	12	0.03
Music Room	2002-11-12 to 2010-01-21	OK	RISK	OK	GOOD	72.5	35	42.6	48	1.58	4.3	10	0
Music Storage	2008-03-27 to 2008-12-30	RISK	RISK	RISK	RISK	68.5	56	51.4	35	1.81	8.3	14.8	1.41
Pottery Storage	2009-10-22 to 2012-02-23	RISK	OK	RISK	GOOD	70.5	50	50.5	33	1.35	6.9	11.8	0.15
Average (6 locations)						69.4	47.9	46.8	40.7	1.6	6.7	12.5	4.85

FIGURE 2.2: Output of the eClimateNotebook.

The outcome of one risk can be compared to the outcome of other risks in lists and maps. One way to represent the outcome graphically is a risk matrix, according to effect (loss of value combined with the fraction of the total collection value) and probability (the period within which damage is expected). A (stacked) bar chart is used when the magnitude of risk is presented as the combination of the ABC scores. Similarly, the effect of different scenarios to mitigate a certain risk can also be visualised in a bar chart (Fig. 2.1).

In the eClimateNotebook and Physics for Monuments, the calculated preservation metrics resulting from the input climate data in repositories are also converted to a risk category to simplify the interpretation of the preservation metric. Regarding the metrics of natural ageing, a TWPI larger than 75 is considered 'good', a TWPI between 46 and 75 'ok', and a TWPI lower than 46 is 'at risk' (Fig. 2.2). In Physics for Monuments, an LM > 1 is seen as a small risk, an LM between 0.5 and 1 as a medium risk and an LM < 0.5 means a high risk (Martens; 2012, p. 370).

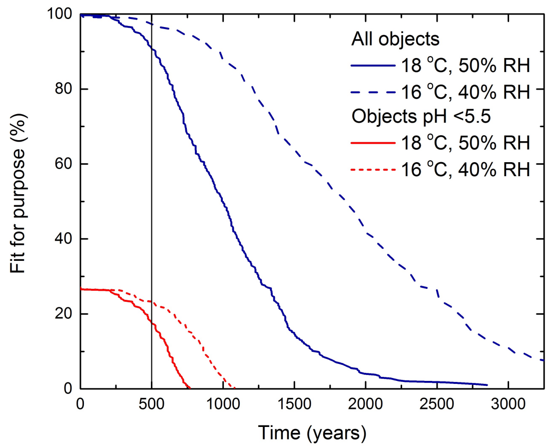


FIGURE 2.3: Demographic plots developed for the collection of the Amsterdam City Archives (Duran-Casablancas et al.; 2017).

The dose–response functions are presented as isoperm and isochrone plots (Strlič et al.; 2015b; Oriola et al.; 2011; Fenech et al.; 2012, 2013). In addition, damage functions are also used to build demographic curves predicting the lifetime of collections over time (Strlič et al.; 2015b; Duran-Casablancas et al.; 2017) (Fig. 2.3). Individual data (the pH and DP of the archival records) are collected in a survey to develop these curves. To inform preservation management decisions, the percentage of collections with a certain characteristic (in this case, fit for purpose) is shown over time. No value, and therefore no concept of risk, is included. The curves show the change over time, but the judgement of whether the change is acceptable is left to the users of the model.

2.6 Combining approaches

In the previous section, risk assessments and modelling have been presented separately, but it is clear that formal models can be a source to inform the A-score, the rate of occurrence of a degradation process, for a certain risk scenario. A good example of this application is given by Świątkowska et al. (2018). The authors analysed several risk scenarios for the collections of works on paper in the National Museum in Krakow. One of the identified risks was the loss of mechanical strength due to chemical degradation. In this study, the end of 'lifetime' is defined as the number of years till an object reaches the critical degree of polymerisation value (DP) of 200. The 'end of life' was calculated for papers representing DP_0 values between 500 and 1000 (characteristic for acidic paper) and around 2500 for bleached pulp papers at different pH values.

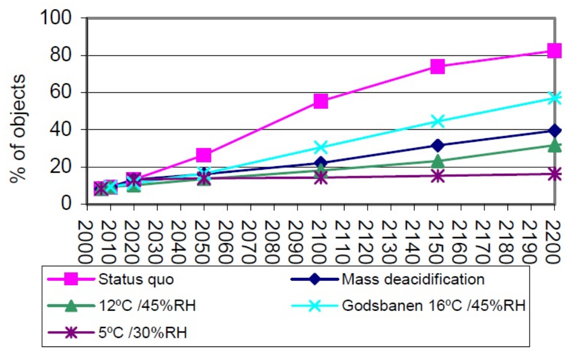


FIGURE 2.4: Comparison of preservation scenarios at the Danish Royal Library collections from 1850–1985 given by Vinther Hansen and Vest (2008).

In the example presented by Świątkowska et al. (2018), generic input about the characteristics of the collections is used in the mathematical model to inform the magnitude of risk. However, mathematical models can also be used to model the lifetime of collections with known characteristics. An example of this use is given by Vinther Hansen and Vest (2008). In this study, formal models are applied in conjunction with the specific characteristics of a collection to explore what-if scenarios for a better understanding of the consequences of the preservation decisions. Based on the age, pH and brittleness of the collection, the authors compare the effect of reducing temperature, RH, and deacidification (Fig. 2.4). Compared to the bar charts of the magnitude of risks, time plots (simulations) showing the lifetime of collections seem a more straightforward way to communicate due to the introduction of the time variable.

2.7 Adapting the existing models

This literature review reflects the major contributions made in recent decades to develop models and tools to support collection keepers with the complex decision-making process of preservation management. Thus far, the focus has been to approach preservation management mainly as single, independent measures. Such an approach does not always meet the complexity of the real world.

The complexity of decision making in archives and libraries is shown by the following examples. In 1993, Frieder observed that libraries and archives tended to adopt a 'wait and see' attitude towards mass deacidification for three main reasons: 1) digitization will make it unnecessary to preserve paper; 2) the maintenance of excellent environmental conditions will produce the same result as deacidification, and 3) if we wait a few more years, the technology will be better (Frieder; 1993, p.58). In 1995, Kiesling (1995) disputed these arguments because thus far, research had not shown evidence that maintaining the collections at a certain temperature and RH would have the same effect as deacidification. She also pointed out that by the time digitisation is a well-embedded preservation measure, 'we may have lost valuable information. Can we afford to wait?' (Kiesling and Stroud; 1995). Thirty years later, the effect of environmental conditions on life expectancy has been extensively studied and translated into models. More evidence has been collected on the effect of mass deacidification. What is still lacking is a tool that allows end users to compare these measures to make well-informed decisions.

Another example of the challenges that institutions face is illustrated by Varlamoff, who observed that in the future, budgeting and optimising costs will become an important issue in institutions (Varlamoff; 2004). Varlamoff remarked on the importance of calculating the costs on a long-term basis and stated that it is even more important to understand the relationship between the costs of microfilming and the money saved by reducing remedial treatments.

Preservation management is closely related to other processes within archival institutions whose goals might differ from each other. Ashley-Smith said institutions are confronted with 'the need to optimize four competing and interconnected outcomes – stability, cost, sustainability, and accessibility. In such a complex system, the optimal solution will probably be sub-optimal to at least one of the components' (Ashley-Smith; 1999). These dependencies between outcomes are well known, but they have not yet been formally explored. The current approaches do not capture this complexity due to the limited number of parameters included in the present models because the current mathematical models are based exclusively on experimental data and because the mental models used in risk assessments are inherently limited.

In 2006, M. Foot (2006, p. 21) stated that '[p]reservation does not exist in isolation. It is linked to other library and archives functions and a preservation policy should take those functions and activities into account'. Therefore, we believe that the next challenge in preservation management modelling is to build models where preservation measures are not approached in isolation but in relation to the other archive and library functions in the broader context of collections management.

2.8 Conclusions

In this chapter, we reviewed the use of risk assessments and other mathematical models developed to inform preservation management decisions. The shared aim of these approaches is to inform decision making without automating the decision. The presented models and tools are built to calculate the risk of occurrence of certain degradation processes and the effect of preservation measures on risk mitigation. The end user uses this information to form opinions and, in the end, make decisions. An important aspect of these approaches is that the outcomes, although presented quantitatively, are not used as predictions but as qualitative comparisons of certain risks and the impact of preservation measures to mitigate a certain risk.

In view of the findings of this literature review, and bearing in mind the limitations as well as the potentialities of each approach, the Archival Preservation Management (APM) model will be developed according to these three key elements:

1. *One formal mathematical model, including experimental models and mental models*:
   The APM model will use existing experimental mathematical models extended with variables and relationships found in the mental models used by experts when conducting risk assessments. One major advantage of formal models is that they force users to be precise and explicit both in terms of the model structure and when reporting the results. Assumptions and uncertainty are still present, but unlike a mental model, they are clearly reported.

2. *An integrated approach to the management of collections, including costs, preservation, and access, among other elements, and, where needed, integrating different functions of the institutions*:
   The present models do not offer a frame wherein the implications of preservation measures are seen in the broader context of collections management. In the next two chapters, the importance of approaching the preservation management of collections as a complex system will be further discussed.

3. *Modelling the effect of the different strategies during the lifetime of collections*: Regarding the time horizon, risk assessment applies a normative (100 years) plan horizon. However, when modelling degradation processes that take place at a very slow rate, time horizons of 500 years or even longer are needed to detect the effect of different preservation measures. If other processes are included in the model, such as digitisation, perhaps a strategic horizon of 30 years would be more appropriate. Therefore, the model needs to include the time variable as output to both accommodate the different time horizons of the preservation measures and as a clear way of showing the condition of the collections evolves over time, depending on the chosen preservation strategy.

In the next chapter, we identify the main variables that should be included in the model when preservation management is approached not in terms of individual, independent measures but as a process that is part of a complex system. Therefore, mental models will be converted into a formal model. As we will see in the next chapters, the major challenge of modelling is to include the main variables without building large models that are not used. As Meadows said: 'The modellers keep asking: "Why don't you use our model?" The policymakers respond: "Why don't you make models we can use?"' (Meadows and Robinson; 1985, p. 7).

References

Ashley-Smith, J. (1999). *Risk Assessment for Object Conservation*, Butterworth-Heinemann.

Bell, N., Cassar, M. and Strlič, M. (2018). Evidence for informed preservation planning and advocacy. A synoptic view, *Studies in Conservation* **63**(sup1): 8–14.

Brokerhof, A. W., Ankersmit, B. and Ligterink, F. (2017). *Risk management for collections*, Cultural Heritage Agency of the Netherlands, Amesfoort.

Brokerhof, A. W. and Bülow, A. E. (2016). The QuiskScan-a quick risk scan to identify value and hazards in a collection, *Journal of the Institute of Conservation* **39**(1): 18–28.

Bülow, A. E. (2010). Collection management using preservation risk assessment, *Journal of the Institute of Conservation* **33**(1): 65–78.

Bülow, A. E. and Brokerhof, A. W. (2011). The QALY in collection care: A cost effectiveness approach to collection management, *in* J. Bridgland (ed.), *ICOM-CC 16th Triennial Conference Lisbon 19-23 September 2011: Preprints*, Critério, Almada.

Bülow, A. E., Stitt, J. and Brokerhof, A. W. (2018). I can see further now. Preventive conservation in a changing heritage world, *Studies in Conservation* **63**(sup1): 35–42.

Coppola, F., Brown, N., Amicucci, F., Strlič, M. and Modelli, A. (2020). Non-destructive collection survey of the historical Classense Library. Part II: Conservation scenarios, *Heritage Science* **8**(1): 1–10.

Cosaert, A. and Beltran, V. (2021). Comparison of temperature and relative humidity analysis tools to address practitioner needs and improve decisionmaking., *in* J. Bridgland (ed.), *Transcending Boundaries: Integrated Approaches to Conservation. ICOM-CC 19th Triennial Conference Preprints, Beijing, 17–21 May 2021*, International Council of Museums, Paris.

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2017). Data mining in collections: From epidemiology to demography, *ICOM-CC 18th Triennial Conference, 2017 Copenhagen*.

Fenech, A., Dillon, C., Ntanos, K., Bell, N., Barrett, M. and Strlič, M. (2013). Modelling the lifetime of colour photographs in archival collections, *Studies in Conservation* **58**(2): 107–117.

Fenech, A., Fearn, T. and Strlič, M. (2012). Use of Design-of-Experiment principles to develop a dose-response function for colour photographs, *Polymer Degradation and Stability* **97**(4): 621–625.

Foot, M. (2006). Preservation policy and planning, *in* G. Gorman and S. Shep (eds), *Preservation management for libraries, archives and museums*, Facet, London, chapter 2, pp. 19–41.

Frieder, R. (1993). CIC Cooperative project. Logistics and managements issues, *Proceedings of the New York state seminar on mass deacification*, Columbia University Libraries, New York.

Garside, P., Bradford, K. and Hamlyn, S. (2018). The use of risk management to support preventive conservation, *Studies in Conservation* **63**(sup1): 94–100.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making, *Annu. Rev. Psychol.* **62**: 451–482.

Henderson, J. and Waller, R. (2016). Effective preservation decision strategies, *Studies in Conservation* **61**(6): 308–323.

Image Permanence Institute (IPI) (2018). eClimateNotebook. Fundamentals. Preservation metrics.
URL: *https://www.eclimatenotebook.com/fundamentals_nl.php*

Jerzy Haber Institute of Catalysis and Surface Chemistry (2020). HERIe. Quantitative assessment of risks to heritage assets.
URL: *https://www.herie.pl/*

Kiesling, K. and Stroud, J. (1995). Mass deacidification of archives and manuscript collections, *Advances in preservation and access. Volume 2*, Learned Information, Medford, pp. 77–100.

Klein, G. A. (1999). *Sources of power. How people make decisions*, MIT Press, Cambridge, Massachusetts.

Ligterink, F. and Di Pietro, G. (2018). The limited impact of acetic acid in archives and libraries, *Heritage Science* **6**(1).

Martens, M. (2012). *Climate risk assessment in museums. Degradation risks determined from temperature and relative humidity data*, PhD thesis, TU Eindhoven.

Meadows, D. H. and Robinson, J. M. (1985). *The electronic oracle. Computer models and social decisions*, John Wiley & Sons, Inc., Chichester.

Menart, E., De Bruin, G. and Strlič, M. (2014). Effects of NO_2 and acetic acid on the stability of historic paper, *Cellulose* **21**: 3701–3713.

Michalski, S. and Pedersoli, J. L. J. (2016). *The ABC Method. A risk management approach to the preservation of cultural heritage*, Canadian Conservation Institute, Ottawa.
URL: *https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf*

Oriola, M., Campo, G., Strlič, M., Csefalvayova, L., Odlyha, M. and Mozir, A. (2011). Non-destructive condition assessment of painting canvases using near infrared spectroscopy., *in* J. Bridgland (ed.), *ICOM Committee for Conservation, 16th Triennial Meeting, Lisbon, 19-23 September 2011. Preprints.*, Criterio - Produçào Gráfica, Lda,, Almada, Portugal, p. paper 1317.

Reilly, J.and Nishimura, D. and Zin, E. (1995). *New tools for preservation assessing long-term environmental effects*, Commission on Preservation and Access, Washington, D.C.

Smulders, H. and Martens, M. (2014). Physics of Monuments.
URL: *http://www.monumenten.bwk.tue.nldefault.aspx*

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015a). Damage function for historic paper. Part II: Wear and tear, *Heritage Science* **3**(1): 36.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015b). Damage function for historic paper. Part III: Isochrones and demography of collections, *Heritage Science* **3**(1): 40.

Strlič, M., Thickett, D., Taylor, J. and Cassar, M. (2013). Damage functions in heritage science, *Studies in Conservation* **58**(2): 80–88.

Świątkowska, B., Skoczeń-Rąpała, Ł., Okrągła, D., Jędrychowski, M. and Czop, J. (2018). Chemical degradation and physical failure: Risk analysis for a paper collection, *Studies in Conservation* **63**(sup1): 251–256.

Tétreault, J., Bégin, P., Paris-Lacombe, S. and Dupont, A. L. (2019). Modelling considerations for the degradation of cellulosic paper, *Cellulose* **26**(3): 2013–2033.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty. Heuristics and biases, *Science* **New Series, 185**(4157): 1124–1131.

Varlamoff, M. (2004). The future, *in* J. Feather (ed.), *Managing preservation for libraries and archives*, Ashgate, Aldershot, chapter 8, pp. 159–174.

Vinther Hansen, B. and Vest, M. (2008). The lifetime of acid paper in the collection of the Royal Library, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, pp. 38–39.

Waller, R. (1994). Conservation risk assessment. A Strategy for managing resources for preventive conservation, *Preventive conservation: practice, theory and research. Preprints of the contributions to the Ottawa Congress, 12-16 September 1994*, IIC, Los Angeles, pp. 12–16.

Waller, R. (1995). Risk management applied to preventive conservation, *in* C. Rose, C. Hawks and H. Genoways (eds), *Storage of natural history collections. A preventive conservation approach*, Society for the Preservation of Natural History Collections, Iowa, pp. 21–27.

Waller, R. (2003). Cultural property risk analysis model. Development and application to preventive conservation at the Canadian Museum Of Nature, *Goteborg Studies in Conservation* **13**.

Zervos, S. (2010). Natural accelerated ageing of cellulose and paper: A literature review, *in* A. Lejeune and T. Deprez (eds), *Cellulose: Structure and Properties, Derivatives and Industrial Uses*, Nova Science Publishers, Inc., New York, chapter 5, pp. 155–204.

Chapter 3

Qualitative modelling

'Causal attributions are a central feature of mental models. We all create and update cognitive maps of causal connections among entities and actors, from the prosaic -if I touch a flame I will be burned- to the grand -the larger the government deficit, the higher interest rates will be.'

Sterman 2000, p. 28

In the previous chapter, we stated that one of the key elements of the Archival Preservation Management (APM) model is that preservation measures will not be modelled in isolation but in the broader context of other archival and library functions. Adopting this integrated approach recognises that the management of collections is part of a larger system. In this chapter, following the Systems Thinking methodology (Dangerfield; 2014), we analyse the behaviour of this system using diagramming techniques. Through participatory sessions, we developed a causal loop diagram (CLD) to identify the key variables and the causal structure behind the system's dynamic behaviour (Sterman; 2000; Pruyt; 2013).

In the second part of this chapter, we use the drawn causal loop diagram to define the parameters (e.g. key variables and time horizon) of the APM model.

3.1 Causal loop diagram

3.1.1 Method

Three participatory sessions with a total of 25 participants from 14 Dutch archives and libraries were organised. The participants were conservators, collection managers, archivists and digitisation experts. As starting point for these sessions, the problem of interest was initially defined as follows:

What are the consequences of conservation measures and other archive and library functions on the preservation of the paper-based archival collections and their physical and digital accessibility?

FIGURE 3.1: Group model building session. On the screen the CLD is built together with the participants.

From the problem formulation, the participants agreed that three main variables could be seen as envisaged outputs of the model: collections `becoming vulnerable`, `digitally available` and `budget`. Taking these three variables as a point of departure, participants were encouraged to identify the variables that, according to their expertise and point of view, needed to be included to model the consequences of actions during the collections' lifetime. Through discussions, the participants sought to capture the underlying causal structure of the system. As facilitator, my role was to moderate the group discussion and find consensus on which variables should be included in the CLD and how these variables were related to those already present in the model. The participants were actively involved in the process of drawing the CLD, which could be followed on the screen (Fig. 3.1).

During each session, a causal loop diagram was drawn according to the following steps (Kim; 1992; de Pinho; 2015):

1. Variables were identified and named. The identified variables corresponded to measurable quantities of the system under study (for example, number of requests in the reading room), as well as soft elements (such as pressure for digitisation).

2. Arrows were drawn between two variables to show the direct causal relation among the variables.

3. A link polarity was assigned for each arrow. A positive link polarity (+) means "that if the cause increases, the effect increases above what it would otherwise have been, and if the cause decreases, the effect decreases below what it would otherwise have been [. . .] A negative link (−) means that if the cause increases, the effect decreases below what it would otherwise have been, and if the cause decreases, the effect increases above what it would otherwise have been" (Sterman; 2000, p. 139).

4. The most important feedback loops (symbolised by a loop identifier) were pointed out. Feedback loops are formed when elements are connected by a directed cycle of arrows. A reinforcing feedback loop will generate an exponentially increasing behaviour, while a balancing loop will generate equilibrating behaviour.

In each session, participants developed their own diagram. The diagrams have been included in Appendix A. During the session, certain areas of the model were developed in more depth, depending on how the discussions in each group evolved, as well as the time available per session. In the three-hour sessions participants identified the main variables and started thinking about the link polarity. Only one group had the opportunity to attend a second session. Having more time, this group could discuss in more depth about the link polarity as well as feedback loops, resulting in a more elaborated model (see Appendix A, group 2).

Since the main differences observed between the three diagrams could be mostly explained due to a difference in completeness rather than in the underlying causal structure of the diagram, which is very similar in all three diagrams, the three diagrams could be emerged to one single CLD (Fig. 3.2). The final CLD follows the most elaborated model built during the participatory sessions (Appendix A, group 2), with a few additional variables discussed by the other groups.

The causal loop diagram was drawn using the modelling software Vensim PLE (Ventana Systems Inc.).

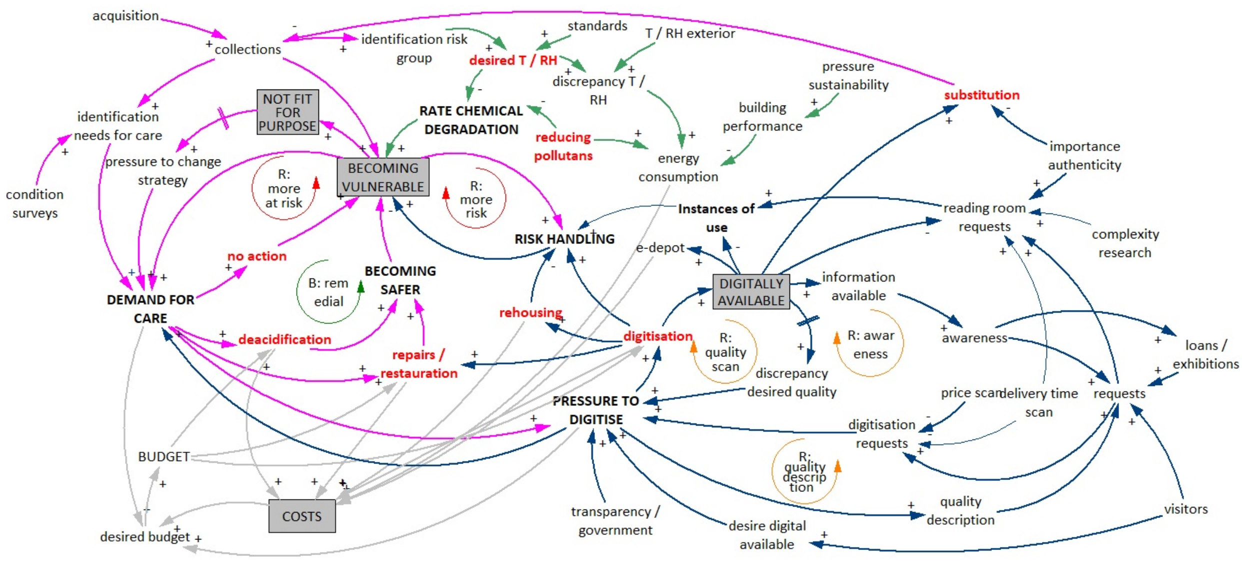


FIGURE 3.2: Causal loop diagram with the causal relationships between preservation management (pink and green arrows), access (blue arrows) and costs (grey arrows). Preservation measures are indicated in red. Grey boxes represent envisaged outputs of the model.

3.1.2 Results and discussion

The final version of the causal loop diagram (CLD) (Fig. 3.2) shows the variables and their relationships that participants deemed relevant to modelling the effect of preservation measures. In this section, following the CLD, we describe the discussions that took place when developing the CLD.

The CLD has been arranged so that the left side of the model includes variables related to the preservation aspects, while the right side captures the dynamic of collections use (Fig. 3.2). These two parts are directly linked to each other. Grey arrows show the variables related to the costs of the included actions. The presence of costs in the model makes clear that, according to the participants, one important aspect of the mathematical model is that it should allow the comparison of different measures with similar results regarding the preservation of the collections, but perhaps with a different impact on the institutional budget.

3.1.2.1 Preservation management

On the left side of the model (Fig. 3.2, pink and green arrows), the dynamics of how the condition of the collections changes during their lifetime are reproduced (`Becoming vulnerable` and `Unfit for purpose`). As the focus of this model is on the preservation of paper-based collections, object condition is defined by two main mechanisms: chemical and mechanical degradation. Participants agreed with the differentiation between `Becoming vulnerable` and `Unfit for purpose`. Whereas the former includes those objects at higher risk of mechanical degradation due to a low degree of polymerisation (DP) (e.g. DP $>$ 300) and/or accumulation of wear and tear), the latter are those for which, at each handling, there is a risk of loss of information due their critical DP (e.g. DP $<$ 300). [In Chapter 6, section 6.1.1, a brief introduction on the degradation of paper is given.]

The diagram shows how chemical and mechanical degradation are primarily determined by the decisions taken in two areas of the model: the storage conditions in the repositories (green arrows), and whether the objects are physically used (blue arrows). However, if the measures taken in these areas are inadequate or insufficient, then the group of objects at risk will inevitably increase (reinforcing loop `At risk`). At the same time, the risk of mechanical degradation will further increase (reinforcing loop `More risk`), due to the poor chemical properties of the materials, increasing the occurrence and accumulation of wear and tear.

Whereas the storage conditions in the repositories (green arrows) are a preventive conservation measure, the other actions that can be taken in this preservation area of the diagram (pink arrows) are remedial, and are taken once degradation has already occurred. The majority of participants observed that in their institutions remedial treatments are mostly undertaken when objects are requested for digitisation (right side of the model). This is because the preservation needs become clear when objects are used, so action is then taken.

Participants observed that it remains particularly difficult to anticipate which objects are at greater risk due to their characteristics, and to identify them within the kilometres of shelving that archives and libraries count. Therefore, participants are concerned that we probably are not using the preservation measures efficiently: vulnerable objects are still being used in the reading room while other objects at a lower risk have been digitised, and not further physically used. For this reason, participants fear that it is likely that, at some future point, the `Unfit for purpose` group will become notably visible. This is seen as an important turning point which will force institutions to critically review their preservation strategies.

3.1.2.2 Access of the collections

According to the participants, digitisation has dramatically changed how we use collections. Institutions feel an enormous pressure to make the collections digitally available. Besides the pressure from visitors and government to provide digital collection access, there are several reinforcing loops that increase this pressure even more. One is caused by the awareness of the information available: the more information available, the greater the awareness of it, resulting in more requests which generate more information as well. In addition, there is another reinforcing loop related to the quality of information available: the pressure to digitise results in also pressure to describe collection content, resulting again in more requests. In addition, in the long term, there is a third reinforcing loop showing that collections may be digitised again in the (near) future because the quality of current scans will not meet the quality expected by the users.

Collections are being digitised for different reasons, and while digitisation as a preservation measure is one of those, it is certainly not the main driver behind digitisation. This means that, although institutions might be expending part of their budgets on making the collections digitally available and, therefore, considerably reducing the need to use them physically in the reading room, it is still possible that those objects with a higher risk of mechanical degradation are not being targeted. That would be the case if, for instance, collections containing mostly rag paper are digitised, while collections with a higher risk of damage during handling, such as those containing (high lignin) wood-pulp paper post-1850, are used intensively in the reading room.

Participants also agreed that, besides the positive effect of digitisation on the reduction of handling, other benefits of digitisation projects include the realisation of conservation treatments, such as tear mending and deacidification, or rehousing. As downsides, it was mentioned that, too often, damage occurs during the digitisation process. Experts working in libraries also noted that, even when the objects are digitally available, visitors can still be interested in accessing the physical object. According to the participants working at archives, as long as scans are available for free, the majority of the readers find, in the digital scan, the information that are looking for, as they are mostly interested in the text, and have the added benefit of being able to access it online.

One of the discussions during the session focussed on the quality of the scans and the financial aspects of digital storage. Here, the question arose whether the model should include the option of substitution (where the physical record is replaced by its digital form) and whether the modelling of digitally born collections should also be included. Regarding substitution, this is (still) a controversial option in archives (Baker; 2001). Although there are guidelines on the selection of collections for substitution and the quality requirements for the digital image (van de Velde et al.; 2016; NVAO; 2020), it was not seen as an option for preservation management. On the contrary, the participants recognise the increasing importance of understanding the preservation of digitally born collections. It is clear that the participants wish to develop a preservation management tool that includes all types of collections, from archival records and bindings to architectural drawings, photography and digitally born collections. The participants agreed that their causal loop diagram, with minor differences, could be applied to broad types of collections.

Besides access to the collections, whether digitally or involving physical records in the reading room, other archive functions included in the diagram were acquisition, cataloguing and exhibitions, as these can increase the (physical) use of the collections.

3.1.3 General observations and conclusion

Participants found building a CLD very intuitive and straightforwards. However, one aspect of the CLD was found confusing and difficult to use correctly. As positive and negative signs are used to indicate the polarity of the links among the variables, a common mistake was to assume that an increase in a “cause variable" will actually result in the increase of the “result variable". Diagram polarities do not refer to the behaviour of the variable, but to the system’s structure (Richardson; 1986). Participants recognised that decisions are very often taken based on assumptions of this type (Dörner; 1990). However, drawing a causal diagram made the participants aware that, in order to understand the output of a certain variable, *all* the input variables need to be taken into account and every single variable needs to be quantified. Participants agreed that causal loop diagrams help to visualise the complexity of the system and the importance of taking all aspects into account in order to evaluate the effect of a certain decision. In addition, due to the visual clarity of the diagram, it was found a useful tool for communicating with colleagues from other departments.

Participants identified the storage conditions in the repositories (green arrows) as the most important leverage point in the system. It is well known that storage conditions (temperature and relative humidity) are key parameters affecting the chemical degradation of materials. However, participants strongly stressed that, although it is also well known that lowering the temperature and relative humidity can significantly reduce chemical degradation, particularly for acidic paper, we are not making full use of this preservation measure. In addition, lowering the temperature is also seen as an effective measure to reduce energy costs as, for instance, in the Netherlands the average external temperature is about 10°C. Although the potential benefits related to adequate storage conditions are clear, the participants noted that institutions often consider introducing changes in the actual storage conditions when there is a plan to build a new repository, which is not necessary driven from a preservation or cost-effectiveness point of view, but more probably by a shortage of storage capacity.

It was stressed that, despite the sound information that can be found in the literature, such as demography plots (Strlič et al.; 2015), the urgency of taking certain actions and how to prioritise preservation actions is a challenge. In participants' opinions, it is still extremely complex to determine at what point objects will become `vulnerable` or `not fit for purpose`, according to the use and the storage conditions in their particular cases and the effects of the preservation measures being taken. Condition surveys and risk assessments are tools meant to inform practitioners about the urgency of taking preservation measures but the information generated with these tools is seen as not precise enough.

Participants acknowledged the importance of mathematical models to support informed decisions. In particular, they recognised the potential of mathematical models to test what-if scenarios, e.g. to explore the effects of different conservation actions (understood as the risk of becoming vulnerable or unfit for purpose) within a certain period of time, taking into account the specific characteristics of the collections as well as their use within a particular institution. Thus, the conclusions drawn in Chapter 2, that identified the main gaps in preservation management modelling, were further supported by the experience and opinions of the preservation professionals.

3.2 Defining the boundaries of the APM model

Based on the findings of the CLD, in this section the purpose, the key variables and the time horizon of the mathematical model are summarised. The CLD forms the blueprint of the mathematical model that will be developed in the next chapters. Note that in this chapter we mainly enumerate the variables, but in the next chapters, we will expend on the justification of the choices made, based on the literature review and the experiments conducted for each part of the model.

3.2.1 Problem definition

The first step of any modelling process starts with establishing a clear purpose for the model (Sterman; 2000, p. 89). Over the course of the participatory sessions, it became clear that it is a matter of great concern for the professionals caring for the collections to develop an understanding of how the collections' state of preservation will evolve depending on what preservation measures are taken during the collections' lifetimes. In agreement with the findings of the Collections Demography project (Strlič et al.; 2015), in participants' opinion the threshold of fitness-for-use should refer mainly to mechanical degradation (tears and/or missing pieces), whereas other degradation phenomena such as discolouration are not necessarily seen as a major threat to collections use either now or in the future.

In addition, participants agreed that collections are preserved in order to be used and, inevitably, the use of the collections will eventually affect their preservation. Therefore, the outputs of the model should be not be in terms of preservation alone, but also in terms of access. Similarly, the modelling of these outputs should take into account other functions within the institutions that might affect collection use. The CLD shows that object condition depends not only on the preservation measures taken, but are also affected by other archive functions, particularly how the collections are accessed and used. The latter refers not only to access to physical objects, but also to whether the collections are digitally available which, in most cases, will have the beneficial side effect of reducing physical use. Likewise, the impact of the possible preservation measures on institutional budgets was also identified as an important output, in order to compare the costs and benefits of the different actions.

Therefore, the Archive Preservation Management (APM) model will explore the effects of preservation measures, taking into account not only the characteristics of the objects but also other activities within the institutions that might relate, directly or indirectly, to collection preservation. The model will be developed for paper collections (bound and unbound) and will report on three main outcomes: chemical and mechanical condition of the collections, physical and digital accessibility, and costs.

3.2.2 Key variables

3.2.2.1 Characteristics of the collections

One of the main outcomes of the model will be to understand at what rate the variable `becoming vulnerable` will increase over time. The CLD identified the following characteristics of the collections (or of the individual records) as relevant:

- mechanical, chemical and physical properties of the paper;
- type of protective enclosure;
- whether the records/collection are digitally available;
- their level of demand in the reading room.

How these initial characteristics change over time will be endogenously modelled in three parts of the model according to three activities: whether conservation and preservation measures are taken, the choice of storage conditions, and whether records are physically used. Looking at the effects of these three activities on the same collection should allow modelling of not only how collections accumulate wear and tear, but also how they do not accumulate tears when they are not used, even when the chemical condition is poor. This will allow to reproduce scenarios that have been observed in studies on actual collections, where, in collections involving paper with a low double-fold resistance, no signs of tears were observed simply because the books had not seen much use (Di Pietro et al.; 2016).

In addition, the chemical and mechanical properties of the collections also depend on the acquisition policy of the archive or library. The acquisition policy determines, in both the past and in the future, not only how many shelf-kilometres of collections need to be taken care of, but also the types of collections with their own management characteristics. In the CLD, acquisition is treated as an exogenous variable.

3.2.2.2 Preservation and conservation options

Keeping in mind that the CLD describes collection preservation mainly in terms of the objects' mechanical properties, two main preservation options were included: environmental storage conditions and rehousing. Rehousing of library and archival collections refers to providing objects with adequate protective enclosures, since enclosures are seen as an effective measure to prevent mechanical damage from occurring and to mitigate the risk of dust, typically by boxing (Wilson and VanSnick; 2017). Regarding the storage conditions, whereas humidity, temperature and pollutants were identified as important variables in the CLD, light and pest management were not regarded as essential for inclusion.

Regarding remedial conservation, measures to correct physical or chemical deterioration, two main measures were mentioned during the participatory sessions: (mass-)deacidification which addresses the chemical degradation of the paper, and repairs which correct physical deterioration, such as tears or missing parts. Mass strengthening treatments (Zervos and Alexopoulou; 2015), combined with deacidification and reinforcement of paper (Dupont and Cheradame; 2008; Vizarova et al.; 2008), have been proposed in the past. However, these treatments, which have not been broadly applied so far, were not mentioned during the participatory sessions, and will not be included in the model.

To sum up, the CLD identifies the following variables for inclusion in the model:

- temperature, relative humidity (RH) and level of pollutants in the repositories;
- rehousing;
- mass-deacidification;
- repairs.

3.2.2.3 Access

There are two main variables in this part of the model:

- reading room requests;
- requests for digitisation.

These two variables are key to modelling the physical use of the collections, which can potentially cause mechanical degradation. If the archival records are not digitally available, then users can choose between accessing the records in the reading room, or requesting them for digitisation if the archive offers a scan-on-demand service.

Following the CLD, `instances of use` is modelled endogenously by taking into account the percentage of the collection `becoming digital available`. The focus in this part of the model is to capture when the digitisation of the collections will result in a reduction of readers in the reading room, due to an increasing part of the collections is becoming digitally available.

Regarding the modelling of the percentage of the collection `becoming digitally available`, the CLD identifies two main variables:

- Pressure to digitise;
- Willingness of the readers to access the collections digitally.

These variables could be defined as soft variables. During the workshop, participants identified the following factors to explain visitors' preferences for accessing the records in the reading room or for requesting digitisation: price of scan (if it is not free), complexity of the research, the delivery time of the scan, and importance of authenticity. In Chapter 5, we will conduct user studies to explore the importance of these variables.

Another interesting question that arises in the access part of the diagram is whether improving access to the collections, whether by making the available online, improving the cataloguing, or displaying the collections in exhibitions, will result in an increase in the (physical or digital) use of the collections. However, this type of variables relate to information that can be difficult to quantify, and, for now, they will be not included in the model.

3.2.2.4 Costs

Including costs in the model provides a financial frame for comparing different options, e.g. costs of digital storage vs physical storage vs deacidification. In addition, energy consumption has become an important issue, not just due to constant pressure on the institutional budget, but also due to the increasing awareness in heritage institutions that they can also make a contribution to major issues such as sustainability (e.g. by lowering energy consumption in both traditional repositories and in digital storage) (Kramer et al.; 2016)

Regarding the costs of activities in the model, some can be directly calculated, such as the costs of deacidification, digitisation or digital storage. However, some other costs, such as energy costs in repositories, are more complex to calculate, where the desired temperature and RH, as well as the characteristics of the building and the climate control system (HVAC) need to be taken into account. As will be discussed in Chapter 7, these calculations will require a submodel within the main model.

3.2.3 Time horizon

In modelling, the time horizon needs to be long enough to capture feedback loops as well as to evaluate the effect of policies (Sterman; 2000). In this model, variables can develop over very different time frames from decades (digitisation programs, lifespans of HVAC systems or building envelopes (Henry; 2013, p. 20)) to centuries (chemical degradation of paper collections, climate change projections). For each part of the model, we will investigate an adequate time horizon and the best way to report results with different time frames.

3.3 Conclusions

Developing a CLD in collaboration with professionals working at archival and library institutions proved to be an effective method to refine the purpose of the model as well as its structure. In view of the CLD, the purpose of the model will be to simulate and compare the effects of different preservation and conservation measures during the lifetime of collections, where mass-deacidification, remedial conservation, digitisation and storage environments are the preservation measures of interest. The model will include other archival functions, such as acquisition and access, to explore potential synergies as well as counter-intuitive outcomes of decisions made within the same institution, but by different departments.

In our project, the purpose of the model is to evaluate different preservation measures according to the following criteria: their preservation effect, impact on the institution's budget, and their effect on the collection accessibility (physical and/or digital) in the short and long term. Therefore, the model will be approached as three sub-models:

1. The preservation sub-model will address the change over time in the percentage of collection materials showing mechanical degradation, understood as the accumulation of tears, and chemical degradation, understood as loss of the DP value, depending on the (preservation) measures taken;

2. The access sub-model will address the use of the collections in the reading room, depending on the digitisation projects;

3. The costs sub-model will calculate the impact of the (preservation) measures on the institution's budget.

The discussions with the practitioners during the development of the causal loop diagram pinpointed, in particular, the following aspects which will require further investigation while developing the mathematical model:

- The required level of detail related to the characteristics of the collections need to be defined for each part of the model. We can think of variables related to the chemical and mechanical properties (e.g. DP and pH), but also whether collections are often requested or are digitally available.

- The CLD-identified (soft) variables, such as the willingness of readers to access the collections digitally instead of in the reading room, which may be first further explored to evaluate whether they should be included in the model.

- The model will include different time horizons, from decades to centuries. A possible option is to merge different time horizons into one model that would be used initially to simulate short-term outcomes, and then to explore longer-term effects.

In the next chapter, we review the main simulation modelling approaches to explore which would most adequately serve to convert the CLD developed in this chapter into a mathematical model.

References

Baker, N. (2001). *Double Fold: Libraries and the Assault on Paper*, Random House, New York.

Dangerfield, B. (2014). Systems thinking and system dynamics. A primer, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discrete-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., chapter 3, pp. 26–51.

de Pinho, H. (2015). Participant guidelines systems tools for complex health systems. A guide to creating causal loop diagrams, *Technical report*, Mailman School of Public Health, Columbia University, New York City.

Di Pietro, G., Ligterink, F., Porck, H. and de Bruin, G. (2016). Chemical air filtration in archives and libraries reconsidered, *Studies in Conservation* **61**(5): 245–254.

Dörner, D. (1990). The logic of failure, *Philosophical Transactions of the Royal Society B* **327**(1241): 463–473.

Dupont, A. and Cheradame, H. (2008). Simultaneous deacidification and reinforcement of paper by treatment with aminoalkylalkoxysilanes, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, p. 37.

Henry, M. (2013). Conservation environments, museum buildings, and sustainability, *Summit on the Museum Preservation Management*, Smithsonian Institution, pp. 20–24.

Kim, D. (1992). Guidelines for drawing causal loop diagrams, *The Systems Thinker* **3**(1): 5–6.

Kramer, R. P., Schellen, H. L. and van Schijndel, A. W. (2016). Impact of ASHRAE's museum climate classes on energy consumption and indoor climate fluctuations. Full-scale measurements in museum Hermitage Amsterdam, *Energy and Buildings* **130**.

NVAO (2020). Besluit vervanging archiefbescheiden NVAO 2019.
URL: *https://wetten.overheid.nl/BWBR0043168/2020-02-14*

Pruyt, E. (2013). *Small system dynamics models for big issues*, TU Delft Library, Delft.

Richardson, G. P. (1986). Problems with causal-loop diagrams, *System Dynamics Review* **2**(2): 158–170.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*, Mcgraw-Hill Education, Boston.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015). Damage function for historic paper. Part III: Isochrones and demography of collections, *Heritage Science* **3**(1): 40.

van de Velde, M., Karelse, J., Kiens, P. and Windhorst, M. (2016). Handreiking vervanging archiefbescheiden, versie 2.0, *Technical report*, Groninger Archieven. **URL:** *https://www.nationaalarchief.nl/sites/default/files/field-file /handreiking_vervanging_archiefbescheiden_v2.0.pdf*

Vizarova, K., Kischnerova, S., Suty, S., Tino, R. and Katuscak, S. (2008). Strengthening and deacidification of acidic groundwood paper with the ternary system chitosane-methyl-hydroxyethyl-cellulose- cationic starch in Mg(HCO3)2Aqueous solution, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, pp. 53–54.

Wilson, H. and VanSnick, S. (2017). The effectiveness of dust mitigation and cleaning strategies at The National Archives, UK, *Journal of Cultural Heritage* **24**: 100–107.

Zervos, S. and Alexopoulou, I. (2015). Paper conservation methods. A literature review, *Cellulose* **22**: 2859–2897.

Chapter 4

Simulation modelling

> *'Systems analysts tend to see analogies between systems with different elements but similar relationships, analogies that would be missed by those who concentrate on elements alone. For example, ... the same simple first-order differential equation can be used to describe the depreciation of industrial capital, the decay of radioactive materials, the inventory-ordering policy of a beer distributor, or the cooling of a cup of hot coffee.'*
>
> Meadows and Robinson 1985, p. 22

In the previous chapter, we approached preservation management processes in archives and libraries as a system. We developed a causal loop diagram to identify the key variables and the causal structure behind the dynamic behaviour of the system. In this chapter, we discuss whether the preservation management of collections can be defined as a complex system, and if so, which model paradigm could most adequately model this system. In this chapter, the concepts of systems, models and simulation will be defined in order to clarify what we mean by these terms, as depending on the field where they are used, they can have a slightly different connotations.

4.1 (Complex) systems, models and simulation

In the standard work *'Thinking in Systems'*, Meadows (2009, p. 11) defines a system as 'an interconnected set of elements that is coherently organized in a way that achieves something.' From this definition, we learn that systems consist of two kind of entities, the elements and the relationship between these elements, having a purpose or function.

The management of archives can clearly be defined as a system in Meadows' sense. As seen in the causal loop diagram drawn in the previous chapter, elements are linked to each other in order to achieve some purpose (e.g. caring of the collection for the present and future generations). However, can it be argued that the system that we defined can be categorised as a *complex* system, and, if so, what does it mean and what are the implications?

It is outside the scope of this thesis to examine the origins of the notion of systems thinking and complexity; for a brief overview, sufficient for our needs, see Marchal et al. (2013, p. 2) and Preiser (2019). Let us move directly to the distinction made between simple, complicated and complex systems, and why this is relevant in the context of the management of collections.

Glouberman and Zimmerman (2002) use the following example when differentiating between *simple*, *complicated* and *complex* systems, according to the problem that the system addresses: whereas following a recipe would be characterised as simple problem, sending a rocket to the moon is a complicated problem, and raising a child is a complex problem. Simple systems require mastering certain techniques or terminology, but once these are acquired, success is highly assured just by following the recipe. In complicated systems, the formulae are critical and necessary, and high levels of expertise in a variety of fields can be necessary. But once the system has been formulated, there is a high degree of certainty in the outcome, and the model is transferable to similar systems. In complex systems, on the other hand, formulae have limited application because different initial conditions will result in different outcomes, and there is no assurance of success when applying the same conceptualisation to the next system, due to the level of ambiguity and uncertainty in the elements. Every system is unique, and must be understood as an individual. Hence, all living systems, psychological systems, and social systems are defined as complex (Poli; 2013).

The distinction between a *complicated* system and a *complex* system lies in knowing what we can expect of the outcomes of the two systems. Whereas in complicated systems the problems can be (permanently) solved, in complex systems they can be modified but not solved, since intervention generates new problems

(Poli; 2013). Therefore, when dealing with complex systems, Meadows' advice is that the best approach is to 'learn dancing with them' (Meadows; 2009, p. 170).

It is worth noting that systems can be complicated and complex at the same time. As Goldenfeld has observed, 'nature can produce complex structures even in simple situations, and can obey simple laws even in complex situations' (Goldenfeld and Kadanoff; 1999). However, a unique characteristic of complex systems is that the behaviour of complex system cannot be explained by the sum of the effects of each part. Such explanation applies the 'reduction principle', which is valid only in those systems with a low level of complexity (Wierzbicki; 2007). We can use simplified (smaller) models to study a part of the system, but this will not necessarily provide a level of description that answers our questions about the overall system's behavior (Goldenfeld and Kadanoff; 1999). However, completeness should not be a goal when modelling. As Oberkampf et al. (2002, p. 338) have observed, 'the predictive power of a model depends on its ability to correctly identify the dominant controlling factors and their influences, not upon its completeness or complexity. A model of limited, but known, applicability is often more useful than a more complex model'.

In complex systems, it is not unusual that sub-purposes act in conflict with the overall purpose. Measures taken in certain parts of the model may produce an undesirable outcome for the whole system. As mentioned in the previous chapters, Ashley-Smith has remarked that heritage institutions are trying to find the balance between four competing and interconnected outcomes—stability, cost, sustainability, and accessibility—and, according to him, a optimal solution for one particular outcome will be a compromise for another (Ashley-Smith; 2013, p. 30). A challenge in complex systems is thus keeping sub-purposes and overall system purposes in harmony (Meadows; 2009, p. 16). By recognising that preservation management of archival collections can be understood as a complex system, we implicitly acknowledge that the problems cannot be solved. Nevertheless, we can still identify the places to intervene in the system which might turn in a leverage point (Meadows; 1999). Here is where Meadows' advice comes into practice: learn dancing with the system.

In general terms, complex systems have one or more of the following characteristics: large numbers of elements, large numbers of relationships among elements, nonlinear and discontinuous relationships, and uncertain characteristics of elements and relationships (Rouse; 2015). In particular, these systems are often characterised by their dynamic complexity. Such complexity arises from the interactions between the parts of the system, often leading to counterintuitive behaviour of complex systems (Forrester; 1971). Sterman (2006) has listed some of the causes of this dynamic complexity and why we are not good at understanding them. For example, one characteristic of these systems is nonlinearity (where the effect is rarely proportional to cause), as not just one factor, but usually a group of factors are involved in determining outcomes. Complexity also arises because the system is history-dependent, in other words, many actions are irreversible. In complex systems, cause and effect are distant in time and space resulting, sometimes, in counter-intuitive effects. Similarly, long- and short-term effects usually differ from each other, leading to policy-resilient decisions where an intended effect may have unintended consequences.

In healthcare, the dynamic complexity of causes of disease emerges from certain characteristics (Marshall and Galea; 2015, p. 93). For the characteristics in these systems, there are parallels in the archival and library collections, supporting the concept that complexity can be applied to collection management in archives and libraries:

- 'Structural non-linear relationships (e.g., phase transitions) between causes and outcomes': in the case of heritage collections, agents of deterioration should not be studied in isolation, and the synergies among them (e.g. chemical degradation and the effects of mechanical forces during handling) should be taken into account (Taylor; 2005).

- 'Feedback loops, such that causal effects are magnified (i.e., positive feedback) or dampened (i.e., negative feedback) as disease processes progress': providing access to collections (in the short term) without allowing for their preservation needs, can have unintended long-term consequences, for instance, increased treatment costs.

- 'Adaptivity, in that individual and population behaviour can evolve based on past history' and ' a high degree of sensitivity to initial conditions': the initial conditions of the objects (paper quality) and their past history (whether the records have been heavily used or rehoused) can lead to different outcomes over the lifetime of the collections.

One of the main characteristics of complex systems is their dynamic behaviour over time, due to nonlinearity. In a causal loop diagram, non-linear behaviour arises from positive–negative feedbacks as well as delays. Senge (1990) has identified some of the most common behaviour patterns that are common in complex systems: archetypes such as 'limits to growth', 'escalation' or 'eroding goals' have been described, and two of them are of particular interest for us because they fit nicely with some of the observations made during the development of the causal loop diagram in Chapter 3:

Shifting the burden: 'A short-term 'solution' is used to correct a problem, with seemingly positive immediate results. As this correction is used more and more, more fundamental long-term corrective measures are used less and less. Over time, the capabilities for the fundamental solution may atrophy or become disabled, leading to even greater reliance on the symptomatic solution' (Senge; 1990, p. 380). Preventive measures are taken (e.g. digitisation), however there is concern that these measures are not being used efficiently (e.g. vulnerable objects are still being used in the reading room, while other objects with a lower risk have been digitised and are hence insulated from further physical use), and that measures to retard long-term chemical degradation of the paper are further delayed. Therefore, in the long term the part of the collection becoming unfit might become increasingly visible.

Success to the successful: 'Two activities compete for limited support or resources. The more successful one of them becomes, the more support it gains, thereby starving the other' (Senge; 1990, p. 385). When digitisation gains more attention and, therefore, more resources, the budgets available for other preservation measures (e.g. remedial treatment) shrinks.

In the case of collection management, most of the single relationships are linear (e.g. DP vs accumulation of wear and tear; temperature and RH vs DP). But, as pointed out during the development of the causal loop diagram, complexity arises because one output is the result of multiple inputs interacting over time (e.g. wear and tear depends on mechanical characteristics *and* use; and mechanical characteristics depend on the chemical condition which depends on storage conditions as well as the initial pH and DP of the paper items).

Complex systems are usually studied through simulation. If a mathematical model is simple enough, framing the problem in a well-understood form, then an exact, analytical solution might be worked out. But, in the case of complex systems, such solution is not possible and numerical simulation is used to investigate how inputs affect the output measures of performance (Law; 2007, p. 6).

Definitions of simulation give a clear emphasis to the concept of 'experiment'. For example, according to Chung (2003, p. 1) 'simulation modelling and analysis is the process of creating and experimenting with a computerized mathematical model of a physical system'. More specifically, simulation has also been defined as 'the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behaviour of the system and/or evaluating various strategies for the operation of the system' (Shannon; 1998, p. 7). Hence, simulation offers the possibility to use the model with the purpose of policy design: the model is used to test a number of policies or hypotheses, and to clarify why different results are obtained, if it is accepted that the model (at least roughly) reproduces the behaviour of the real system. Simulation models are regarded as 'tools for systematically attaining a better understanding of the real world and thereby for providing support for addressing the real-world management decision-making problems' (Rotaru et al.; 2014, p. 85). The results are intended to inform policies and decision making and, in some cases, to optimise the system (Preston White and Ingalls; 2016). Such purpose is very close to the purpose defined for the APM model, where the aim is to compare the effects of collection management options over time.

4.2 Simulation paradigms

In this section, we review the main approaches developed to model complex systems. First, a brief introduction is given of the three major paradigms in modelling: discrete-event simulation (DES), system dynamics (SD), and agent-based modelling (ABM). Then, while reviewing the different aspects of the simulation approaches, we hypothesize about how the most distinctive characteristics of each paradigm could be applied in constructing our model. We also introduce the hybrid model, which combines one or more approaches in a single environment. For all the reviewed modelling approaches, a mathematical formulation with equations is needed to permit a quantitative approach to the problem under study. Therefore, in the last part of this section, we briefly review the type of data required by each approach.

Although literature on SD, DES and ABM can be found in a very broad number of fields, for the literature review presented in this section we mainly focusses on the modelling of healthcare systems. Healthcare deals with the challenges of managing (preventive and remedial) measures for groups with certain diseases and, therefore, the parallel can be easily made to the management of heritage collections.

It is already widely recognised that it is beneficial to model healthcare problems as complex systems (Plsek and Greenhalgh; 2001; Homer and Hirsch; 2006; Lipsitz; 2012), evidenced by the increasing number of studies using approaches such as SD and ABM over the last two decades (Liu et al.; 2018). The systematic literature review by Salleh et al. (2017, p. 940) distinguishes three main applications of system dynamics in healthcare: (1) resource management to optimise health service flow, (2) modelling the effects of policy interventions for effective decision making, and (3) modelling infectious disease, which comes in two distinct kinds: biological modelling of disease progression within an individual, and epidemiological modelling over populations. Point (2) and (3) in the epidemiological sense align precisely with the aim of this research, which can also be expressed via the words of Cassidy et al. in another review of simulation models of healthcare systems: 'to reduce undesirable patient outcome (mortality and hospitalisation)' and 'to reduce the total cost of care' (Cassidy et al.; 2019). In summary, the common conceptual framing between healthcare systems and heritage systems indicates scope to use simulation modelling to improve decision-making in collection management.

The discussed articles in this section have been selected based on the presented models which may be potentially seen as examples for models in the heritage field.

4.2.1 Paradigms description

There are three major paradigms in modelling (Fig.4.1):

1. *Discrete-event simulation, DES* (middle–low abstraction; meso–micro level). The DES approach was originally developed to improve the design and the operation of manufacturing plants. Although DES has found a broad set of applications, there is generally a strong focus on improving the design and operation of the system at a workflow level. For example, DES models have typically been used to improve supply chains, capacity in hospitals, call centres and other processes where the operation of the system can be improved by a better understanding of the relation between queuing and available resources, caused by events that occur in discrete time.

The building blocks in DES models are entities that flow through a network of queues and activities, where resources are shared between activities. Entities typically represent customers arriving at some service facility, or workpieces arriving at a machine. The model is built as a set of interconnected activities and queues that are subjected to random (stochastic) variation. In a DES model, the dynamic behaviour of the system emerges from the queuing structure and from variability in the system (Robinson; 2014).

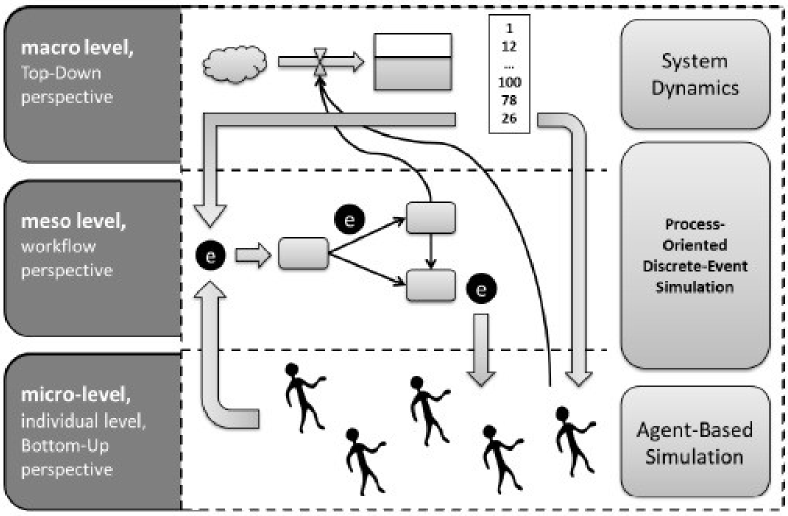


FIGURE 4.1: Scheme describing the three major paradigms in simulation given by Djanatliev (2013). Whereas in system dynamics the system is approached at a macro-level, discrete-event simulation focuses on the workflows at a meso-level, and a micro-level approach is adopted by agent-based modelling. The scheme also introduces the possibility of creating a multi-paradigm model in a common environment, where different approaches are combined in one model.

2. *System dynamics, SD* (high abstraction; macro-level): SD was originally developed to simulate industrial supply chain problems. In his *Industrial Dynamics,* J.W. Forrester formulated the SD approach to study the 'information-feedback characteristics of industrial activity to show how organizational structure, amplification (in policies), and time delays (in decisions and actions) interact to influence the success of enterprise' (Forrester; 1961, p. 13).

SD model represent processes in terms of stocks (of material or knowledge), flows between these stocks, and information (auxiliaries) that determines the value of the flows. The stock levels constitute the state of the system, and the flow rates are influenced by external decision processes, and therefore might change the system state. An important aspect of SD-modelled systems is that their behavioural patterns emerge endogenously, due to feedback loops (Schieritz and Milling; 2003). The structure of a system determines the system's dynamics.

3. *Agent-based modelling, ABM* (from micro- to macro-level): ABM concepts are based on cellular automata (Weimer et al.; 2016) and complex adaptive systems (Macal and North; 2009). This approach has mostly been applied to model human social and organizational behaviour and individual decision-making (flow simulation, organizational simulation, market simulation, and diffusion simulation) (Bonabeau; 2002).

   ABM models complex systems 'approaching them as a composition of interacting, autonomous 'agents'... Agents have behaviours, often described by simple rules, and interactions with other agents, which in turn influence their behaviours. By modelling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviours can be observed as it gives rise to the behaviour of the system as a whole' (Macal; 2010, p. 371). The behaviour emerges 'as a result of many (tens, hundreds, thousands, millions) individuals, each following its own behaviour rules, living together in some environment and communicating with each other and with the environment' (Borshchev and Filippov; 2004). A key aspect of ABM is to capture the heterogeneity of agents across a population (Macal; 2010).

Table 4.1 summarises the key characteristics of the three major modelling approaches.

The characteristics and applications of these three paradigms have been extensively compared in the literature. In health care, the suitability of SD and DES, depending on the study's aim, has been analysed (Chahal and Eldabi; 2010; Brailsford et al.; 2014), and taxonomies have also been developed, providing guidance to modellers choosing between SD and DES (Barton et al.; 2004; Brennan et al.; 2006; Cooper et al.; 2007). SD and ABM have been also compared in several studies (Scholl; 2001; Schieritz and Milling; 2003; Rahmandad and Sterman; 2008; Macal; 2010; Lättilä et al.; 2010; Luke and Stamatakis; 2012; Swinerd and McNaught; 2012) and literature can also be found where all three paradigms are

TABLE 4.1: Key characteristics of the three major modelling approaches.

	DES	SD	ABM
Basic building block	entities, queues, activities, resources	levels (stock), rate (flows), explanatory variables (auxiliaries), feedback loops	agents, agent relationships (rules), environment
Unit of analysis	queuing structure	structure	rules
Level of modelling	meso	macro	micro
Level of information	individual entities with attributes	aggregates	heterogeneous individuals
Perspective	workflow	top-down	bottom-up
Handling of time	discrete	continuous	discrete
Elements	stochastic	deterministic	stochastic
Mathematical formulation	–	ordinary differential equations (ODEs)	thresholds, algorithmic rules, if–then rules
Origin of dynamics	events, causing queuing	internal causal structure, causing accumulations	many individuals following their own behavioural rules

described and compared (Borshchev and Filippov; 2004; Gunal; 2012). Furthermore, the literature abounds with examples of systems that can be modelled with either SD or ABM, taking advantage of each approach's capabilities (Macal; 2010; Cimler et al.; 2018; Ahmed et al.; 2012; Borshchev and Filippov; 2004; Swinerd and McNaught; 2012). But the most profound difference is that 'different modelling paradigms cause their practitioners to define different problems, follow different procedures, and use different criteria to evaluate the results' (Meadows and Robinson; 1985, p. 20).

A good example of how a topic can be modelled by different approaches, but with slightly different focus depending on decision makers' requirements and the type of problem, is the study by Currie et al. (2020). In this study, simulation modelling methods (DES, SD, ABS and hybrid simulation) are suggested depending on the identified decisions regarding how the impact of COVID-19 could be reduced. This publication shows very clearly that, although same topic can be modelled by the three approaches, the appropriateness of each method depends on the decision to be made. For example, for the decision regarding targeting of vaccinations, SD (for a population-level model) as well as ABS (for models that capture individual behaviour) are applicable. However, for the delivery of vaccination, a DES approach is suggested.

To illustrate that the same topic can be modelled by the three simulation methods, but with a slightly different focus, we conducted a small experiment. We searched in Google Scholar for articles since 2020 (search conducted on 26-11-2021) on simulation modelling in COVID-19 studies. A search of the name of the modelling method and COVID (e.g. 'Agent based' AND 'COVID') resulted in 5740 matches for 'system dynamics', 8190 for 'agent based', and 1330 for 'discrete event simulation'. Then a third search term was added (policy, transmission, resources, capacity, strategy), terms typically associated with one of the methods. The results of this search, shown in Fig. 4.2, indicates that the concept of policy making is more present in SD studies, operational terms as resources and capacity are more frequently found in DES studies, whereas transmission between individuals is more frequently included in ABS studies. Interestingly, the concept of strategy is equally found for all three methods.

In the next sections, to gain insight into how a single problem might be approached by each paradigm, the following distinctive characteristics of each paradigm

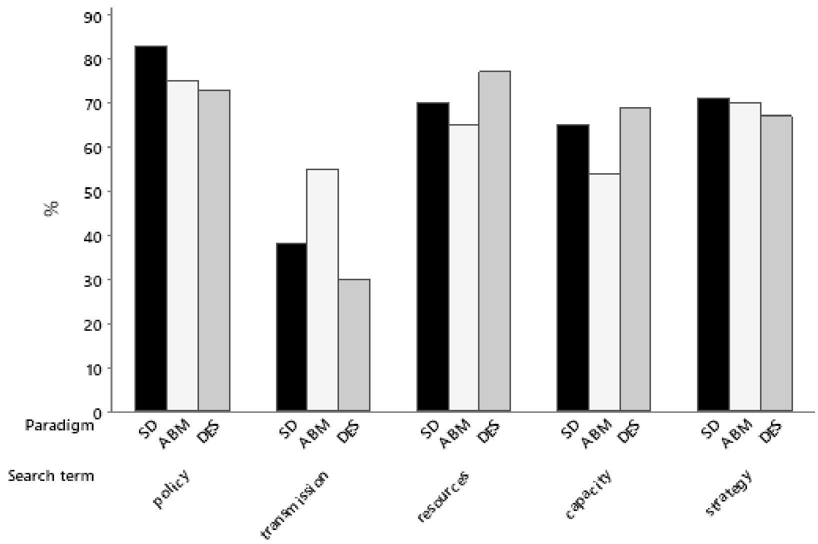


FIGURE 4.2: Results of a search in Google Scholar (conducted on 26-11-2021), showing the percentage of articles since 2020 that include the term 'Covid' and the name of one of the simulation methods in combination with one of the search terms.

are discussed: heterogeneity, randomness, handling of time, and how the dynamic behaviour of the system under study emerges.

4.2.2 Heterogeneity: Aggregates, individual entities or agents

According to Barnes and Chu (2015, p. 19), 'a system is heterogeneous if: (i) it consists of different parts, and (ii) these parts do not necessarily behave according to the same rules/laws when they are in different states'. Most of the systems have a certain level of heterogeneity and, in some systems, in this case in a biological system, 'it would make no sense to assume an 'average' lion with some 'mean' behaviour [..] Lions behave differently depending on their age, their rank within the group and their gender' (Barnes and Chu; 2015, p. 19). When modelling health care systems, Brailsford et al. (2014, p. 111) also observed 'a fascinating difference in the way that clinicians engage with DES and with SD. Initially, people are more attracted to DES because it accords with their fundamental belief that all patients are different and should be treated as individuals'. It might seem that capturing this heterogeneity is necessary in order to build a model closer to the system in the real world, but, as has been pointed out in section 4.1, modelling should not attempt to reproduce exactly the system under study, but to identify the dominant controlling factors. In some models, that may involve the heterogeneity present in the system, but this is not necessarily the case for all complex systems. A misconception in modelling is that 'more-detailed models are necessarily better models' (Barnes and Chu; 2015, p. 7).

Heterogeneity is approached very differently by each modelling paradigm, resulting in different levels of detail when describing stocks, entities or agents:

- SD takes an aggregate view. Stocks are accumulations which are altered by inflows and outflows. These accumulations are regulated by the flow rates. However, as inflows and outflows are regulated by different variables (decisions), they differ from each other, causing dynamic disequilibria (Sterman; 2000). In SD, items in any one stock are indistinguishable; they are not individually described. Stocks are conceptualised as tanks holding a liquid: they have no attribute and are regulated only by the rates of flow.
- In DES, the entities are individual items that flow through the system. The individual items are not individually described, but attributes (specific features) can be assigned for individual entities. 'The attributes can be used to determine the logic of the model, for instance the time an individual entity will spend in an activity, its priority in a queue, or its route through the system' (Robinson; 2014, p. 14).
- Agents in ABM present the highest level of detail. Usually agents are described by a state chart, and events trigger transitions between states. Agents are not homogeneous, and differences between them lead to dynamic behaviour in the system (Weimer et al.; 2016). Therefore, it is important to determine the numbers of agents needed, yet making the simulation computationally efficient and giving correct results.

Disease-related models based on system dynamics are referred as 'compartmental models', and use an a 'aging chain' structure (Darabi and Hosseinichimeh; 2020). Reinforcing/balancing feedback loops are important to capture dynamics of infectious diseases in SIR (susceptible/infectious/recovered) models (Ahmed et al.; 2012). Stocks represent populations with average properties that undergo different stages of the illness, and flows between groups can be probabilistic (e.g. rates of infection and recovery) (Davahli et al.; 2020). Activities or events can have different effects depending on the states of the items. SD copes with such structural detail by using extra stocks. In health care models, for example, age group, or gender, or disease status are associated with different probabilities of becoming infected or mortality. In order to include the different states of items with different probabilities, the models can become extremely complex, with as many as 80 population stocks (Hirsch et al.; 2010). This is a well-known difficulty of SD (Borshchev and Filippov; 2004; Brailsford et al.; 2014). However, this aggregate view of SD allows abstraction from single events to concentrate on policies instead. In SD, each causal relationship in the system is formulated mathematically. The mathematical formulation makes it possible to explore how changes in parameter values impact the model's overall behaviour.

In order to explore the effects of different policy decisions, SD has been broadly applied in health care to model groups with different risk probabilities. In these models, the health and economics impacts of policies or interventions are modelled for groups that differ in exposure or/and risk probabilities. One example is the HealthBound model developed to simulate intervention scenarios in the US health system (Milstein et al.; 2010, 2009). HealthBound includes a software interface that allows stakeholders to simulate the impacts of diverse scenarios over a 25-year period on morbidity, mortality, health equity, and cost. This model offers a clear example of SD whose emphasis is on policy rather than predictions. The aim of the model is not optimisation or prediction, but to identify how small changes to one part of the system can have considerable impacts elsewhere. In this model, the entire US population is divided into two groups—the advantaged and the disadvantaged—and further subdivided among three states of health.

A more complex example is given by Mahamoud et al. (2013), who developed a simulation model for changes in health, social determinants, and disparities in Toronto, Canada, to project forward from 2006 to 2046 under different assumptions. In this case, there are 30 population segments represented, stratified by ethnicity, immigration status, and gender. The model has over 400 active equations, 300 input constants and time series, and over 2300 output elements (due to the breakdown of the model variables into 30 subgroups).

Simulation models have been used to examine alternative interventions for chronic illness (Homer et al.; 2004) or chronic conditions such as diabetes (Jones et al.; 2006) and heart failure. The Prevention Impacts Simulation Model (PRISM), for example, simulates the multiyear impacts of 50 intervention levers (more than 4000 variables) aimed at reducing cardiovascular disease risk (Hirsch et al.; 2014).

An interesting aspect of this study is that the effect of the interventions is compared for the entire U.S. versus a high-poverty county, showing that the sensitivity of the results depends on the initial conditions.

These examples illustrate that, when subgroups are created to capture the heterogeneity of the stocks, the complexity of the model greatly increases. Another factor that increases the complexity of the model is the number of variables, for example, to compare different interventions. In the model proposed by Diaz et al. (2015), the model is simplified by building two co-flows that simultaneously consider the health and financial consequences by contrasting the impacts pre- and post-intervention.

When a model needs to capture the behaviour of the individuals and interactions among individuals, it can become quite complex to model the system mathematically. In those cases, a computational model, such as ABM, can build a model closer to the system under study, for example, in epidemiology studies (Galea et al.; 2009; Auchincloss and Roux; 2008). The traditional 'regression models do not allow us to take into account the inter-relations, reciprocity, or discontinuous nature of the relations that likely underlie the determination of behaviour in the 'real world" (Galea et al.; 2009, p. 4). One of the advantages of the ABM approach is that this approach can capture the dynamic complexity of disease causes, for example, (positive) feedback loops that magnify causal effects as disease processes progress, and the degree of sensitivity to initial conditions (Galea et al.; 2010). ABM is usually used when there is interaction among the agents, typically in a pre-determinated space. A clear example is the importance of modelling interference (i.e., one person's exposure affects the outcomes of others) (Marshall and Galea; 2015). The agents are placed in a grid, where the spatial dimensions and the number of agents will affect the interrelations among agents.

ABM has been broadly used in the field of infectious disease transmission (*MIDAS. Models of infectious disease agent study*; n.d.). However, object-orientated microsimulation has also been developed as causal system modelling in chronic disease epidemiology where there is no interaction among the agents (Ness et al.; 2007). In this case, microsimulation is used to capture the heterogeneity of the agents. One example is the Archimedes model, which includes 'physiology models' and 'models of care processes', originally designed for health care applications, but also applicable to natural systems in general (Schlessinger and Eddy; 2002). In Archimedes, agents have features, and every feature has a value which changes over time. The 'trajectory' of the feature is defined mathematically as its value as a function of time. The trajectory of a feature can be affected by the agent's characteristics, behaviours and other features (risk factors). As a feature progresses, it can cause certain 'events' to occur (outcomes). Interventions can change the value and/or the rate of progression of one or more features. The occurrence of outcomes can set in motion a wide variety of 'logistic events'. ABM allows a description of each individual and to explore the dynamic interactions among agents and environment, reflecting emergent behaviours (Nianogo and Arah; 2015)

Homer et al. (2014) have briefly compared the results of the system modelling approach of PRISM to the microsimulation of Archimedes (Kahn et al.; 2008) and concluded that 'the simplified compartmental structure of PRISM means that it cannot capture certain detailed comorbidity effects and distributional effects that one might see in a microsimulation like Archimedes. The advantage of the compartmental approach is that a simulation can be performed in an instant, and thorough sensitivity testing of dozens of assumptions thereby becomes practical' (Homer et al.; 2014, p. 5).

Since DES has typically been used to model on an operational level with the aim of prediction or optimising specified performance criteria, the number of articles related to the modelling of diseases (as seen in the previous examples) is more limited. However, DES has also been used to model patients going through different states in a SIR model for spread of disease (Brailsford and Hilton; 2001).

The SIR models simulate susceptible, infected and recovered individuals in terms of infection risk or preservation measures, and are an excellent example of how SD and ABM will approach the same topic very differently (Macal; 2010). At the same time, it is a good example for finding parallels to the case of archival and library items that may change their state (from good condition to critical condition). In SD terms, the archival records (single sheets or inventory numbers) form the stocks. As SD does not distinguish between the individual features of the objects, then the stocks represent average features of the records (for example, average DP below, or above, 800). If other features are added, for example, whether the records are digitised or not, then more stocks need to be created to combine the different features. In contrast, the strength of ABM is that the heterogeneity of the agents is an essential element of the model. In an individual agent's state chart, rich information can be held. This information (eg. on the condition of the agent (DP, tears), access mode (reading room or digitisation)) forms the agent's history, and determines the behaviour of the agent—an archival record—over time.

In the case of archival and library collections, it is clear that representing archival items as individual agents with their own behaviours and state charts, instead of using aggregates, is closer to the real world. Doing so, the (initial) characteristics of the records (e.g. type of paper) are taken into account, and different events (e.g. physical use) may also have different effects depending on which state the records are in. To model this level of detail, ABM might be the more appropriate approach. The combination of state (physical and chemical) with use (level of demand) might be challenging to translate into a SD model.

4.2.3 Randomness: Deterministic vs stochastic elements

Of the three paradigms under study, only SD is a deterministic model. SD is a mathematical model whose equations capture the mean behaviour of a system or the probability of a specific event taking place at a specific time. Delays can only be represented as pipeline or batch delays, or as exponential delays (Brailsford and Hilton; 2001).

SD models are based on the assumption that system structure leads to system behaviour, and that any observed variation in that behaviour is a result of interactions between system structures. Once the (causal) interactions are formulated in equations and the initial conditions are fixed, the one possible outcome is calculated within seconds. There is no need to run a SD model many times, unless the values assigned to system parameters vary. However, some precaution is needed when interpreting the results from variables with uncertain parameters. This limitation is solved by some SD software which allows the modeller to specify that the value of a parameter is sampled from a probability distribution (Pidd; 2014).

In SD, random features in the real world are averaged out to deterministic equations. As Barnes and Chu observe, 'this means that there is discrepancy between the assumptions made in deterministic models and reality. This discrepancy by necessity leads to an error of the model whenever nature happens to be discrete' (Barnes and Chu; 2015, p. 207).

In contrast, DES and ABM models are built to capture the randomness in the system. In ABM, the randomness originates in the heterogeneity (variability) of the agents, their rules and interactions. In DES models, randomness typically occurs in the length of time an activity takes or in the arrival of entities (Robinson; 2014).

The randomness in the system is described by stochastic features. 'Stochastic' denotes something that varies in a probabilistic manner through time, and is represented in a simulation model by appropriate probability distributions. The simulation is run over time and repeated numerous times to obtain a distribution of possible outcomes for the specified system (Auchincloss and Roux; 2008). Due to the stochastic features, the outcome of two simulation experiments, even if they use the same configuration parameters, may be quite different. How different they are will depend on the particular variations that occur in each run and on how sensitive the system is, given its initial conditions (Pidd; 2014). It is necessary, therefore, to repeat experiments multiple times in order to gain (statistical) confidence in the significance of the results obtained. When using DES models, Monte Carlo simulation is usually used to obtain these multiple-run outcomes.

The major drawback of ABMs is their computational cost (Barnes and Chu; 2015). Depending on how complex the system is and the number of agents, ABMs can take a very long time to simulate. For example, a typical calculation for a medical centre serving 300.000 people, described by the Archimedes model discussed above, takes 24–72 hours (Schlessinger and Eddy; 2002). It is not expected that such long run times will be needed for the model that we are developing, but it is clear that, due to the amount of information to be processed, the average run time is expected to be higher in ABM models compared to SD models whose simulations are typically run within seconds.

In the course of this thesis, we will learn that, in the case of archival system, examples of deterministic and stochastic events can be found. For example, the dynamics of DP and pH can be approached using a deterministic model: once the actual DP and pH of the paper, as well as the storage conditions (temperature and relative humidity), are set, it is possible to predict the degradation rate of the paper collections. However, other events or activities (e.g. digitisation or reading room requests) occur erratically, and can be modelled as a stochastic process during the life time of the collections.

4.2.4 Handling of time: Continuous or discrete?

In SD the state changes are regarded as (quasi-)continuous because the rates and levels change only at discrete, predefined points in time (as the simulation takes a time-slicing approach) and are held constant between these points (Brailsford and Hilton; 2001). In other words, the rates, which represent an activity or decision and control the levels, are constant over a time interval (*dt*). In DES, the events occur at irregular intervals. Therefore, the simulation proceeds from event to event, and the interval between those events vary (Pidd; 2014). While an entity is being worked on during an activity, the state of the system remains unchanged. If events occur regularly (e.g. once each day), then DES will be simulated by time-slicing.

ABMs can be built as continuous (time-driven) or discrete (event-driven) models. In the case of time-driven models, the handling of time is similar to SD models, where time progresses in discrete time steps of fixed duration, resulting in a quasi-continuous handling of time. In the event-driven approach, the time between successive events can vary (Barnes and Chu; 2015).

4.2.5 How does the dynamic behaviour emerge?

The three paradigms clearly differ on how they approach the generation of dynamic behaviour. In the DES paradigm, the dynamic behaviour is explained by the events which occur at discrete times, resulting in queuing due to limited resources. Therefore, this approach is usually used for optimisation or prediction. For example, if the aim of the model is to reduce the delivery time of requested archival records in the reading room, DES would be the most appropriate approach.

In the SD paradigm, dynamic behaviour is explained by the structure of the system, and particularly by feedback loops. Therefore, an exponential process can be explained as due to the presence of a reinforcing feedback loop. Oscillation, on the other hand, is usually due to a balancing feedback loop followed by a delay. In section 4.1, we introduced the system archetypes defined by Senge (1990), which help us to recognise patterns in complex systems. These archetypes are used in SD to show how small changes to one part of the system can have a considerable impact elsewhere in the system. Once the structure is fixed, then the parameters can be varied to understand the effects of various measures. The focus of this approach is on policy design rather than optimisation or prediction.

SD models do not serve well for optimisation or prediction, as they rarely yield accuracy above 40% (Lane; 2000).

As a bottom-up approach, ABM 'uses the model to observe the emergent behaviour instead of using the model to explain the behaviour' (Weimer et al.; 2016, p. 66). The dynamics emerge not only due to the interactions among the agents but also due to agents' past experiences (Gilbert; 2011). In this approach it could be possible, for instance, to model how mechanical degradation in archival collections accumulates due to physical use, in how the agent's behaviour varies with experiences that, in this case, arise from being handled in the reading room. The characteristics of the individual archival items thus explain the dynamic behaviour of the system, rather than preservation decisions such as reducing physical use.

As earlier discussed, '[i]t is suggested that the choice of modelling method is made based on decision makers' requirements, type of problem and system complexity and its characteristics' (Currie et al.; 2020, p. 84). One might argue that characteristics of the preservation management system in archival collections are more favourable for an ABM approach, because the heterogeneity of the collections are a key element of the model. If we take the decision makers' requirements and problem type as formulated in Chapter 3, where we defined the aim as being to gain a better understanding of the effect of the preservation measures on the collections, a bottom-up approach seems also to meet these requirements. In the ABM approaches, the collections will be central in the model.

Were the aim of the model to address policy making, then SD models might be a more suitable approach (Ghaffarzadegan et al.; 2011). In that case, we would be more interested in knowing how preservation decisions generates. In that case, SD will offer the right elements to build the model: first to capture the behaviour patterns (exponential growth/decline, s-shaped growth or oscillations) due to the feedback loops and delays present in the model structure, and, secondly, to introduce the information variables in the causal loop diagram (e.g. using pressure to digitise or readers' willingness to use digital scans to model the increase in digitisation requests).

4.2.6 Hybrid models

Hybrid models use different simulation paradigms in a common environment. This approach is still in development as it has emerged only recently. Traditionally, there have been two different communities for SD and DES, working separately, with their own journals and conferences (Lane; 2000). The potential of combining SD and DES has been discussed since 2001 (Brailsford and Hilton; 2001). In the same year, a clear call was made for cross-study and joint research due to the potentials of combining SD with ABM (Scholl; 2001). According to Heath et al. (2011), the best results have been achieved when DES is combined with another paradigm. Good overviews of studies discussing the potential of combining SD and ABM as practical examples of this hybrid approach are given by Swinerd and McNaugth (2012) and Guerrero et al. (2016), and, for SD and DES, by Morgan et al. (2017).

According to Swinerd and McNaugth's (2012) classification of hybrid models (specifically for SD and ABM), integrated hybrid systems is just one group within the hybrid models, alongside interfaced and sequential models. The integrated class includes continuous, fluid processes incorporating feedback between modules. In both interfaced and sequential models, modules for ABM and SD run separately, but, in the former, the outputs of the two modules are combined at different points of the run to represent the desired output as a function of time, and in the later one module is run by itself, and its output then fed to the next module. One of the advantage of running separate modules is that computational costs can be reduced when the (growing) number of agents considerably affects the computation time (Vincenot et al.; 2011). For example, only a limited number of agents can be simulated in ABM, so the model swaps to SD when a performance threshold is breached.

Within the integrated hybrid models, Swinerd and McNaught (2012) further distinguish three architectures:

1. In an ABM module, the internal structure of the agents is modelled as an SD model (agents with rich internal structure);
2. A stock within the SD module is an aggregate measure of an ABM module (stocked agents);
3. A parameter within the SD module is an aggregate measure of an ABM module (parameters with emergent behaviour).

In the classification given by Vincenot et al. (2011), the stocked agents and the parameters associated with emergent behaviour are described together as a single architecture: agents interacting within a single SD model. To illustrate this class, Vincenot et al. (2011) use the example of fishes (individual agents) in a lake (modelled as SD). The fishes' behaviour varies according to the dynamic lake properties. SD stock triggers a state chart transition which, in turn, modifies the stock value.

Borshchev's work (2013; 2004) shows all the possibilities for combining DES, SD and ABM (Fig. 4.3).

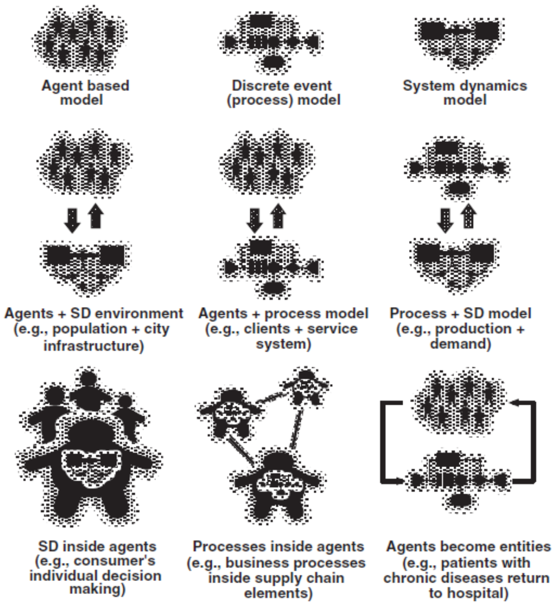


FIGURE 4.3: Hybrid model architectures according to Borshchev (2013)

Similar examples, using the same architecture, can be found in the literature. Djanatliev et al. (2012; 2013) have developed the Prospective Health Technology Assessment (ProHTA) to simulate the effects of new healthcare innovations. The proposed model consists of four modules. Three of these—population dynamics, disease dynamics and health care financing—are modelled by the SD approach. The fourth module, Health Care, is developed as an agent-based simulation where events trigger state changes in each agent's state chart, i.e. in each person's behaviour, and there are also sub-modules for each healthcare phase (e.g. prevention, pre-treatment, etc.) The main modules run independently, but at the same time they affect other parts of the model through well-defined interfaces. The continuous SD environment and the discrete agent-based model are loosely coupled, and the values are synchronized annually during the run.

Gao's diabetes study (2014) models the impact and the costs of different treatments combining three simulation modelling approaches. For the whole of the population, an aggregate model in the SD framework was chosen, while ABM was used to represent diabetic individuals, their characteristics, their possible states, and the possible treatments. Then, a third module was used to model resource availability during the patients' progression through health care processes.

Of all the architectures, agents in a SD environment seems closest to the system that we want to model: an architecture to model the policy decisions (SD) that trigger state chart changes of the archival records (agents).

4.2.7 What type of data is required?

The SD model uses ordinary differential equations (ODEs) to represent the rates of change in each stock. These ODEs are solved numerically using a discrete time step, usually called dt. ABM is defined as a computational model due to the use of computer-based algorithms. In ABM, the agents' individual behaviours are captured using thresholds, if–then rules, or nonlinear coupling (Bonabeau; 2002). As Vincenot et al. (2011, p. 212) have noted, it can be 'difficult to create a robust mental model of how an algorithm works. On the contrary, SD is able to represent processes in an intuitive and straightforward way'.

DES and ABM use probabilistic distributions, from statistical data. For example, each agent, at each time step, has a certain probability of experiencing an event or not. These probabilities are estimated from individual-level characteristicswhich are, for example, developed from survey data. A good example on how functions and updating probabilities are included in a ABM model is given by Cerdá et al. (2015).

SD uses averages, however some software permits distributions to be introduced as well. The models are built on the available data. In the case of health care models, prevalence, incidence or hazard rates are used. These data are also used for validating the model. In SD, it has been stressed that not having data on a particular variable it should not stop modellers from including the variable in the model (Sterman; 2000). If data are not available, then assumptions can be made and equations can be formulated in terms of a virtual feature (value) to reproduce the observed phenomena with the greatest accuracy possible (Schlessinger and Eddy; 2002).

Regarding archival collections, and as we will show in the next chapters, a large body of data is available at a micro-level: from survey data to usage histories automatically collected by collections management systems. This type of data could be used as input to characterise the agents in an ABM model. In contrast, there is a clear lack of data needed to determine the 'information variables', for example how the use of collections generates more use, or to quantify readers' willingness to use digital scans instead of the physical collections.

4.3 Conclusions

In this chapter, we argued that the preservation management of collections can be defined as a complex system. However, whether we are dealing with simple, complicated or complex models does not change how the model is created, since the key to successful modelling is understanding which level of detail needs to be included in order to capture the system's behaviour. It should be stressed that, in acknowledging that we are dealing with complex systems, we understand that such systems are frequently composed of parts whose purposes are not necessarily the same as the whole system's. We must hence be aware that new interventions might result in new problems.

The literature review covered three systems modelling paradigms and their applications. Although, theoretically, the same problem can be modelled using any one of the approaches, depending on the problem to be addressed and the characteristics of the system to be modelled, one particular approach might be more appropriate. In the case of DES, queues and activities are the main building blocks of the model. DES models have proven to be valuable to solve operational problems, where the randomness of activities, waiting time and available resources are key elements that define the system. For example, in our case, if we would like to improve processes within an archive (e.g. delivery of records in the reading room, or digitisation processes, from request to the delivery of the scan), then DES would be the most appropriate approach. However, as this is not the topic of this research, the application of DES will not further pursued herein.

As our problem is to explore the effects of different preservation measures on certain metrics, SD might provide useful information to analyse long-term policy decisions. Due to the aggregate view of SD models, it is possible to explore the effects of decisions by changing the parameters. At the same time, these aggregate views need a strong simplification of the system under study. In SD models, it can become rather complex to capture the heterogeneity of the individuals and their different behaviours, and also the randomness of events. Other approaches, such as ABM, have features that allow the inclusion of randomness in the system.

One important consideration is the level of detail needed in the model in order to adequately reflect the system and its stochastic elements. In the case of archival collections, individuals (archival records) undergo different stages (digitised, accessed in the reading room), which can be modelled with both SD and ABM. At the same time, archival collections are not uniformly requested, and the popularity of items can be difficult to represent in finite categories. This kind of heterogeneity may be easier to model at individual level, lending itself to ABM. Other capabilities of ABM, however, are not needed to model this system. Specifically, there is no interaction among agents, as they only change state due to exogenous factors.

The literature review revealed that the two approaches may be combined in a single model, where agents interact within a single SD model (environment). In such a model, agents' behaviours vary according to the decisions taken in the SD model.

Like SIR models, our model focusses mainly on how preservation measures affect the individuals (viz. archival records), rather than how preservation decisions are made. In this regard, the choice between a SD or ABM approach should be based on whether we need the micro- approach of ABM, or if approaching collections as aggregates is sufficient to capture the system's dynamics. In the next chapter, we will explore this question when developing the APM model's first part, which will focus on the use of the physical collections and how digitisation is affecting this use.

4.4 Model building methodology

Up to this point in the thesis, we have identified the variables to be included in the model during the development of the causal loop diagram, and the literature review presented in this chapter have lead us to the selection of SD and ABM as the two modelling approaches to be further explored in this research. From here we can now start building the Archival Preservation Management (APM) model.

The proposed APM model is composed of three sub-models. Figure 4.4 shows the order how the three sub-models have been built, as well as how they are interrelated.

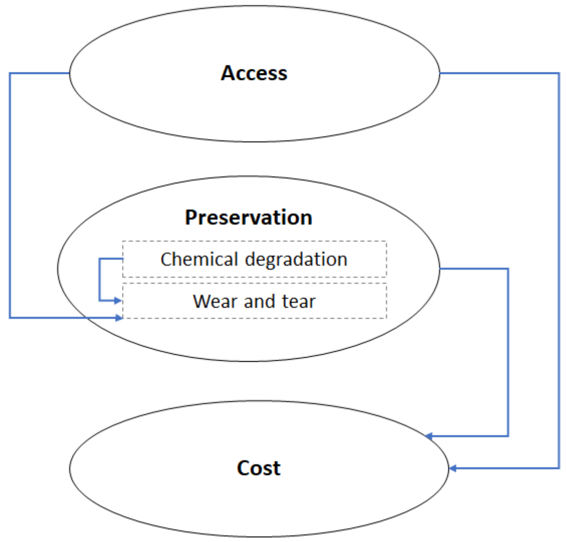


FIGURE 4.4: The proposed APM model composed of three interconnected sub-models.

1. Access: This sub-model is discussed in the first of the three chapters where the APM model is developed, as the outputs of this part of the model are required as input variables in the other two sub-models, but not vice versa. In addition, as it will be shown, this part of the model provides a good opportunity to test two different modelling approaches on the same topic.

2. Preservation: This part includes the modelling of chemical degradation of paper collections *and* the accumulation of wear and tear. Whereas the chemical degradation model works independently from the other models, the wear and tear model uses the output of the chemical degradation model as well as the access model as input variables.

3. Cost: This part of the model calculates the cost of the activities modelled in the other two sections. The inputs of this part of the model are thus related to the variables in the other two parts of the model.

The methodology applied to develop the three sub-models is as follows:

- A literature review is first conducted reviewing models and data that could be of use for each part of the model.
- Historical data and survey data is analysed for further hypothesis development, and to parameterise the model under development. Where no data was available, data collection is conducted.
- Based on the results of the literature review and the data analyses, a mathematical model is developed for each of the three parts of the model.
- Simulation experiments, mostly sensitivity analyses, are conducted to validate the sub-models, and to test parametric uncertainty.
- Simulation experiments on case studies are conducted to explore the application as well as the limitations of the APM model.

References

Ahmed, A., Greensmith, J. and Aickelin, U. (2012). Variance in system dynamics and agent based modelling using the SIR model of infectious disease, *Proceedings - 26th European Conference on Modelling and Simulation, ECMS 2012* .

Ashley-Smith, J. (2013). British Publicly Available Specification (PAS) 198, "Specification for managing environmental conditions for cultural collections", *Summit on the Museum Preservation Management*, Smithsonian Institution, pp. 29–32.

Auchincloss, A. H. and Roux, A. V. D. (2008). A new tool for epidemiology. The usefulness of dynamic-agent models in understanding place effects on health, *American Journal of Epidemiology* **168**(1): 1–8.

Barnes, D. J. and Chu, D. (2015). *Guide to simulation and modeling for biosciences*, second edn, Springer International Publishing, London.

Barton, P., Bryan, S. and Robinson, S. (2004). Modelling in the economic evaluation of health care : selecting the appropriate approach, *Journal of Health Services Research and Policy* **9**(2): 110–118.

Bonabeau, E. (2002). Agent-based modeling : Methods and techniques for simulating human systems, *PNAS* **99**(Suppl. 3): 7280–7287.

Borshchev, A. (2013). Multi-method modeling, *in* R. Pasupathy, S. H. Kim, A. Tolk, R. R. Hill and M. Kuhl (eds), *Proceedings of the 2013 Winter Simulation Conference*, pp. 4089–4100.

Borshchev, A. and Filippov, A. (2004). From system dynamics and discret event to practical agent based modeling: reasons, technques, tools, *The 22nd International Conference of the System Dynamics Society*.

Brailsford, S., Churilov, L. and Dangerfield, B. (eds) (2014). *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd.

Brailsford, S. and Hilton, N. (2001). A comparison of discrete event simulation and system dynamics for modelling healthcare systems, *in* J. Riley (ed.), *Proceedings from ORAHS 2000*, Caledonian University, Glasgow, pp. 18–39.

Brennan, A., Chick, S. E. and Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies, *Health Economics* **15**: 1295–1310.

Cassidy, R., Singh, N. S., Schiratti, P. R., Semwanga, A., Binyaruka, P., Sachingongu, N., Chama-Chiliba, C. M., Chalabi, Z., Borghi, J. and Blanchet, K. (2019). Mathematical modelling for health systems research: A systematic review of system dynamics and agent-based models, *BMC Health Services Research* **19**(1): 15–17.

Cerdá, M., Tracy, M., Keyes, K. M. and Galea, S. (2015). To Treat or to Prevent? Reducing the population burden of violence-related post-traumatic stress disorder, *Epidemiology* **26**(5): 681–689.

Chahal, K. and Eldabi, T. (2010). A multi-perspective comparison for selection between system dynamics and discrete event simulation, *International Journal of Business Information Systems* **6**(1): 4–17.

Chung, C. a. (2003). *Simulation modeling handbook. A practical approach*, CRC Press, Boca Raton.

Cimler, R., Tomaskova, H., Kuhnova, J., Dolezal, O., Pscheidl, P. and Kuca, K. (2018). Numeric, agent-based or system dynamics model? Which modeling approach is the best for vast population simulation?, *Current Alzheimer Research* **15**(8): 789–797.

Cooper, A. K., Brailsford, S. C. and Davies, R. (2007). Choice of modelling technique for evaluating health care interventions, *The Journal of the Operational Research Society* **58**(2): 168–176.

Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A. and Tako, A. A. (2020). How simulation modelling can help reduce the impact of COVID-19, *Journal of Simulation* **14**(2): 83–97.

Darabi, N. and Hosseinichimeh, N. (2020). System dynamics modeling in health and medicine: a systematic literature review, *System Dynamics Review* **36**(1): 29–73.

Davahli, M. R., Karwowski, W. and Taiar, R. (2020). A system dynamics simulation applied to healthcare: A systematic review, *International Journal of Environmental Research and Public Health* **17**(16): 1–27.

Diaz, R., Behr, J. G. and Britton, B. S. (2015). Estimating cost adjustments required to accomplish target savings in chronic disease management interventions. A simulation study, *Simulation Transactions of the Society for Modeling and Simulation International* **91**(7): 599–614.

Djanatliev, A. and German, R. (2013). Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation, *in* R. Pasupathy, S. H. Kim, A. Tolk, R. R. Hill and M. Kuhl (eds), *Proceedings of the 2013 Winter Simulation Conference*, pp. 270–281.

Djanatliev, A., German, R., Kolominsky-rabas, P. and Hofmann, B. M. (2012). Hybrid simulation with loosely coupled system dynamics and agent based models for prospective health technology assessments, *in* C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A. Uhrmacher (eds), *Proceedings of the 2012 Winter Simulation Conference*, pp. 770–781.

Forrester, J. W. (1961). *Industrial dynamics*, M.I.T. Press, Cambridge.

Forrester, J. W. (1971). Counterintuitive behavior of social systems, *Theory and Decision* **2**: 109–140.

Galea, S., Hall, C. and Kaplan, G. A. (2009). Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research, *International Journal Drug Policy* **20**(3): 209–216.

Galea, S., Riddle, M. and Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology, *International Journal of Epidemiology* **39**(2009): 97–106.

Gao, A., Osgood, N. D. and Dyck, R. F. (2014). A tripartite hybrid model architecture for investigating health and cost impacts and intervention tradeoffs for diabetic end-stage renal disease, *in* A. Tolk, S. Y. Diall, I. O. Ryzhov, L. Yilmaz, S. Buckley and J. Miller (eds), *Proceedings of the 2014 Winter Simulation Conference*, pp. 1676–1687.

Ghaffarzadegan, N., Lyneis, J. and Richardson, G. P. (2011). How small system dynamics models can help the public policy process, *System Dynamics Review* **27**(1): 22–44.

Gilbert, N. (2011). *Agent-based models*, SAGE Publications, Inc.

Glouberman, S. and Zimmerman, B. (2002). Complicated and complex system. What would successful reform of Medicare look like?, *Discussion Paper N° 8*, Commission on the Future of Health Care in Canada, Ottawa, Ontario.

Goldenfeld, N. and Kadanoff, L. P. (1999). Simple lessons from complexity, *Science* **284**(5411): 87–89.

Guerrero, C. N., Schwarz, P. and Slinger, J. H. (2016). A recent overview of the integration of system dynamics and agent-based modelling and simulation, *Proceedings of the 34th International Conference of the System Dynamics Society, July 17-21, 2016 Delft, Netherlands*, p. 1153.

Gunal, M. M. (2012). A guide for building hospital simulation models, *Health Systems* **1**: 17–25.

Heath, S. K., Brailsford, S., Buss, A. and Macal, C. M. (2011). Cross-paradigm imulation modeling. Challenges and successes, *in* S. Jain, R. Creasey, J. Himmelspach, K. White and M. Fu (eds), *Proceedings of the 2011 Winter Simulation Conference*, pp. 2788–2802.

Hirsch, G., Homer, J., Evans, E. and Zielinski, A. (2010). System dynamics model for planning cardiovascular disease interventions, *American Journal of Public Health* **100**(4): 616–622.

Hirsch, G., Homer, J., Trogdon, J., Wile, K. and Orenstein, D. (2014). Using simulation to compare 4 categories of intervention for reducing cardiovascular disease risks, *American Journal of Public Health* **104**(7): 1187–1195.

Homer, J. B. and Hirsch, G. B. (2006). System dynamics modeling for public health: Background and opportunities, *American Journal of Public Health* **96**(3): 452–458.

Homer, J., Hirsch, G., Minniti, M. and Pierson, M. (2004). Models for collaboration. How system dynamics helped a community organize cost-effective care for chronic illness, *System Dynamics Review* **20**(3): 199–222.

Homer, J., Wile, K., Yarnoff, B., Trogdon, J. G., Hirsch, G., Cooper, L., Soler, R. and Orenstein, D. (2014). Using simulation to compare established and emerging interventions to reduce cardiovascular disease risk in the United States, *Preventing Chronic Disease* **11**(140130).

Jones, A. P., Homer, J., Murphy, D. and Essien, J. (2006). Understanding diabetes population dynamics through simulation modeling and experimentation, *American Journal of Public Health* **96**(3): 488–494.

Kahn, R., Robertson, R., Smith, R. and Eddy, D. M. (2008). The impact of prevention on reducing the burden of cardiovascular disease, *Diabetes Care* **31**(8).

Lane, D. (2000). You just don't understand me. Modes of failure and success in the discourse between system dynamics and discrete event simulation, *Operational Research working papers, LSEOR 00.34*, Department of Operational Research, London School of Economics and Political Science, London.

Lättilä, L., Hilletofth, P. and Lin, B. (2010). Expert systems with applications hybrid simulation models – When , Why , How ?, *Expert Systems with Applications* **37**: 7969–7975.

Law, A. (2007). *Simulation modeling and analysis*, fourth edn, McGraw-Hill, New York.

Lipsitz, L. A. (2012). Understanding health care as a complex system: The foundation for unintended consequences, *JAMA - Journal of the American Medical Association* **308**(3): 243–244.

Liu, S., Xue, H., Li, Y., Xu, J. and Wang, Y. (2018). Investigating the diffusion of agent-based modelling and system dynamics modelling in population health and healthcare research, *Systems Research and Behavioral Science* **35**(2): 203–215.

Luke, D. A. and Stamatakis, K. A. (2012). Systems science methods in public health: Dynamics, networks, and agents, *Annual Review of Public Health* **33**: 357–376.

Macal, C. M. (2010). To Agent-based simulation from system dynamics, *in* B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan and E. Yucessan (eds), *Proceedings of the 2010 Winter Simulation Conference*, pp. 371–382.

Macal, C. M. and North, M. J. (2009). Agent-based modeling and simulation, *in* M. Rossetti, R. R. Hill, B. Johansson and R. Ingalls (eds), *Proceedings of the 2009 Winter Simulation Conference*, pp. 86–98.

Mahamoud, A., Roche, B. and Homer, J. (2013). Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto , Canada, *Social Science & Medicine* **93**: 247–255.

Marchal, B., Van Belle, S., De Brouwere, V. and Witter, S. (2013). Studying complex interventions. Reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco, *BMC Health Services Research* **13**(1).

Marshall, B. D. L. and Galea, S. (2015). Practice of epidemiology formalizing the role of agent-based modeling in causal inference and epidemiology, *American Journal of Epidemiology* **181**(2): 92–99.

Meadows, D. H. (1999). *Leverage points. Places to intervene in a system*, The Sustainability Institute, Hartland.

Meadows, D. H. (2009). *Thinking in systems*, Earthscan, London.

Meadows, D. H. and Robinson, J. M. (1985). *The electronic oracle. Computer models and social decisions*, John Wiley & Sons, Inc., Chichester.

MIDAS. Models of infectious disease agent study (n.d.).
URL: *midasnetwork.us*

Milstein, B., Homer, J., Consulting, H. and Hirsch, G. (2009). The " HealthBound " policy simulation game. An adventure in US health reform, *International System Dynamics Conference, July 26-30, 2009, Albuquerque, New Mexico*.

Milstein, B., Homer, J. and Hirsch, G. (2010). Analyzing national health reform strategies with a dynamic simulation model, *American Journal of Public Health* **100**(5): 811–820.

Morgan, J. S., Howick, S. and Belton, V. (2017). A toolkit of designs for mixing Discrete Event Simulation and System Dynamics, *European Journal of Operational Research* **257**(3): 907–918.

Ness, R. B., Koopman, J. S. and Roberts, M. S. (2007). Causal system modeling in chronic disease epidemiology. A proposal, *Annals of Epidemiology* **17**(7): 564–568.

Nianogo, R. A. and Arah, O. A. (2015). Agent-based modeling of noncommunicable diseases: A systematic review, *American Journal of Public Health* **105**(3): e20–e31.

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V. and Alvin, K. F. (2002). Error and uncertainty in modeling and simulation, *Reliability Engineering and System Safety* **75**(3): 333–357.

Pidd, M. (2014). The ways forward : A personal view of system dynamics and discrete-event simulation, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, first edit edn, John Wiley & Sons, Ltd., Chichester, chapter 15, pp. 318–336.

Plsek, P. E. and Greenhalgh, T. (2001). The challenge of complexity in health care, *British Medical Journal* **323**(7313): 625–628.

Poli, R. (2013). A note on the difference between complicated and complex social systems, *Cadmus* **2**(1): 142–147.

Preiser, R. (2019). Identifying general trends and patterns in complex systems research: An overview of theoretical and practical implications, *Systems Research and Behavioral Science* **36**: 706–714.

Preston White, K. and Ingalls, R. (2016). The basics of simulation, *in* T. Roeder, P. Frazier, E. Szechtman, E. Zhou, T. Huschka and S. E. Chick (eds), *Proceedings of the 2016 Winter Simulation Conference*, pp. 38–52.

Rahmandad, H. and Sterman, J. D. (2008). Heterogeneity and network structure in the dynamics of diffusion. Comparing agent-based and differential equation models, *Management Science* **54**(5): 998–1014.

Robinson, S. (2014). Discrete-event simulation : A primer, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., Chichester, chapter 2, pp. 10–25.

Rotaru, K., Churilov, L. and Flitman, A. (2014). Philosophical positioning of discrete-event simulation and system dynamics as management science tools for process systems. A critical realist perspective, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., Chichester, chapter 5, pp. 76–104.

Rouse, W. B. (2015). *Modeling and visualiztion of complex systems and enterprises. Explorations of physical, human, economic, and social phenomena*, John Wiley & Sons, Inc., New Jersey.

Salleh, S., Thokala, P., Brennan, A., Hughes, R. and Booth, A. (2017). Simulation modelling in healthcare: An umbrella review of systematic literature reviews, *PharmacoEconomics* **35**(9): 937–949.

Schieritz, N. and Milling, P. (2003). Modeling the forest or modeling the trees: A Comparison of system dynamics and agent-based simulation, *in* R. Eberlein, V. Diker, R. Langer and J. Rowe (eds), *Proceedings of the 21st international conference of the System Dynamics Society : July 20 - 24, 2003, New York City, USA*, pp. 1–15.
URL: *http://www.systemdynamics.org/conferences/2003/proceed/PAPERS/140.pdf*

Schlessinger, L. and Eddy, D. M. (2002). Archimedes. A new model for simulating health care systems — the mathematical formulation, *Journal of Biomedical Informatics* **35**: 37–50.

Scholl, H. J. J. (2001). Agent-based and system dynamics modeling. A call for cross study and joint research, *Proceedings of the 34th Hawaii International Conference on System Sciences*.

Senge, P. (1990). *The fifth discipline. The art and practice of learning organization*, Doubleday/Currency, New York.

Shannon, R. E. (1998). Introduction to the art and science of simulation, *Proceedings of the 1998 Winter Simulation Conference*, pp. 7–14.
URL: *http://cecs.wright.edu/~fciarall/ISE195/Readings/ShannonSimulationART.pdf*

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*, Mcgraw-Hill Education, Boston.

Sterman, J. D. (2006). Learning from evidence in a complex world, *American Journal of Public Health* **96**(3): 505–514.

Swinerd, C. and McNaught, K. R. (2012). Design classes for hybrid simulations involving agent-based and system dynamics models, *Simulation modelling practice and theory* **25**: 118–133.

Taylor, J. (2005). An integrated approach to risk assessments and condition surveys, *Journal of the American Institute for Conservation* **44**(2): 127–141.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K. and Mazzoleni, S. (2011). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology, *Ecological Modelling* **222**(1): 210–218.

Weimer, C. W., Miller, J. O. and Hill, R. R. (2016). Agent-based modeling: an introduction and primer, *in* T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka and S. E. Chick (eds), *Proceedings of the 2016 Winter Simulation Conference*, pp. 65–79.

Wierzbicki, A. P. (2007). Modelling as a way of organising knowledge, *European Journal of Operational Research* **176**(1): 610–635.

Chapter 5

Access

> *'Anecdotal evidence on what is "obvious" to practitioners, however, cannot be relied upon to convince others, and so the archive community needs hard facts and sound statistical evidence to paint a proper picture of the domain and to illustrate trends.'*
>
> Pickford 2002, p. 33

Institutions have adopted digitisation of analogue collections to provide access to archival and library collections (Campagnolo; 2020). After almost four decades of digitisation projects, analogue and digital access coexist in most institutions. In this chapter, we explore how digitisation is affecting the use of the physical collections. First, we provide some background on how digitisation has become part of the strategy of providing access to the collections in institutions, particularly the drivers and selection criteria for digitisation. Further, we analyse 20 years of usage data from the reading room at the Amsterdam City Archives (SAA) to obtain evidence on how digitisation is affecting the use of collections in the reading room, understood as the number of access requests. We also discuss the results of a survey conducted among the readers to collect information on their preferences for accessing archival records in the reading room or using scan-on-demand (SoD), a service that allows readers to request the digitisation of records that are not digitally available yet.

In the second part of this chapter, the dynamic hypothesis developed from the data analysis of the SAA data set and the survey results are taken together as the basis to convert the variables identified during the development of the causal loop diagram in Chapter 3 into a simulation model. As discussed in Chapter 4, the dynamics of archival collections can be reproduced with different modelling approaches. For the access model, we test two modelling approaches: systems dynamics (SD) and agent-based modelling (ABM).[1]

5.1 Digitisation strategies: An introduction

The digitisation process in libraries and archives started in the 1970s with the development of electronic catalogues. In the 1980s, the early digitisation projects were an attempt to test whether digitisation could be a convenient alternative to microfilms, a common approach at that time to create surrogates of frequently requested records. Terras remarks that, even in this period, prior to the World Wide Web (WWW), 'solutions were seen as means to make their holdings more accessible, searchable, and available, for both government bodies and the general public' (Terras; 2011, p. 6). At the beginning of the 1990s, small-scale projects were mostly related to treasures, but over the decade, there was a shift to large-scale projects, stimulated by the introduction of the Internet (Terras; 2011). The attention to online access and finding aids, as well as the preservation of digital information, became more prominent. In the 2000s, the awareness of the cost of digitisation arose. More institutions started developing programs that offer strategic benefits to institutions and users instead of carrying out opportunistic and short term projects (Hughes; 2003, p. 8). However, at the same time, initiatives from large commercial firms, such as Google, began to emerge. For the first time, the term 'mass digitisation' was introduced to describe the 'conversion of whole libraries without making a selection of individual materials' (Coyle; 2006, p. 641). Some authors make a distinction between mass digitisation and large-scale digitisation. The latter is described as 'more discriminating that mass-digitisation projects. Although they create a lot of scanned pages, they are concerned about the creation of collections and about reproducing complete sets of documents' (Coyle; 2006, p. 642). The term mass digitisation has also been used in archives to denote these structured projects (Jones; 2008; Holtman; 2017). There is no doubt that the large-scale digitisation of collections will continue in the next decades, but a counter movement, 'slow digitisation', has already been advocated for (Prescott and Hughes; 2018).

[1]Parts of this chapter (section 5.1, 5.2, 5.4, 5.5, including figures 5.1, 5.2, 5.21, 5.23, 5.29) have been published in Cristina Duran-Casablancas, Marc Holtman, Matija Strlič, & Josep Grau-Bové, The end of the reading room? Simulating the impact of digitisation on the physical access of archival collections (2022), Journal of Simulation, https://doi.org/10.1080/17477778.2022.2128911

5.1.1 Drivers for digitisation

Even before the introduction of the WWW, the potential benefits of digitisation had been recognised, and through the decades, its benefits have become more apparent in a broad range of fields (Hughes; 2003; Rikowski; 2011; Tanner and Deegan; 2011). There are two main drivers to digitisation: access and preservation. However, enhanced access to the collections is frequently seen as the primary purpose of digitisation (Astle and Muir; 2002). Enhanced access is the result of creating digital content and other aspects (e.g. the content needs to be indexed and preferably freely available on the Internet). The Guidelines for Digitisation Projects written by the International Federation of Library Associations and Institutions (IFLA) and the International Council on Archives (ICA) pointed out that '[t]he key point is to evaluate the contribution that increased access could make to a defined user community' (IFLA; 2002, p. 11). We illustrate this point with an example. In 1970, the Public Record Office (PRO) conducted its first market research survey of readers. Of the total respondents, 83% of researchers were academics, students or professional researchers who were conducting historical research leading to a publication, and 25% had leisure interests such as genealogy (Shepherd; 2009, p. 86). However, genealogical researchers had become the largest group in the subsequent years, accounting for 50–90% of all users in North America and Europe (Tucker; 2007). To serve this group, numerous projects have been initiated: creation of websites with directories and census materials and conservation projects for microfilm or digitised major genealogical sources (Lindsay; 2003).

Although targeting only the records that are of interest to the readers seems like a well-founded strategy, an increasing pressure to digitise the entire archival holdings comes from user expectations that have changed over the years (Bantin and Agne; 2010; Oliver; 2011). The pressure to digitise entire collections has been further increased by government bodies that are requesting the release of more information online (Dalgleish; 2011), making digitisation and digital dissemination topics of conversation in cultural policies (European Commission; n.d.; Valtysson; 2017). It has also been argued (Jones; 2008) that only through digitising complete collections can the desired outcomes of providing digital access to collections be achieved (e.g. enabling the serendipitous discovery of materials (Thylstrup Bonde; 2018, p. 7)). Hence, other digitisation strategies, such as the digitisation of the collections by acquisition, have also been proposed (Miller; 2013, p. 529).

The digitisation of entire collections together with new technology facilitates new forms of access and use (IFLA; 2002). For example, the traditional cataloguing in archives is now partly done with the collaboration of volunteers in crowd-sourced projects (Andro; 2018). Artificial intelligence is also now used for text recognition. This new use of the collections, where the entire collection is searchable, indicates that the entire collection needs to be digitised.

In addition to access, preservation is seen as a driver or a benefit because digitisation can provide access to materials that, due to value or condition, would not otherwise be accessible to readers. Digitisation projects can also be initiated to reduce handling of high-demand materials and the subsequent risk of wear and tear. This preservation strategy of making surrogates of originals started in the form of microfiches or microfilms. Although the benefits of reducing handling are evident, it has been argued that digitisation is not the equivalent to preservation, as wear and tear is only one aspect, and by not taking further preservation measures, the risk of obsolescence increases (Oliver; 2011; Maroso; 2003; Riley-Reid; 2015). Although this involves a risk, digitisation also provides an opportunity to treat the collection and stabilise the condition of the objects before digitisation (O'Brien Miller and Tedone; 2011; Lindsay; 2003). This opportunity was also observed during the development of the causal loop diagram presented in Chapter 3.

In addition to access and preservation, Hughes has remarked that digitisation can (indirectly) lead to other benefits, such as institutional and strategic benefits, or support research and education purposes (Hughes; 2003) .

5.1.2 Costs of digitisation

While digitisation has multiple uses/advantages, large-scale digitisation of collections comes at a cost. In 2010, the cost of digitisation was calculated for heritage collections, including libraries and national archives. The total number of pages to be digitised in European libraries was estimated to be between 1.47 and 2.36 billion, which costs between €4.79 and 11.76 billion (Poole; 2010). According to the same report, national archives collections contain 17.27 billion pages eligible for digitisation, which would cost €41.87 billion.

These are the costs of the first step of digitisation, namely creating digital scans of physical records. However, in addition to scanning, digitisation comes with a long-term obligation to preserve and provide access to the digital content. However, this is not an easy task. For example, the United States National Archives and Records Administration of de National Archives (NARA) report states that 'NARA has nearly 235 million pages of records digitised, but only about 15 percent of those digital records are currently available to the public through the National Archives Catalogue' (NARA; 2018, p. 6). Keeping digital information catalogued, accessible and usable requires constant effort and budget. In Chapter 7, we will demonstrate that each stage in the life cycle of digitisation needs to be taken into account when calculating the cost of providing digital access to the collections: creation or purchase, acquisition, ingest, metadata creation, bit-stream preservation, content preservation and access (Ayris et al.; 2010). According to the Collections Trust report (Poole; 2010, p. 3), most estimates put the costs of providing digital access to collections for a period of 10 years at 50-100% of the initial costs of creating digital assets.

Considering the workload and financial resources involved, complete digitisation of heritage collections may not be feasible and likely unnecessary. According to the Collections Trust report (Poole; 2010, p. 3), national archives in Europa account for 26.98 billion pages of archival records, of which approximately 17.27 billion are eligible/appropriate for digitisation (Poole; 2010). In 2017, the Europeana report (Nauta et al.; 2017, p. 28) inspected the percentage of European heritage that has already been digitised at European memory institutions. They estimated that approximately 10% of the collections in archives and 17% in libraries are digitally available. Interestingly, the report also mentions that 40% and 33% of the archival and library collections, respectively, '[do not] need to be digitally reproduced' (Nauta et al.; 2017, p. 28). However, the criteria on which the latter was defined were not included.

Given the cost of digitisation and the vast amount of collections in memory institutions that are potentially eligible to be digitised, the sustainability of digitisation programs can be guaranteed only if these projects are planned based on well-defined criteria for the holdings to be digitised.

5.1.3 Criteria of selection

According to the results of the Enumarate project in 2017, 44% of the archives and 37% of the libraries had a digital strategy in place (Nauta et al.; 2017, p. 17). Up to 90% of these strategies concern the digitisation of analogue collections. Long-term digital preservation is also mentioned in 74% of the strategies. However, these strategies usually include basic principles without elaborating on the holdings that need to be prioritised for digitisation (Bantin and Agne; 2010, p. 245). Ooghe also noted that digitisation initiatives often respond to ad hoc decisions and available funding (Ooghe and Moreels; 2009).

To support the selection of holdings to be digitised, guidelines and evaluative tools have been developed. For example, the Harvard Model (Brancolini; 2000) includes a decision-making matrix similar to the one used in the acquisition process. General guidelines, such as the IFLA Guidelines, have also been developed for digitisation projects. According to the IFLA Guidelines (IFLA; 2002, p. 13), digitisation projects should start with a clear selection policy that specifies what material will be included and for what purpose. The selection criteria should take the following into account: (1) the intellectual value of the items based on content, (2) the level of demand, and (3) the physical condition of the material to be digitised and the level of cataloguing and descriptive data.

A survey including over 150 libraries and archives showed that the main selection criteria were increased access, historical value and academic importance, but '[these figures] appear to indicate that little attention is paid to user demand, although value and importance may be determined by demand to some extent' (Astle and Muir; 2002, p. 68). High demand seems like a logical selection criterion, as it directly responds to the interest of the readers. However, as pointed out in the previous section, there are different drivers to initiating digitisation programs. For example, thematic digitisation projects or format-based projects (e.g. photographs, posters) are chosen due to practicality (Lindsay; 2003). Institutions are constantly in search of the right balance between user needs and institutional priorities (Mills; 2015).

5.1.4 User studies

The need for a better understanding of the use of the collections and users has been continuously stressed in the last forty years (Duff et al.; 2010; Parkinson; 2005). The Joint Information Systems Committee (JISC) report *Digitisation in UK* states that '[f]uture developments in digitisation need to respond more directly to user demand rather than supply, yet researchers' and other user needs (including searching behaviours) are still not fully understood' (Parkinson; 2005, p. 5). In 2011, Dobreva (2011) remarked that users are not sufficiently involved in the digitisation process. Rhee (2015) also noticed that fewer user studies have been conducted on archives than on libraries. Rhee suggests two possible reasons. First, while libraries have a stronger focus on serving user needs, the main focus of archives has mainly been preservation, due to the uniqueness of the material. Second, there is a preference for using limited resources for processing and description than conducting user studies. Often, user studies in archival institutions are conducted as annual surveys to monitor user satisfaction (see for example ARA n.d.).

An overview of sources and tools designed to collect user information and the use of the collections is given in Dobreva et al (2011), Kelly (2014), Rhee (2015), McAvena (2017). The main research topics of archival user studies can be categorised as follows: information needs, information seeking and information use. Regarding user studies methods related to digitisation, Dobreva distinguishes two main types: (1) methods based on direct user involvement such as questionnaires, interviews or experiments to study user behaviour aspects; (2) methods based on indirect observation such as deep log analysis that examines the traces of user activities in the use of web sources (duration of the visit, search terms used, etc.) (Dobreva et al.; 2011).

Statistics has been traditionally used to collect information on users (number of readers, age, gender, ethnicity and education level) and the use of the collections (requests in the reading room and requests of reproduction) (Pickford; 2002). The analysis of statistics is now shifting to web metrics (Kelly; 2014). Web data can be used to analyse usage and search patterns. Identification of the items viewed the most or the search findings in catalogues could be used to justify the prioritisation of the digitisation of the collections. However, most studies have focused on the number of readers, the pages readers enter and time spent on pages or elements of the page. Google Analytics is used in archives, whereas COUNTER online metrics, provided by Vendor, is used to analyse usage data of ebooks, (COUNTER; 2021). The potential of web metrics can be found in the 2009 report by the JISC (Meyer et al.; 2009). The report presents the toolkit for the Impact of Digital Scholarly Resources, which provides open access to a set of approaches and tools available to measure and potentially improve the usage and impact of current and future digitisation projects . The report also provides an overview of quantitative and qualitative methods, supported by the analysis of five case studies to illustrate the type of information that can be extracted from each method. Among the quantitative measures, the following methods are discussed: web metrics (link analysis), web analytics (Google Analytics), log files, bibliometric analysis and content analysis. The report stresses the importance of log files to detailed information on user behaviour and access statistics, such as 'the most popular pages on the website, the most downloaded files, the most downloaded images, the most displayed images, the most requested directories, and the top entry pages' (Meyer et al.; 2009, p. 58).

5.1.5 Data driven decision making

The sources of data discussed in the previous sections, from statistics on the use of analogue and digital collections to satisfaction surveys, can be potentially used to drive decision making. A good example of data supporting decision-making is the use of circulation statistics to support management practices related to collection development, acquisition, storage, preservation, digitisation or staffing needs (Morrisey; 2010; Tyler et al.; 2013; Hughes; 2016). It is well known that the level of demand within collections can vary greatly. For example, statistics suggest that a very limited number of e-journals are intensively used, whereas the rest of the journals are occasionally used (Brown; 2003, p. 146). The justification for developing strategies based on the most frequently demanded records is clear because a high access indicates that the records are of interest to the readers. By digitising highly demanded records, the selection has already been done by the readers (Erway and Schaffner; 2017). This is the case for the service scan-on-demand (SoD), for which the readers select the material to be scanned (Ling and McLean; 2004; Schaffner et al.; 2011; Kemp; 2016). Accordingly, only records that are of interest to the readers are digitised. By letting readers select the records, single records are digitised, and an entire collection is rarely digitised. However, from the perspective of the completeness of digital content, this can be seen as a limitation. In addition, a tailored service, such as the SoD, calls for more resources for selection and retrieval from the repositories than when linear metres of archives are selected for digitisation.

Web metrics can be used similar to the circulation statistics to identify collections of interest for the readers. However, website statistics are usually used to demonstrate a higher number of readers than the traditional access in the reading room and justify the investment of digitisation projects (Lejeune; 2007). Although web metrics are difficult to compare with circulation data of physical collections (Yuan et al.; 2018), institutions use statistics to compare the two services. In 2010, based on the statistical data, the UK National Archives provided 221 documents online for every document provided in reading rooms (Macpherson; 2010). In National Archives Australia, 22,290 visits in the reading room were counted against more than 5 million visits to the online RecordSearch database, which is the method used for accessing digitised images of records (Macpherson; 2010).

In view of these numbers, institutions are faced with the dilemma of whether they should prioritise online access instead of the traditional service of the reading room (McCausland; 2011; Jeurgens; 2013). It has been argued that the majority of the (archival) readers are mostly interested in the text and, therefore, digital access is preferred (Macpherson; 2010). At the same time, readers still require access to physical records in cases where other values, such as the material aspect of the objects, matter (Jeurgens; 2013).

Likewise, it has been argued that after digitisation, digital copies would be preferred by the majority of the users, as digital copies can provide the information they are seeking, mostly related to the content and less to the material aspect of the objects (Hirtle; 2002, p. 46). However, the opposite has also been argued: digitisation will eventually result in an increased interest in access to physical items, as digitisation promotes the use of originals among readers who value the experience of handling originals (Oliver; 2011; Astle and Muir; 2002). However, during discussions with practitioners when developing the causal loop diagram (Chapter 3), it turned out that these arguments are frequently based on anecdotal evidence. Meanwhile, a study conducted at the National Archives in the UK showed the importance of procedural improvements post-digitisation to ensure blocking access to originals post-digitisation, which will result in a reduction of handling after digitisation (VanSnick and Ntanos; 2018).

It is worth mentioning that in addition to digitisation, other factors might affect the number of access requests. In the causal loop developed in Chapter 3, professionals noted that it is reasonable that the use of the collections will result in more use and that an important condition that needs to be met is the level of metadata. It is expected that creating metadata will increase the use of the collections (Yuan et al.; 2018). However, it is also worth noting the point raised by Lejeune that detailed information in the catalogues can also help readers to check whether the records meet their requirements before accessing the records, suggesting that records are sometimes unnecessarily requested for access (Lejeune; 2007) .

Although different factors determine how the number of requests in the reading room might evolve over time, it has become clear that digitisation has an impact on the number of readers visiting the reading room for certain types of collections. For example, referring to academic library collections, Martell noted in 2008 that '[t]here is no end in sight to the declines in circulation and reference that many libraries are experiencing.' (Martell; 2008, p. 406). Following the example of genealogical studies presented earlier in this chapter, for this group, the impact of providing digital access is also becoming visible. Surveys conducted in the UK archives show a downward trend from 49% in 2014 to 41% in 2018 (ARA; 2019) in family history research, likely as an outcome of more content becoming available online.

From the perspective of practitioners, it might seem 'obvious' that digitisation will eventually change the way we access collections as the share of the collection becoming digitally available increases. However, anecdotal evidence is not sufficient to understand the importance of this rate of change and its relationship with management strategies. As Chapman et al. have pointed out, '(w)hile special collections and archives managers have at times recognised the importance of using data to drive decision making, translating this objective into reality and integrating data analysis into day-to-day operations has proven to be a significant challenge' (Chapman and Yakel; 2012, p. 129). Chapman et al. reviewed case studies to asses how institutions incorporate (operational) data analysis into decision-making and management practices. The authors concluded that 'data collection and analysis—while perhaps understood to have value—are often prioritized behind many other essential activities. They also indicate that anecdotal and observational evidence has proven easier for librarians and archivists to understand and use than methods involving quantitative data analysis' (Chapman and Yakel; 2012, p. 137).

This literature review has shown that although data collection is well-embedded in archival institutions, further data analysis is needed to move from anecdotal evidence to sound (statistical) evidence. In the next section, we begin by analysing the usage data from actual collections to obtain evidence for the impact of digitisation on the number of access requests in the reading room. Once records are digitally available, the option to order the original is usually removed (VanSnick and Ntanos; 2018). Consequently, it can be expected that digitisation will eventually result in a decrease in the number of access requests in the reading room. The collections of the SAA are taken as a case study, as this archive has been collecting usage data for almost 20 years and has undertaken numerous digitisation projects over the last 15 years.

5.2 Analysis of actual collection data

5.2.1 Introduction

Since the introduction of computers to archives, reliable data on the number of times an object is accessed in the reading room have become available. Information on the usage of the analogue records is automatically generated and stored by collection management systems. During this study, we found that archives have usage data from the last 5—10 years. It is not unusual to lose information collected using an old system when an institution upgrades to a newer system.

The case of the SAA is an exception with over 20 years of usage data. The collection management system (ABS-Archeion) was introduced in 1998 and remained in use until 2021. The SAA is responsible for 50 kilometres of the city's documentary heritage, from parchment documents to digital files, dating from 13^{th} century to the present. To increase accessibility and preservation of the collections, the SAA has been digitising its collections for several years and now digitises around 1.5 million pages of archival material per year. Given the availability of usage data on the analogue collections for more than 20 years, as well the digitisation activity in the last 15 years, the data set provided by the SAA was used in this study to investigate the impact of digitisation on the access of the collections in the reading room.

Digitisation is an important aspect of the SAA strategy (Kemp; 2016). Following a trial period of two years, in 2006, the SAA started the digitisation of the archival collections, according to two programmes. One involves the SoD service, wherein readers request the digitisation of archival records that are not digitally available yet. The other programme consists of digitisation of blocks of (popular) archives. The SAA chooses subsets of archives that will be digitised as a block. The primary aim of digitisation is to enhance access to the collections. SoD offers the opportunity to digitise records that are of interest to the readers. The selection of subsets of archives to be digitised in totality is based on the level of reader demand.

The aim of this section is to elucidate usage data to obtain evidence for the following hypothesis: the availability of digital collections will eventually reduce the number of requests in the reading room at a rate that depends on how often records have been requested in the reading room before being digitised. This analysis will then be used to develop a mathematical model that simulates the impact of digitisation on the use of the analogue collections.

5.2.2 Methodology

The following data on the (requested) records were obtained from the collection management system (ABS-Archeion):

- Inventory number of the record
- Archive number to which the record belongs
- Reader's identification number who accessed the record
- Date of record access
- Date of record digitisation (if applicable)

The analysed data were collected from 1999 to 2018. Requests regarding collections or library collections were excluded because, unlike linear metres of archives, collections are counted as single items. Records accessed by staff were also excluded to obtain the requests of the readers in the reading room alone.

Data were analysed at three levels:

- Records identified by a unique combination of archive number and inventory number
- Requests: the number of times that record is accessed
- Readers: readers in the reading room

Microsoft Access Database was employed to make queries by filtering the data, performing calculations and combining data from various tables.

5.2.3 Results and discussion

The usage data of the SAA shows that since 2006, there has been a steady downward trend in the number of access requests in the reading room: a decrease of 36%, from 24,782 requests in 2006 to 15,769 requests in 2018 (Fig. 5.1). In the reported period, there were three outliers: 2007, when the archive moved to a new location and remained closed for several months, and 2011 and 2012, when there were several collaboration projects with educational programmes in archival sciences, resulting in an increased use of collections by students.

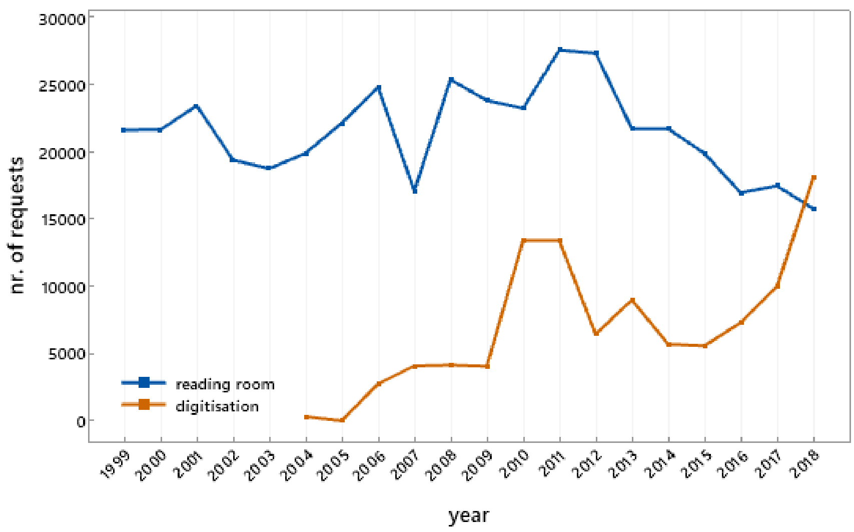


FIGURE 5.1: Number of access requests per year in the reading room of the SAA between 1999 and 2018 and number of archival records digitised per year since 2004.

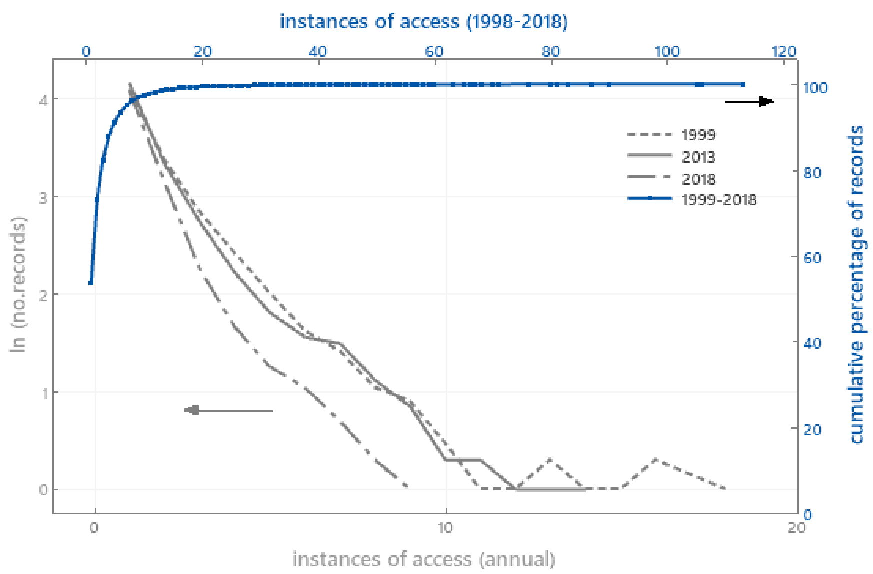


FIGURE 5.2: Frequency distribution of records that had been requested in the reading room according to the number of access requests in 1999, 2013 and 2018, and cumulative percentage of the number of records that had been requested in the reading room between 1 and 113 times for 20 years at the SAA.

In 2006, after several pilot projects, the SAA started digitising the collections within two projects: SoD and the digitisation of blocks of (popular) archives. The total number of archival documents digitised per year is reported in Figure 5.1. The peaks in 2010 and 2011 were due to large digitisation projects when two archives were fully digitised. The sharp increase since 2017 is also due to the notable increase in digitisation projects of archival blocks. Regarding the SoD production, ca. 4,000 records were digitised each year from 2008 to 2015, and this figure steadily increased to 6,000 records in 2018. By 2018, 104,169 archival records had been digitised, of which 42% (44,045 records) were part of the SoD programme.

5.2.3.1 Instances of use in the reading room

From 1999 to 2018, 167,690 archival records were requested 429,108 times. The accumulation of access requests of the archival documents follows a Pareto distribution, which means that more than a half of the archival records had been accessed once, 99% have been accessed less than 15 times in 20 years (Fig. 5.2), and the distribution showed a long tail (records accessed between 15 and 113 times in 20 years). A Pareto distribution is also obtained when the frequency of requests for archival records is plotted per year.

Despite the observed fluctuations in the number of requests through the period of 20 years (Fig. 5.1), the number of access requests per record in one year seemed to remain stable through this period. Approximately 14,000 records were requested once in a year, and approximately 2,000 records were requested twice in a year (Fig. 5.3). Records requested 3—5 times were in the hundreds, and those requested 6—8 were in the tens. Only a few records had been requested 9—22 times in the same year. However, Figure 5.4 shows that the proportion of records requested once has been slightly increasing since 2013, whereas the proportion of records requested more than once has been decreasing (i.e. the difference between the total number of requests and the number of records requested once became smaller). This can also be seen in Figure 5.2, where the tail of the distribution is notably shorter in 2018 compared to the preceding years.

These data suggest that the decrease in the total number of requests since 2013 is mostly due to the slight decline in records that used to be frequently requested in the same year.

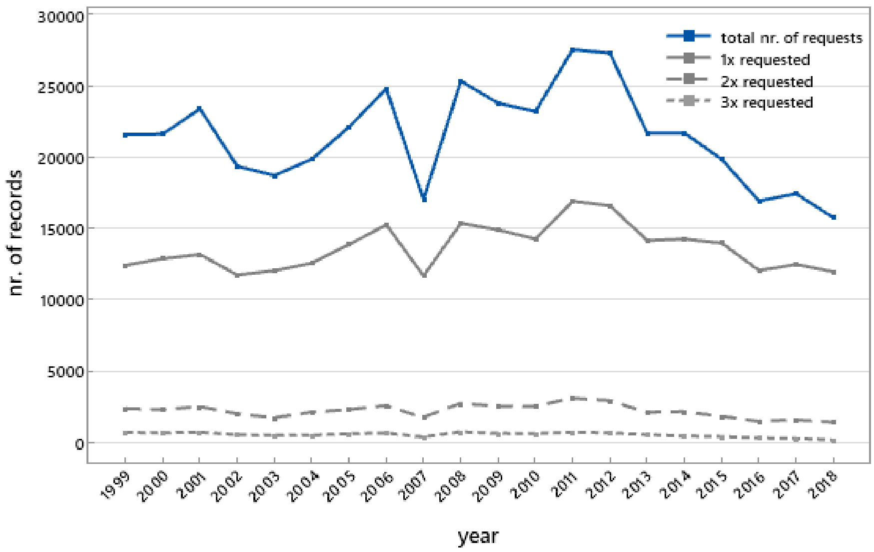


FIGURE 5.3: Number of records requested 1–3 times per year. The number of annual requests made between 1999 and 2018 is shown as a reference at the SAA.

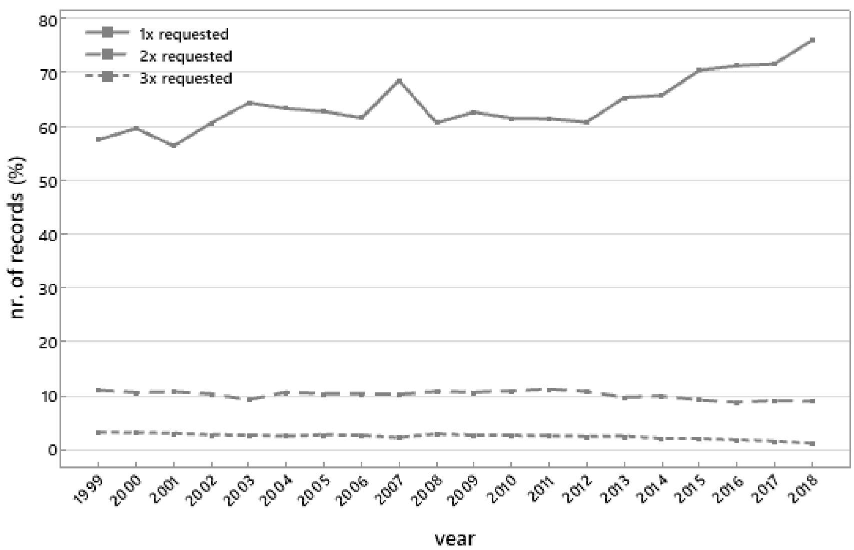


FIGURE 5.4: Number of records requested in the reading room between 1 and 3 times per year, indicated as percentage of the total number of annual requests at the SAA.

5.2.3.2 Impact of digitisation on the use of physical collections

Because more than 20 years of data are available, we can also analyse how many records have been requested only once during this period Approximately half of the records that had been requested once in a year were not further requested in the 20-year-period . From 2014, the proportion rose to two-thirds of the records. If the records are classified according to the number of times they have been requested in the last 20 years (Fig. 5.5), then a change in the pattern can be observed from 2012, confirming the results discussed in the previous section. A decrease can be observed for records that have been requested more than thrice (Fig. 5.5). This change in the pattern could be a result of the digitisation programmes where popular records have been digitised, and therefore, their use in the reading room has decreased. At the same time, digitisation seems to have notably increased the number of records that are only accessed once.

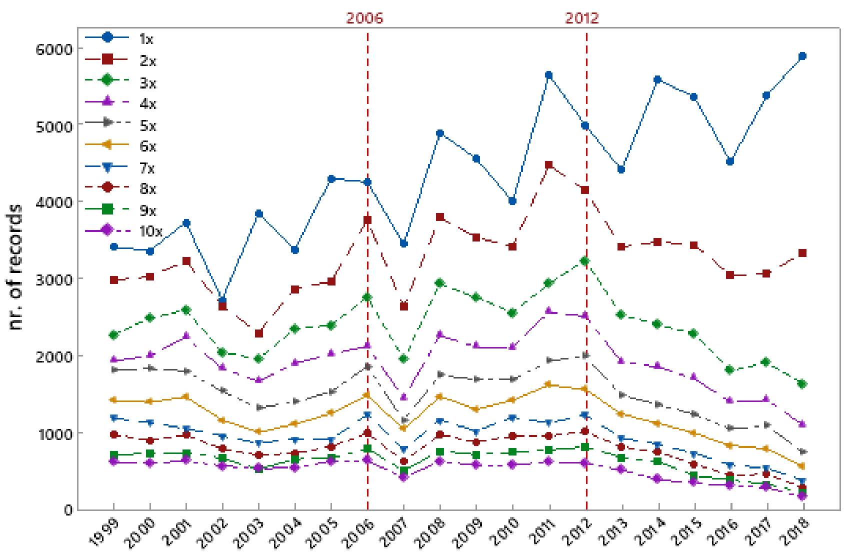


FIGURE 5.5: Number of records classified according to the total number of access requests during the 20-year-period at the SAA.

To summarise, although the frequency distribution of access requests remained moderately stable throughout the years (Fig. 5.2), two changes in the pattern of the distribution can be observed: (1) the number of archival records that had only been accessed once increased over the years, (2) whereas access to frequently requested documents has decreased since 2013. This change in access pattern may be explained by two possible reasons: (1) every year, new archival documents are requested for the first time, but the earlier it occurs, the higher the chances of another request, and (2) at the same time, popular records are being digitised; therefore, their use in the reading room is decreasing.

5.2.3.3 Digitisation match

In this section we introduce the term 'digitisation match' as an indicator of how digitisation is succeeding in making the most popular records in the reading room digitally available. In this case popularity is defined by the number of access requests accumulated by a record. From here, we can calculate the percentage of the accumulated requests that accounts for records that have been digitised. Those records becoming digitally available with a high number of accumulated access requests will make a greater contribution to the digitisation match than those records only requested once before being digitised.

The importance of considering the frequency of access requests of archival documents is further stressed in Figure 5.6, in which the variable 'digitisation match' is introduced. When analysing usage data in archival collections, we distinguish between archival records and access requests of the archival records. Figure 5.6 shows the records that have been accessed in the reading room and later digitised. So far, this is the case for 34% of the digitised records. As discussed in the next section, this relatively low percentage is partly due to the digitisation programme, in which archive blocks are digitised. However, this percentage does not tell whether those records becoming digitally available are those that are also popular among the readers. To investigate this, we need to calculate the 'digitisation match', in order words, the accumulated access requests of digitised records out of the accumulated number of access requests for all records (digitised or not). According to Figure 5.6, for example, in 2010 and 2011, the same number of records were digitised, but the match between digitised records and their popularity in the reading room was notably higher in 2010 than in 2011. In 2014, a similar number of records was digitised as in 2013 and 2015, but with a higher match.

If the digitisation match is calculated as the cumulative percentage, then the slopes of the decrease in requests since 2006 and the 'digitisation match' are rather similar ($R^2 = .941$, $p < .001$) (Fig. 5.7), supporting the hypothesis that the observed decrease in the number of access requests since 2006 is correlated with digitisation of the collections. This hypothesis will be further tested in the simulation models.

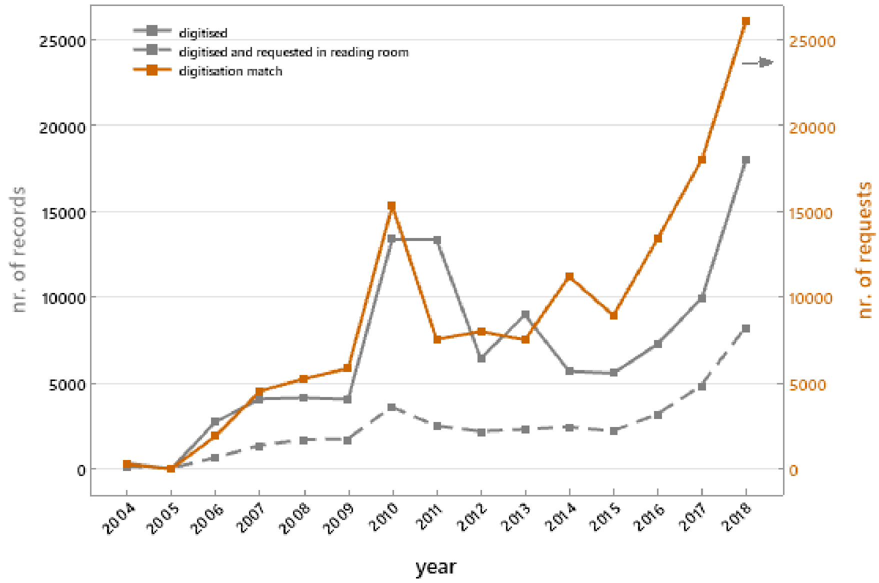


FIGURE 5.6: Number of records digitised per year (grey line); number of records requested in the reading room that have been digitised (grey dashed line); and sum of requests history in this group (accessed and digitised), referred to as 'digitisation match' (orange line) at the SAA.

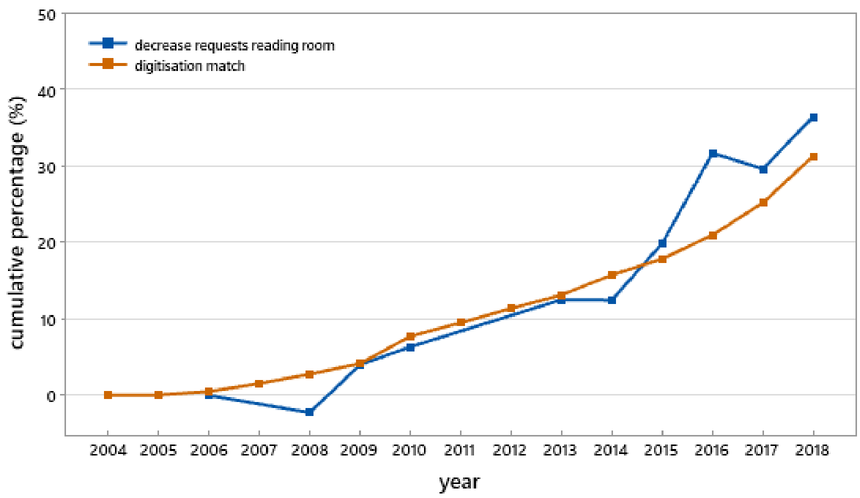


FIGURE 5.7: 'Digitisation match', sum of requests history of requested *and* digitised records, given as a cumulative percentage (orange line) and percentage decrease in requests compared to 2006 (outlier years 2007, 2011 and 2012 are not included) (blue line) at the SAA.

5.2.3.4 Scan on Demand versus block digitisation

Further analysis showed that, compared to the the digitisation of blocks of (popular) archives, SoD targets the archival records that have ever been accessed in the reading room more effectively (42% of the SoD records versus 33.5% for block digitisation) (Fig. 5.8).

In absolute terms, Figure 5.9 shows that annually, ca. 1,600 records requested in the reading room have been digitised using SoD, whereas the figures for block digitisation vary between 500 and 5,500. If the popularity of the records (number of times requested in the reading room) is also plotted (Fig. 5.9, dashed line), it shows that block digitisation in certain years was more successful at targeting popular archives than in certain other years. In the case of SoD, the correlation between the number of digitised records that have been requested in the reading room and their popularity (number of requests) remained stable through the years.

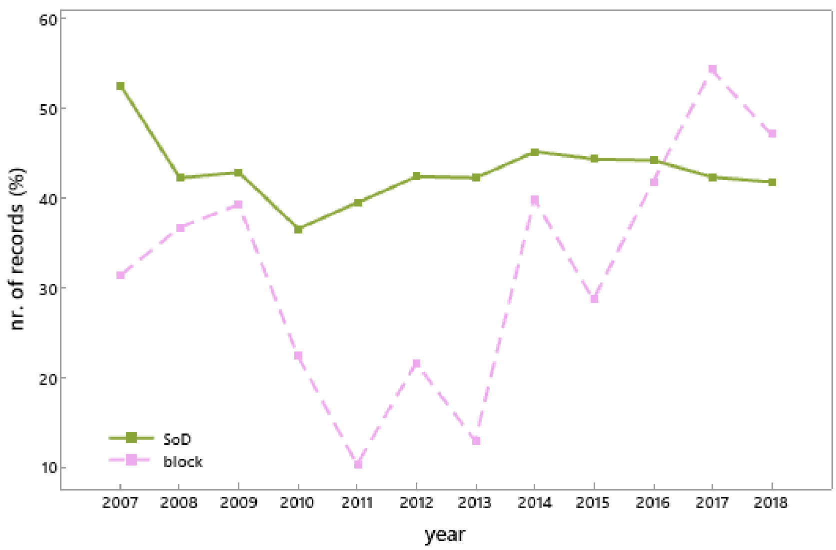


FIGURE 5.8: Percentage of records requested in the reading room *and* digitised according to the two digitisation programs at the SAA.

Regarding the popularity of the digitised records, Figure 5.10 shows that the records chosen to be digitised follow the same distribution as the distribution seen in the access requests in the reading room. Completely by chance, the two digitisation programs have equally contributed to the digitisation of popular records: the sum of the number of times requested through the years (Fig. 5.9, dashed line)

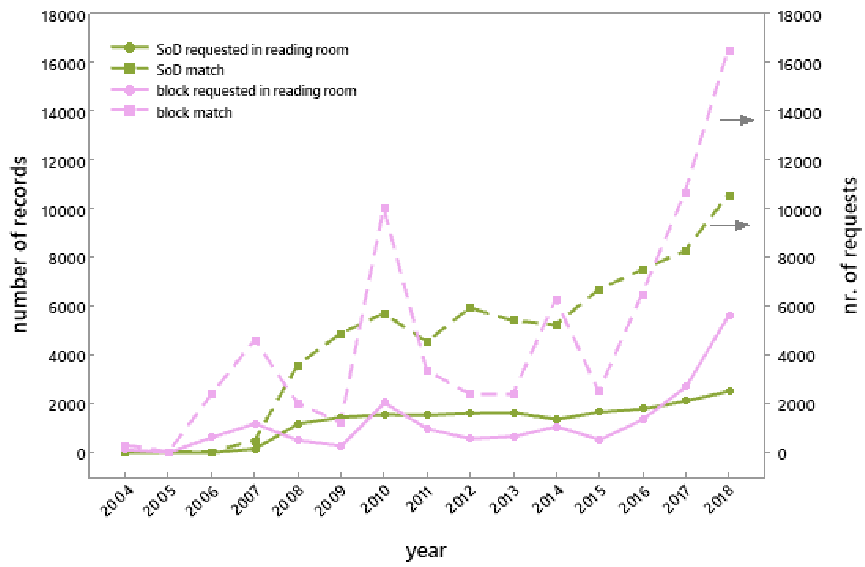


FIGURE 5.9: Number of records requested in the reading room that have been digitised (lines) and sum of requests history in this group (accessed and digitised) (dashed lines), according to the two digitisation programs at the SAA.

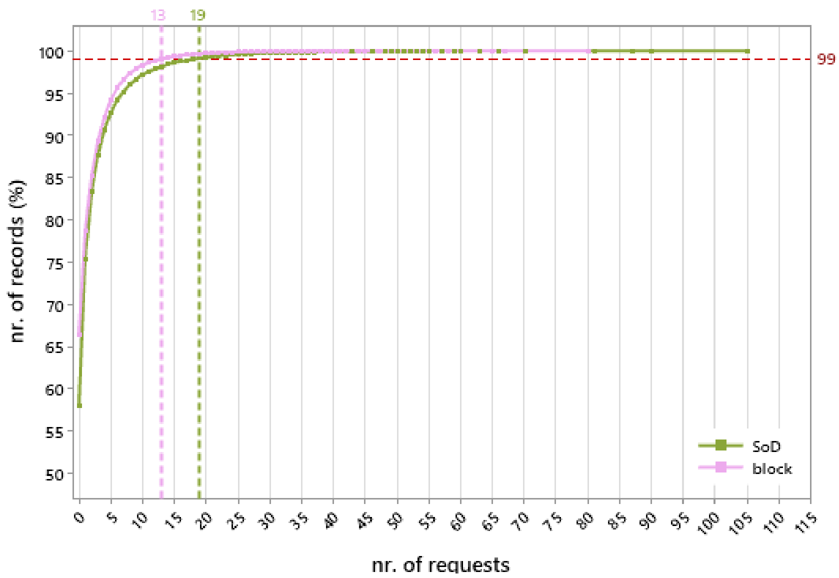


FIGURE 5.10: Cumulative percentage of number of records digitised within the two digitisation programmes, grouped by accumulated instances of access at the SAA.

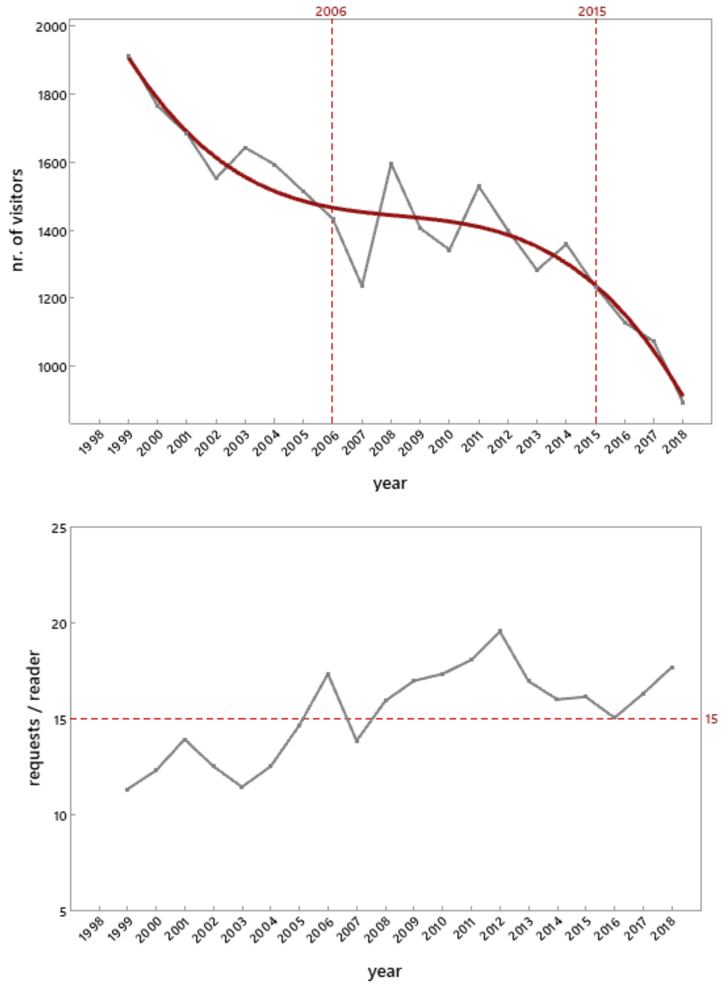


FIGURE 5.11: (*Top*) Number of individual readers between 1998 and 2018, with fitting line. (*Bottom*) Average number of requests per reader at the SAA.

is 68,717 and 71,036. However, as only 42% of the records have been digitised using SoD, this programme seems to be more effective at targeting the (popular) records requested in the reading room. If this difference is corrected, then the SoD has digitised 65% of the popular records compared to the 35% due to block digitisation.

5.2.3.5 Readers' behaviour

Since 1999, the number of individual readers had steadily been decreasing (Fig. 5.11, *top*), long before the start of digitisation programs in 2004, but at the same time, the average number of access requests per reader has increased over the years (Figure 5.11, *bottom*).

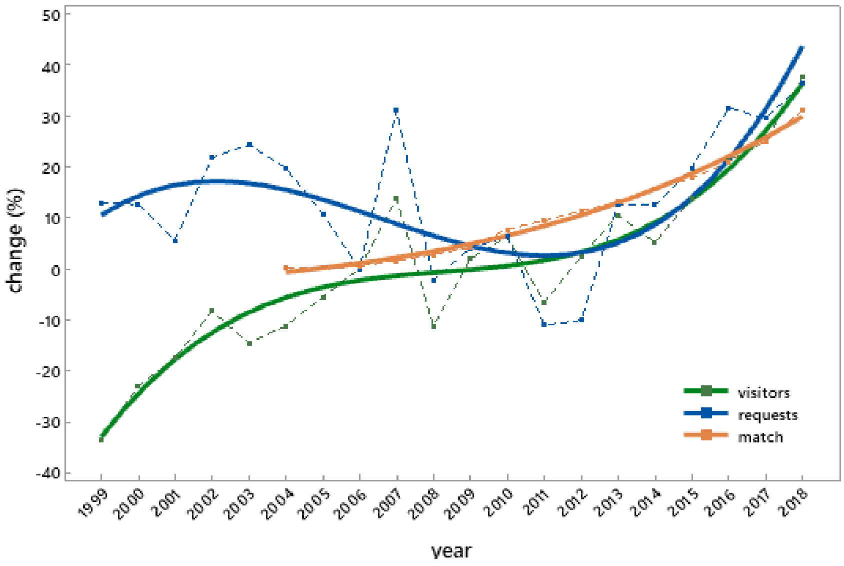


FIGURE 5.12: Annual change (percentage) in the number of readers and requests, and 'digitisation match', compared to the reference year (2006) at the SAA. Fitting lines are included.

Figure 5.12 shows the relationship between the change in the number of readers and the number of requests, as well as the digitisation match. The data show that before 2006, the number of readers had already begun to decrease, but this decrease did not affect the number of access requests. Around 2006, the number of visitors and instances of access seemed to plateau, and the correlation between the decrease in the number of readers and requests became stronger. Since 2015, there has been a more rapid decrease in the number of readers and requests. This decrease is occurring in parallel to the increase in the percentage of digitised records *and* requested in the reading room (digitisation match).

To study the behaviour of the readers, we classified them into five groups according to the number of requests made by each reader per year (1, 2–9, 10–49, 50–99, $>$ 100 annual requests per reader). Figure 5.13 indicates that regardless of the number of readers, the proportion between the readers groups has remained stable during the 20-year-period. Two patterns related to the behaviour of the group with more than 100 requests per year have been distinguished: from 1999 to 2003, this group accounted for 28% of requests, but from 2004, the percentage of requests increased to 48% (Fig. 5.14). If the second group is added (readers that requested 50—99 records per year), then ca. 6% of the readers are responsible for 55% of requests, while in 1999, the percentage of requests by these two groups was ca. 44%.

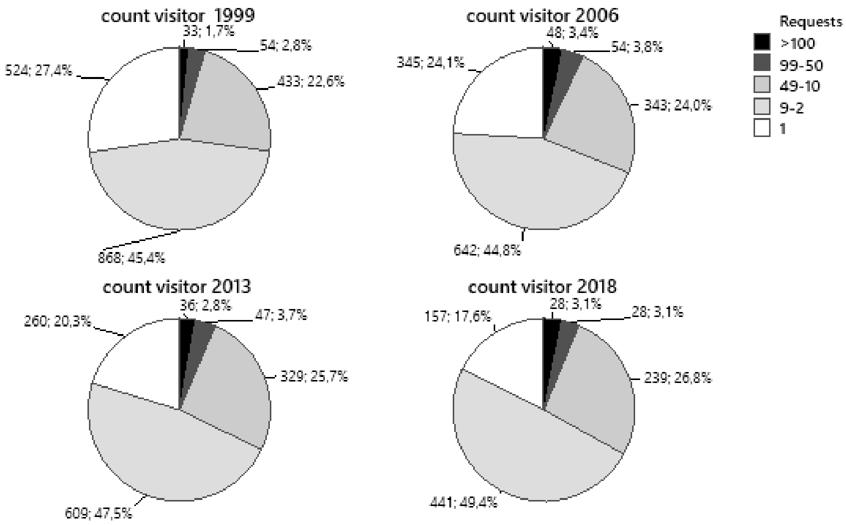


FIGURE 5.13: Number of readers grouped by the number of access requests per reader in 1999, 2006, 2013 and 2018 at the SAA.

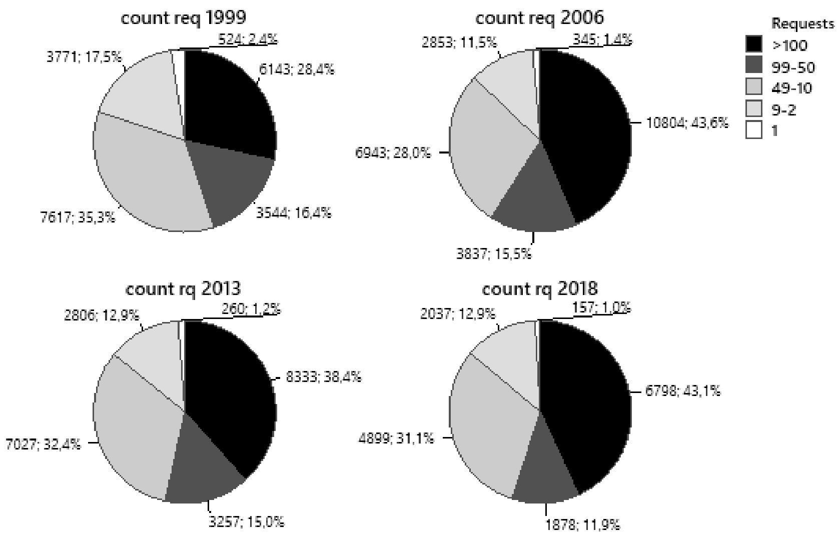


FIGURE 5.14: Number of access requests in 1999, 2006, 2013 and 2018, grouped by reader type, based on the annual number of access requests per reader at the SAA.

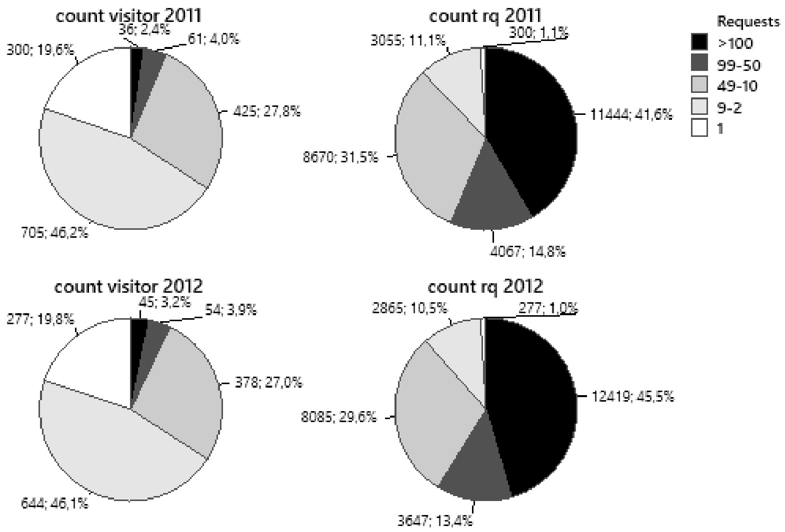


FIGURE 5.15: Number of readers and requests in 2011 and 2012, grouped by reader type, based on the annual number of access requests per reader at the SAA.

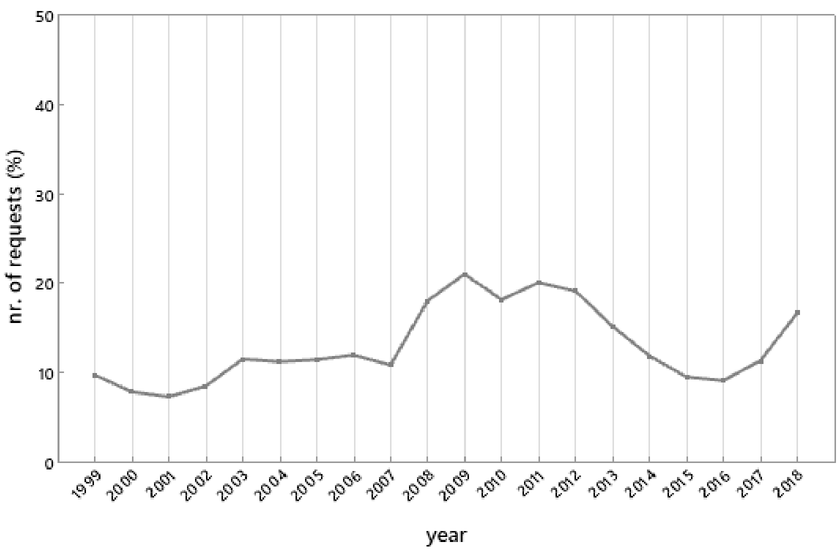


FIGURE 5.16: Percentage of records requested by the top five readers between 1999 and 2018 at the SAA.

The years 2011 and 2012 followed the same patterns as 2004 (Fig. 5.15). The graphs indicate that these two years are outliers due to the slightly higher number of readers, particularly those requesting a high number of records.

Traditionally, a small group of readers (55—100) has been responsible for almost half of the requests. Further analysis of this group of readers showed that between 2008 (after the archive moved to a new location) and 2012, the top five individual readers accessed yearly ca. 20% of the requested records. After a decrease in the activity of this group, a slight increase was observed since 2017. It is not clear whether the decrease in the number of requests from this small group is a direct effect of digitisation. Another explanation could be that some of the readers finished their research at the archive. It is important to note that a slight

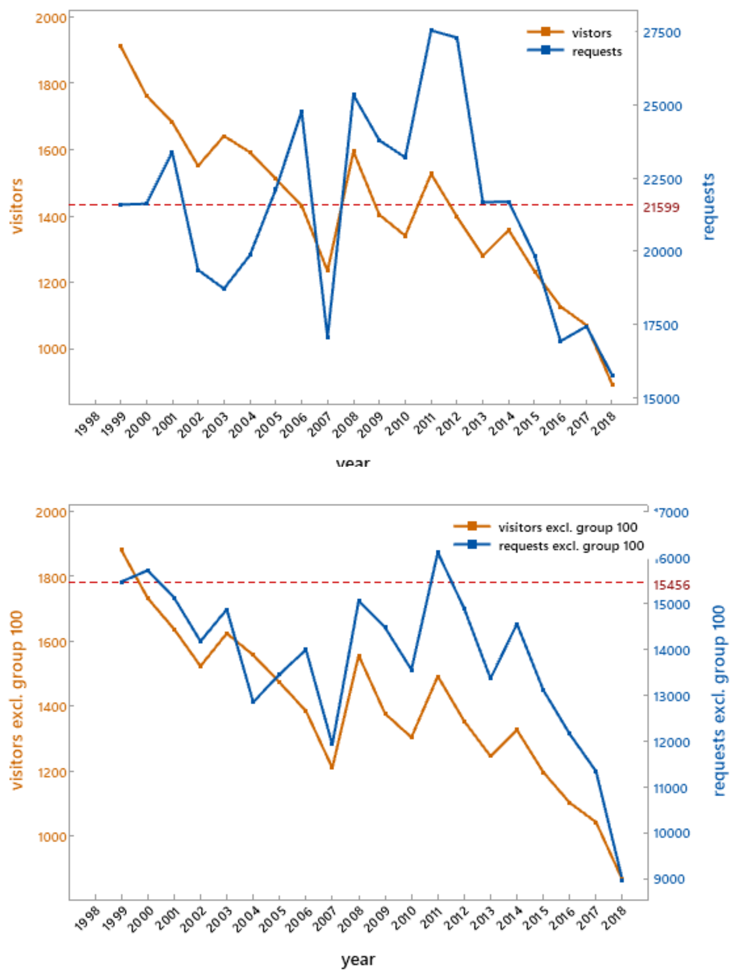


FIGURE 5.17: (*Top*) Number of individual readers and requests at the SAA. (*Bottom*) Number of individual readers and requests, not including the group of readers with more than 100 requests.

change in the behaviour of this small group of readers might have a noticeable impact on the total number of requests in the reading room. For instance, a slight change in the behaviour of the readers in 2018 might result in an even sharper drop in the total number of requests in the reading room (Fig. 5.16).

The importance of this small group of readers requesting more than 100 numbers per year is further stressed in Figure 5.17. The activity of this group explains the difference between the number of visitors and the expected number of access requests. As a small group of visitors has a notable impact on the number of requests, any forecasting of the expected number of requests is subject to a high level of uncertainty.

5.2.4 Conclusions

The aim of this study was to obtain evidence for the hypothesis that digitisation of the collections reduces the use of the collections in the reading room. No evidence has been found that rejects the hypothesis; however, other factors and trends that might have influenced the number of requests in the reading room were also observed. The most important findings are as follows:

- The downward trend in the number of readers and requests had already started before the introduction of digitisation. This trend seems to be driven by external socio-cultural factors, such as a decrease in the general interest in archives.

- This ongoing downward trend slowed down from 2006 to 2015. During this period, although the number of readers continued to slightly decrease, the average number of requests per reader increased, resulting in a higher number of requests in the reading room than before 2006. This stabilization could be explained by the introduction of a new digital platform to access the collections.

- However, from 2015, there was a rapid decrease in the number of readers and requests. That might be a direct result of the digitisation of the collections. During this period, a clear decrease in the use of frequently requested records was observed as a result of digitisation.

- It is important to note that the number of requests was strongly influenced by a small group of readers (5% of the readers who are responsible for over 50% of the requests). Therefore, there is a high level of uncertainty because any small change in the activity of this group has an evident impact on the number of requests.

- The digitisation of frequently requested records in the reading room added to the decrease in the use of collections in the reading room. So far, the SoD service appears to be more effective to target this group of frequently records. The effectiveness of block digitisation strongly varies depending on the chosen archives to be digitised. However, block digitisation has the potential to be an effective measure to further reduce the physical use of the collections. The study of other case studies is essential to evaluate if this relationship can be found in other archives.

Regarding the expected future trends, the analysis of the data from 1999 to 2018 seems to indicate that the downward trend in the number of readers and requests will continue in the next years. It can be expected that the decrease will progressively slow down as the percentage of popular items not digitally available further decreases after digitisation. However, a high level of uncertainty is present, owing to the small but highly active group of readers and the difficulty in anticipating how readers will further adopt the SoD service in the coming years.

Regarding the dynamic hypothesis to be modelled, the data indicate that the main factor related to the decrease in access requests is digitisation, which is also the most straightforward variable to be modelled. Based on the analysis of actual collection data, the most relevant variables to be included in the model are as follows:

- increase in the percentage of records requested for the first time;
- effectiveness of the SoD and block digitisation regarding the digitisation of records accessed in the reading room;
- accumulated instances of the use of digitised records.

The data analysis also indicates that other factors might have an effect on the number of requests; however, it may be difficult to quantify these factors. This is the case for the downward trend of readers or the factors contributing to the increase in the average number of requests per reader between 2006 and 2015 as a result of improving the digital access tools. As already noted, both the digital content and metadata and search tools are crucial to provide access. Similarly, it is also difficult to quantify whether the digitisation of the collection that has not been accessed in the reading room also leads to a decrease in the number of readers. From 2015, we observe that the decrease in the number of readers is slightly higher than the percentage of digitised accessed records. This difference might be the consequence of an increasingly larger part of the collection becoming digitally available. However, more research is needed to obtain reliable data to quantify these variables before being included in the model.

5.3 Survey of readers' preferences

5.3.1 Introduction

Since 2017, digitally available records have been accessible and free of charge on the SAA website. To access records that are not digitally available, readers have the choice between visiting the reading room or requesting (free of charge) digitisation of the records. So far, the SAA has no data on readers' preferences regarding accessing the records at the reading room or requesting digitisation and accessing the scans.

The readers at the reading room were surveyed to collect information on readers' preferences for accessing archival records in the reading room and to identify when readers would be willing to consider the use of the SoD service instead of accessing the records in the reading room. The aim of the survey was two-fold: to inform the SAA about how different options could promote the use of the SoD among the readers and to investigate whether the studied variables (regarding the factors associated with readers' preference for using the SoD service or visiting the reading room) should be included in the APM model.

5.3.2 Methodology

During a six-week period from 18 June to 26 July 2019, readers in the reading room were requested to participate in the survey by completing a (written) anonymous questionnaire (see Appendix B). The potential participants were approached by the reading room staff when the archival record was handed on to the reader. Forty-two readers completed the questionnaires, of which two were incorrectly filled.

The questionnaire consisted of two types of questions: multiple-choice questions and fixed-sum scale questions.

In the first part of the questionnaire, multiple-choice questions were formulated to identify the type of reader:

- type of reader (e.g. historian, researcher, teacher)

- how often the respondent visits the reading room
- how often the respondent uses the SoD service

Information on the number of records requested by the respondent on that particular day was also sought. In addition, a question about the type of research they were conducting was asked to identify readers who needed to access a certain archival record first to determine which record will be requested next.

In the second part of the questionnaire, fixed-sum scale questions were designed to reflect readers' preference for the SoD service or accessing records at the reading room based on the following factors:

1. opening days of the reading room
2. maximum number of requests for digitisation per week
3. delivery time for the scanned copies
4. preference for accessing original records instead of scanned copies
5. no choice as records are not eligible for digitisation.

If the survey participants responded to one of the first three items, then an open question was asked to quantify the responses.

5.3.3 Results and discussion

During the survey period, 150 readers were approached, of whom only 42 readers returned the questionnaires. Due to the small population size and the low response rate, it is not possible to use these results to make inferences about the population due to a rather wide confidence interval (95%CI = 50% ±11%). To obtain ±5% margin of error (95%CI), the sample size should be 100 (two-thirds of the population). It is, however, difficult to obtain a high response rate in a small population, considering the voluntary nature of the survey. Therefore, the results only describe the preferences of the sample, and the trends observed should only be interpreted qualitatively.

Most of the respondents identified themselves as historians (Table 5.1). The low number of genealogists among the readers is interesting, which might be an outcome of large-scale digitisation programmes to make records digitally available that are of interest to genealogists. Statistics also show that 76% of the research

TABLE 5.1: Type of reader given by respondents of the questionnaire ($n = 42$).

Reader type	Respondents (%)
Historian	29 (72.5%)
Genealogist	5 (12.5%)
Researcher	2 (5.0%)
Journalist	3 (7.5%)
Student	1 (2.5%)

studies do not require the use of indexes, and eight archival records were requested per reader (Fig. 5.18). These results can be seen as indirect factors that might be related to readers' preferences for visiting the reading room instead of using the SoD service. Depending on the type of research, the reader will first access a record, and from this record, it will become clear which record needs to be requested next. For this type of research, it is more efficient to visit the reading room than having to wait up to two weeks to obtain a scan.

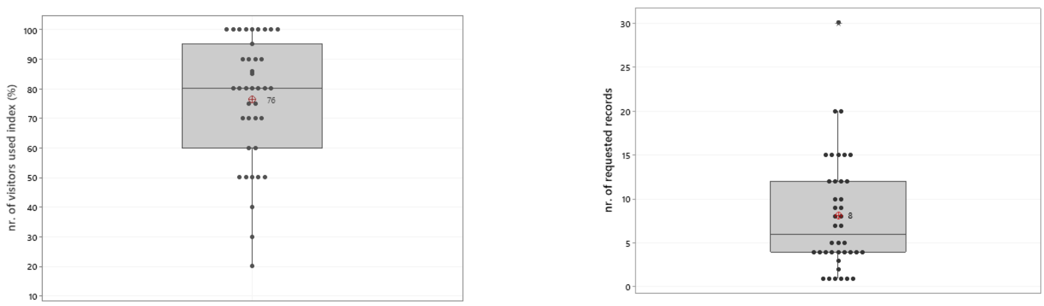


FIGURE 5.18: (*Left*) Percentage of research studies that do not use indexes, conducted by the respondents. (*Right*) Number of records requested by the respondents on the day they participated in the survey.

If the responses are grouped by how often the respondent visited the reading room, it appears that there were SoD users in all groups. The group that visits the reading room yearly use the service more frequently (Fig. 5.19, *left*). Regarding whether the respondents would be willing to use the SoD, most of the respondents in all groups were willing to consider it to a greater or lesser extent, provided the delivery time of the scan is shortened (Fig. 5.19, *bottom*). The yearly readers of the reading room were the most willing to choose the SoD, and their preference for accessing the originals was the lowest. The weekly readers seemed to prefer accessing the originals, and therefore, they were less willing to consider the SoD. However, it is worth noting that the largest group in all cases comprised respondents who choose both options: they were willing to consider SoD, but they also valued access to the originals.

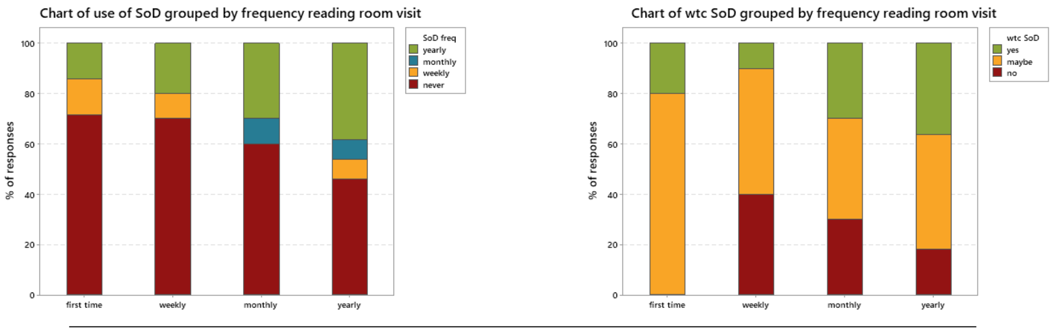


FIGURE 5.19: The actual use of SoD (*left*) and respondents' willingness to use the SoD (*right*), grouped by the number of times the respondents visited the reading room.

Regarding respondents who were willing to consider the SoD provided the scan delivery times were shortened, Table 5.2 shows how long the respondents were willing to wait for a scan. Almost 60% of these respondents reported an ideal wait time of up to five days to consider the SoD.

TABLE 5.2: Increase in the percentage of respondents willing to consider using the SoD based on the scan delivery times.

Delivery scan (days)	**wtc (cum. %)**
10	4.2
7	25.0
6	29.2
5	58.3
4	70.8
3	79.2
2	95.8
1	100

Table 5.3 shows that respondents' preference for accessing the originals and the scan delivery times were the most important factors, followed by opening hours. Table 5.4 shows how reducing the number of open days of the reading room would affect the choice for the SoD.

Regarding other factors, although 50% of the respondents had ever come across

TABLE 5.3: Responses to the survey items regarding the factors associated with the decision to use the SoD instead of visiting the reading room.

	Responses	Weight
preference original	27 (67.5%)	1330 (33.3%)
delivery time scan	27 (67.5%)	886 (22.2%)
open days	22 (55.0%)	590 (14.8%)
no eligible for digitisation	20 (50.0%)	477 (11.9%)
max. nr. digi. requests/week	18 (45.5%)	347 (8.7%)
otherwise	7 (17.5%)	370 (9.3%)

TABLE 5.4: Increase in the percentage of respondents willing to consider using the SoD based on the opening hours of the reading room.

Opening days	wtc (cum. %)
3	35.7
2	78.6
1	92.2
0	100

records that were ineligible for digitisation and therefore only accessible in the reading room, it is not the main reason why respondents visited the reading room. The maximum number of requests per week for the SoD only appeared to play a weak role compared to other factors.

5.3.4 Conclusions

A group of respondents preferred originals to digital scans if they were given a choice. Respondents who visited the reading room more frequently seemed to have a stronger preference for original records than those who visited the reading room yearly. However, among both frequent and infrequent groups, more than half of the respondents were willing to consider the use of the SoD.

Among respondents who were willing to consider using the SoD, the actual scan delivery times seemed to play a major role. Most of them were willing to use the SoD if the delivery time was five days or fewer. This relationship seems to be linear.

However, when evaluating the potential effect of reducing the scan delivery time on the decrease in the number of requests at the reading room, it is important to keep in mind that the frequent users of the reading room were less likely to choose the SoD and were responsible for the half of the requests at the reading room (as seen in the previous analysis of reading room requests from 1999 to 2018). Therefore, the effect on the decrease in the frequency of requests in the reading room might be less noticeable than expected.

The willingness of readers to use the SoD over the traditional method of accessing records in the reading room is a crucial factor that might further intensify the downward trend in the number of access requests. If some of the following variables change through the years, there might be a visible effect on the number of requests:

- delivery scan
- complexity research
- opening hours reading room

However, although these results are a good indication of the main factors that would be worth including in the model, more research is needed to quantify them. For now, these variables could simply be introduced into the model to explore the magnitude of change when different inputs values are used, in a straightforward manner, without taking into account more complex relationships (e.g. the observed difference in responses depending on how often the reader is using one of the two services).

5.4 Modelling access with system dynamics

Based on the findings in the previous sections, the hypothesis of the model is that the main driver behind the decrease in the number of requests in the reading room is the digitisation of records that had been requested in the reading room. We developed first the model in SD, where records are modelled as aggregates based on whether they had been requested in the reading room and digitised. In the next section, the development of a second model in ABM is described, where the archival records are modelled individually, and therefore, the popularity of records, defined as the instances of access in the reading room, is also modelled.

5.4.1 Model boundaries

Because this is the first time a simulation model was built in this research field, we began by building the simplest model in SD: the dynamics of the number of requests in the reading room were modelled endogenously, but other variables (e.g. the number of requests for digitisation) were modelled exogenously. For these variables, we directly used the data provided by the management system of the archives as exogenous inputs, and a further endogenous explanation for their dynamics was not sought.

Consequently, certain elements of the system were excluded from the model. The model only included variables related to the digitisation of the collections. Other dynamic hypotheses identified in the causal loop diagram (Chapter 3), such as the positive feedback of the use of the collections that results in more use, were not part of this model. Similarly, other dynamics seen in the data set (e.g. the ongoing downtrend of the use of the collections before digitisation started or the temporary stabilization due to the improvement of the digital platform) were not included.

Another element that should be added to a comprehensive model is the behaviour of the readers: both the number of readers and the frequency of requests and their preferences for digital or analogue access. However, there is an evident gap in information, and further research is warranted to obtain more data to support the development of a complex model. These variables are needed if the effect of modifying the access to the collections needs to be explored. For now, the model assumes that the reading room and digitisation service remain unchanged.

5.4.2 A brief reflection on the system dynamics application

Before a detailed description of the model, it is worthwhile to reflect on the implications of the proposed simplification in the context of the system dynamics application. Figure 5.20 shows that, by omitting variables that were originally included in the causal loop diagram, an important foundation of the system dynamics is lost, namely the endogenous view of the system.

In Chapter 4 we discussed that the system dynamics approach seeks to find the explanation of the dynamic behaviour of the system within the system boundary. The aim of the causal loop diagram was therefore to describe the system mainly by endogenous variables (variables affected by other variables of the system). With the simplification of the model, variables which were originally described as endogenous become exogenous variables. For the access part of the model, this means that information on the dynamic behaviour of the system is inevitably lost. Regarding the reading room requests, in the mathematical model only one hypothesis is modelled: when more records become digitally available fewer requests to access the records in the reading room are made. Consequently, possible oscillations of the number of access requests caused by other dynamics are lost (e.g. when other reinforcing or balancing loops might be met). Moreover, in the mathematical model the digitisation requests have become exogenous variables, omitting for now the dynamics behind these variables related to the readers behaviour. Herewith the mathematical model is mostly applying an exogenous perspective, with no rich feedback structure.

The implications of choosing an exogenous or endogenous perspective of the system have been argued by Richardson (2011). A system can be approached from a exogenous perspective, where dynamics are governed by external forces outside our control, or endogenous perspective, where 'the act of trying to govern/manage/control generates system dynamics of its own"(Richardson; 2011, p. 239). As Richardson makes clear with different examples, the search for the endogenous sources of the system behaviour is what defines the system dynamics approach. By applying an endogenous perspective, dynamic behaviours (e.g. oscillations) can be explained which otherwise would have been missed or misunderstood.

In our case, it is clear that the fully endogenous modelling of the system, as presented in the causal loop diagram, would help to explain the oscillations seen in the historical data (see Fig. 5.1). However, the same historical data seems to indicate that digitisation is the main source of the predominant dynamic behaviour (a decrease in requests made to the reading room), which justifies that only a limited number of variables are included in the most simple model. In doing so, the core of the model only includes variables related to numerical data on the use of the collections, as an exogenous approach is sufficient to explain the main dynamics observed in the historical data. However, in a more elaborated model, the omitted behavioral variables would need to be included, restoring the endogenous perspective of the system.

Although for the moment one of the foundations of the system dynamics will not be further explored in the mathematical model, other aspects are worth to investigate which are distinctive of the system dynamics approach compared to other simulation approaches, for instance the level of aggregation in the system dynamics compared to the micro-modelling in the agent based modelling approach. This topic, the level of aggregation, will be the main focus of the comparison of the two approaches presented in this chapter.

5.4.3 Stocks and flows

In the proposed SD model, the system, formed by stocks, flows and auxiliaries that determine the value of the flows, includes two interconnected parts: one part represents the collection itself, whereas the second part models the use of the collections (access requests in the reading room and requests for digitisation) (see Fig. 5.20).

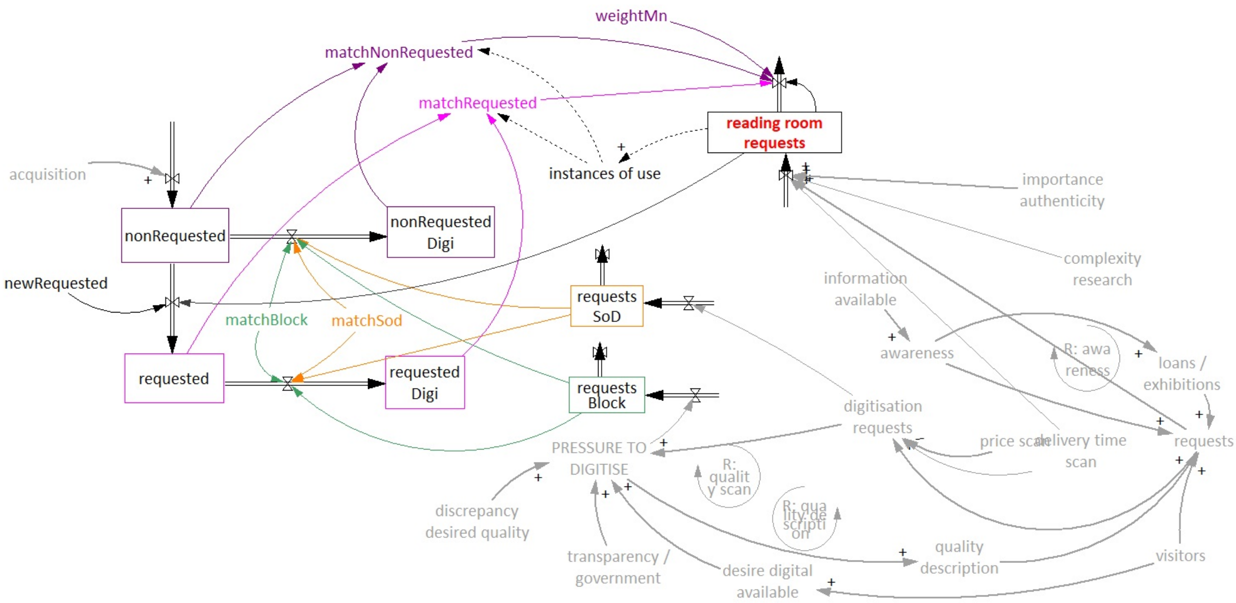


FIGURE 5.20: Visualisation of the access part of the APM model. The stock flow diagram shows the relationship between the level of demand in the reading room and the digitisation of the physical collections. In grey the variables identified in the causal loop diagram that for now are not modelled.

Regarding the use of the collections, the different types of requests are represented by three stocks:

1. Requests in the reading room, R_r
2. Requests within the digitisation programme Scan on Demand, R_s

3. Requests to digitisation initiated by the institution, R_b.

In this model, the stocks represent archival documents as aggregates that can flow between four stocks, according to whether they have been digitised or requested. No information on the number of requests per record is included. The collection can be described by four stocks:

1. Records never requested or digitised, C_n
2. Records requested but not digitised, C_a
3. Records never requested but digitised, C_{nd}
4. Records requested and digitised, C_{ad}

The unit of all stocks is an archival record, including the digitisation stocks. There is only one exception: the stock representing requests in the reading room. As an archival record can be accessed more than once in one year in the reading room, the unit of the stock representing the accesses in the reading room is an access request.

The purpose of the model is to simulate the decrease in R_r due to the digitisation programmes R_s and R_b. However, as is shown in Figure 5.20, the impact of digitisation depends on whether the records that become digitally available have been requested in the reading room (C_{ad}) or not (C_{nd}). Figure 5.20 also shows that there is only a feedback loop in the model. How the records from the state 'non requested' (C_n) flow to the state 'requested' (C_a) depends on the portion of the requests in the reading room R_r requested for the first time. In the experiments section, we will show that this feedback loop is a crucial part of the model. Inputs related to the number of requests for digitisation are exogenously modelled for now, and the input values are based on the data set (see Table 5.6 in the experimental results section). Herewith we the endogenous view, as described in the causal loop diagram, is missing.

5.4.4 Mathematical model

The main output of the model is how the annual number of requests (R_r) changes during the computational experiment. We assume that the main mechanism that explains the change in R_r, in this case, the decrease, is related to the fraction of the collection that annually becomes digitally available, and it is defined as follows:

$$\frac{\partial R_r}{\partial t} = -R_r(M_a + M_n) \tag{5.1}$$

TABLE 5.5: Nomenclature access model

	Name	Unit	Definition
R_r	requestsRR	records	requests reading room
R_s	requestsSoD	records	requests for SoD
R_b	requestsBlock	records	requests for block digitisation
C_n	nonRequested	records	records neither requested nor digitised
C_a	requested	records	records requested but not digitised
C_{nd}	nonRequestedDigi	records	records never requested but digitised
C_{ad}	requestedDigi	records	records requested and digitised
M_a	matchRequested	%	fraction of requested records that become digitally available
M_n	matchNonRequested	%	fraction of never requested records that become digitally available
M_s	matchSoD	%	fraction of requested records for SoD that have been requested
M_b	matchBlock	%	fraction of requested records for block digitisation that have been requested
I	increaseRequested	%	fraction of R_r requested for the first time
w	weightMn	%	fraction of M_n
P	newRequested	%	fraction of R_r that are requested for the first time
t	time	year	

where M_a refers to the annual fraction of records requested in the reading room that became digitally available, and M_n is the fraction of digitally available records that had not been requested in the reading room.

Let us define M_a as the annual rate of requested records that have been digitised, expressed as a fraction:

$$M_a = \frac{\frac{\partial C_{ad}}{\partial t}}{C_{ad} + C_a} \tag{5.2}$$

If M_a is closer to one, it means that almost all records digitised had been requested in the reading room prior to digitisation.

A part of the collection that has not been requested is also digitised. We can assume that there is a chance that records that might have been requested for the first time have already been digitised. Therefore, the model also includes the impact of the subset of the collection that has become digitally available, even for those records that have not been requested in the reading room prior to digitisation. However, as we can only speculate about how many records would have been requested for the first time if they had not been digitally available, there are no data available on this group of records, M_n can only be expressed as a function of the collection becoming digitally available:

$$M_n \sim w(\frac{C_{nd}}{C_n + C_{nd}}) \tag{5.3}$$

where the constant w (any value between 0 and 1) modifies the fraction of non-requested records that have been digitised, denoting that an unknown number of records would have been accessed in the reading room for the first time if they had not yet been digitally available. If the value of M_n is closer to 1 (being w also closer to 1), it means that most of the records selected for digitisation, even if they had not already been requested in the reading room, are of interest to the readers and would have been requested in the reading room if they had not been made digitally available.

To calculate Equation 5.2 and Equation 5.3, we need to define how the four stocks that form a collection change over time. We start defining the part of the collection that has been requested but not digitised, C_a, as follows:

$$\frac{\partial C_a}{\partial t} = \frac{\partial I}{\partial t} - \frac{\partial C_{ad}}{\partial t} \tag{5.4}$$

where I, the increase of records requested for the first time, is calculated using R_r, assuming that a fraction of R_r is requested for the first time:

$$\frac{\partial I}{\partial t} = R_r P \tag{5.5}$$

where P can take any value, between 0 and 1, to modify R_r.

The requested records, C_a, that become digitally available, C_{ad}, are the result of the two digitisation programmes:

$$\frac{\partial C_{ad}}{\partial t} = R_s M_s + R_b M_b \tag{5.6}$$

where R_s is the number of records digitised within Scan on Demand and R_b is block digitisation. As this part of the model only represents the requested records, R_s and R_b are respectively modified by M_s and M_b.

Regarding the records that have not been accessed, C_n decreases due to the access of records for the first time and due to the digitisation of this part of the collection, according to the following equation:

$$\frac{\partial C_n}{\partial t} = -(\frac{\partial I}{\partial t} + \frac{\partial C_{nd}}{\partial t}) \tag{5.7}$$

To simplify the model, it is assumed that the collection size is constant, as the total number of records in the collection is not a key parameter in the model.

Similarly to the digitisation of requested records, the digitisation of non-requested records depends on the number of requests within the two digitisation programmes, SoD and block digitisation:

$$\frac{\partial C_{nd}}{\partial t} = Rs(1 - M_s) + R_b(1 - M_b) \tag{5.8}$$

The SD model was solved in Microsoft Excel, which makes it accessible to a wide user base including decision-makers in archival institutions.

5.4.5 Experiments

We conducted four simulation experiments with the aim of exploring whether the observations made in Section 5.2 could be reproduced. Therefore, the data source is the data set of the SAA. In addition, the experiments were conducted to identify those variables that contribute the most to the output.

5.4.5.1 Experiments description

These experiments simulated the period from 2006, when the digitisation programmes started, to 2018. At model initialisation, the collection consists of one million records, and the fraction of records requested is 0.1. The collection size is assumed to be constant, and the growth of the collection is neglected because no data were available on the total number of records that formed the collection during this period.

In Table 5.6, the input values used for these experiments are summarised. The data source is the data set of the SAA, except for those parameters that are an assumption and where the values are specified.

A total of four experiments were conducted. In each experiment, a new factor of the model was added as follows:

Experiment 1: All digitised records have a direct impact on the reading room requests, R_r.

Experiment 2: Only digitised records that have been requested in the reading room, C_{ad}, have an impact on R_r.

Experiment 3: The same as Experiment 2, but taking into account that every year,

TABLE 5.6: Input variables of the system dynamics model. See Appendix F for SAA data set.

Variable		Value	Type
Records in the collection	C_i	1,000,000	integer
Initial requests	R_r	SAA data set	integer
Initial requested	P_i	0.1	fraction
New requested	P	SAA data set	fraction
Requests SoD	R_s	SAA data set	integer
Requests Block	R_b	SAA data set	integer
Match SoD	M_s	SAA data set	fraction
Match Block	M_b	SAA data set	fraction
Weight match	w	0.3	fraction

new records are accessed for the first time, P (Equation 5). At the end of the experiment, 15% of the collection has been accessed.

Experiment 4: The same as Experiment 3, but the impact of the records never digitised in the reading room, C_{nd}, is also included and modified by w, a hypothetical weight (Equation 3).

5.4.5.2 Results and discussion

The results show that Experiment 1 is not a realistic scenario. In this experiment, where all digitised archival records (M_a and M_n) contribute equally, the effect of digitisation on the decrease in the access requests in the reading room (R_r) is overestimated (Fig. 5.21). Interestingly, a better fit to the actual requests (Experiment 2) is obtained when only the archival records that have been requested and digitised (M_a) are included. However, a factor is still missing in Experiment 2: every year, new archival documents are requested for the first time (P). This factor is important as it modifies the fraction of requested records that are digitally available (M_a), according to Equations 4 and 5. Therefore, Experiment 3 is a more realistic scenario than the one presented in Experiment 2. Although Experiment 3 reproduces a steady decrease in requests from 2006 to 2014, it fails to reproduce the sharper drop beginning in 2015.

We can think of three different hypotheses that may explain why Experiment 3 does not show the drop in requests, looking at the model boundary and the factors excluded in the model:

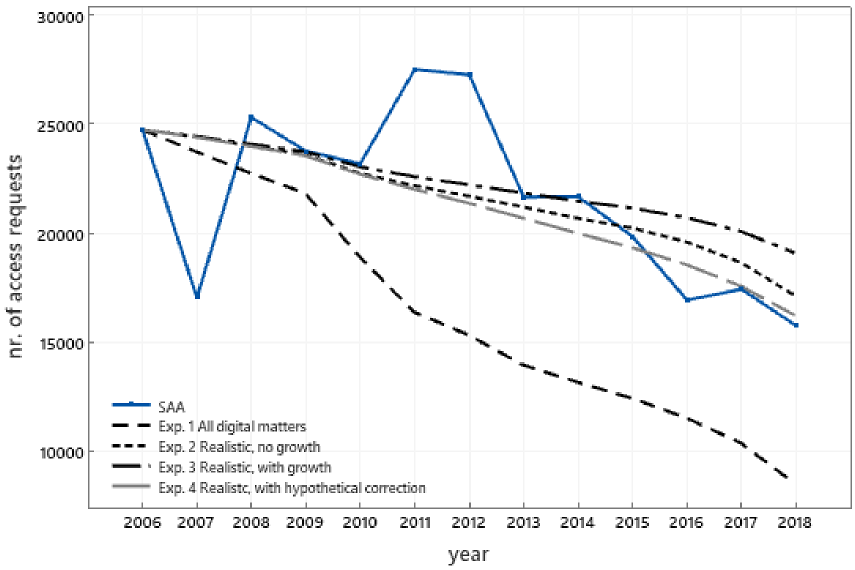


FIGURE 5.21: Comparison of the decrease in access requests according to the SAA data set and the output of four experiments conducted according to the SD approach.

1. The impact of the digitisation of the subset of the collection that has not been requested in the reading room.
2. The impact of the digitisation of the records that have been repeatedly requested, namely the increase in the digitisation match from 2004 (Fig. 5.22).
3. The decrease in requests due to social or economic reasons.

In Experiment 4, we tested whether a better fit could be obtained throughout the simulated period if a hypothetical correction was applied. In this experiment, the never-requested-but-digitised archival records were included. However, instead of contributing equally to the decrease in requests (Experiment 1), this contribution was modified by a hypothetical constant value (w) in this experiment (Eq. 5.3). Different values of w were tested until the best fit was found (Table 5.6).

According to Experiment 4, w should have a value of 0.3 for the best fit. In other words, each year, 30% of the never-requested-but-digitised archival records would have been requested in the reading room if they had not been digitised. However, no historical data are available to support this hypothetical value of w which seems not to be realistic. In contrast, in the analysis of the SAA data set, we found evidence that both the requested archival documents and the number of times that they have been accessed contribute to the decrease in the requests

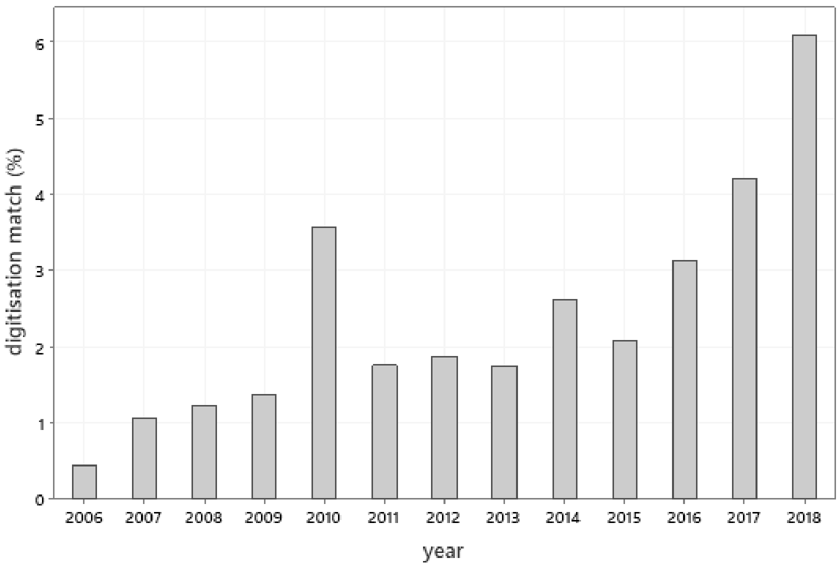


FIGURE 5.22: Annual 'digitisation match', sum of requests history of requested *and* digitised records, given as percentage, in the SAA data set.

in the reading room. From these experiments, we can conclude that Experiment 3 (including P, but not w) is the most realistic, with a relatively good fit. This experiment will be used from now on to validate the SD model in the case studies.

In the next section, we test whether the sharp decrease in requests from 2015 that could not be reproduced in the SD approach (Experiment 3) is obtained when the individual number of instances of access to the archival documents is also included in the model.

5.4.5.3 Limitations of the system dynamics model

From the series of experiments, it can be concluded that the experiments of the SD model underestimate the decrease in requests in the reading room: 30% decrease in the SAA data set versus 22% in the SD results. The SD results are only valid as a broad indication of the impact of the digitisation on the use of the collections. It is important to note that this will be true for collections similar to the SAA data set, where although ca. 60% of the requested records have been accessed more than once, this group accounts only for 20% of the total requests. But even more relevant, almost 50% of the total requests is the result of records requested between two and six times in 20 years. We can expect that an accurate fit will be even more difficult to meet in the case of collections where the group of popular records account for a high percentage of the total requests. In addition, the SD will lose further accuracy when the group of frequently accessed are part of the digitisation strategy, as the impact of the digitisation of the frequently requested records is not included in the SD model.

Most of the input values of the model are derived from data collected by the collection management system through the years. However, two variables are based on assumptions: the fraction of never-requested records that are digitally available (M_n) and the fraction of annual requests that are requested for the first time (P). Regarding the former, from Equation 5.3, we assume that there is a chance that records would have been accessed in the reading room if they had not been digitised before they were accessed for the first time. This chance is quantified by w, a speculative value. Figure 5.21 shows that the impact of M_n is expected to be small when C_n (never-requested records) is considerably larger than C_a (requested records), and w has a low value. Also, it is important to note that the model is based on the assumption that M_a (fraction of records digitised and requested) should be greater than M_n (fraction of records digitised but not requested). Therefore, if C_a (records digitised and requested) were larger than C_n (records digitised but not requested), then w should be equal to 0 or close to 0 to prevent M_n from becoming the determinant factor in the equation of the decrease in requests (Equation 5.1).

Hence, the emerging hypothesis of the SD model is that M_n plays a minor role and the decrease in requests is mostly explained by the digitisation of the requested records (M_a). This implies that when the group of ever requested (C_a) becomes digitally available (C_{ad}), then M_a equals 0 and the decrease in requests (R_r) stops, which also happens if R_r has not reached the value of 0. To illustrate this point, we ran two experiments using the same inputs as in Experiment 3 (Section 5.4.5.1), but this time, the number of digitisation requests per year are constant ($R_s = 5,000$ and $R_b = 5,000$) but the input values of M_s and M_b are changed to 0.5 (first experiment) and 0.8 (second experiment). In the first experiment, it would take 40 years to digitise the records requested in the reading room and by then, the number of requests will be 6723. If the digitisation match is increased (second experiment), then within 18 years, C_a will be 0, but the number of requests in the reading room (R_r) will still be 7508. These examples illustrate that other factors need to be included in order to model at what point the requests will definitely stop.

Another uncertainty in the model is the value of P, the fraction of records requested for the first time. The model assumes that the group of records requested in the reading room (C_a) increases yearly. This means that the slower the rate of digitisation, the greater C_a become and, therefore, the higher the number of records that need to be digitised. For example, if M_s and M_b are 0.5, then 205,000 records (20% of the collection) are digitised before C_a equals 0. If M_s and M_b are 0.8, then only 152,000 records (15% of the collection) need to be digitised. The model assumes that all the records have the same chance to be repeatedly requested. However, from the data set of the SAA, we know that ca. 50% of the records are only requested once and becoming digitally available will not necessarily have an impact on the reading room. Similarly, if not all records requested and digitised (C_{ad}) contribute equally, then not all records requested but not digitised (C_a) should be counted. From the data available for 20 years, it seems that these two groups compensate for each other. To be able to evaluate the role of P, data for a longer period of time are needed. From there, more complexity could be introduced in the model by adding a new variable that modifies M_a as function of the probability of being repeatedly requested.

In conclusion, despite the limitations, the model can nonetheless be satisfactory for those collections where the group of repeatedly requested records accounts for a relatively small percentage of the total of requests. To explore further we need individual agents.

5.5 Modelling access with agent based modelling

5.5.1 Model description

Like the SD model, the aim of the ABM model is to model the decrease in requests in the reading room since the digitisation of the collection began. Whereas we made the distinction between accessed and non-accessed records in SD, with ABM, we explore the importance of the 'digitisation match' (see Fig. 5.7) by including the number of times that individual records have been accessed in the model. Rather than using stocks representing a group of records, we can describe the individual characteristics of the archival records in ABM. In this case, we can consider both whether records have been accessed and how they accumulate instances of access through time.

The first step of the modelling process consists of creating a population of agents.

In this model, the population is the whole archival collection, while agents are the archival records. Different variables characterise the agents (Table 5.7). Agents are characterised by a Boolean variable (random TRUE (p), where p is the probability of `true`) that defines whether an agent has been requested (`requested`). A second Boolean variable (`popular`) states whether the agent is repeatedly requested. Whereas the instances of use for agents that are requested, but not popular, will equal one, the instances of access will accumulate for those that are popular as long as the object is not digitally available. The annual instances and the accumulation of instances of access are also variables that characterise the agents (`requestsYear`, `requestsTotal`). In addition, a statechart is created with two states: `inUse` and `digitallyAvailable`. By default the agents are in state `inUse`, and the transition to the next state (`digitallyAvailable`) occurs when the Boolean variable `digitised` becomes true. When the agents become digitally available, then the accumulation of instances of access stops (`requestsTotal`).

TABLE 5.7: List of variables assigned to the agents in the ABM model.

Variable	**Type**
`requested`	Boolean
`popular`	Boolean
`requestsYear`	integer
`requestsTotal`	integer
`digitised`	Boolean

At initialisation, the agents representing the records requested once are set using a random Boolean function. During the run, the same random function will set new records as requested once. Only those agents with instances of access (`requestsTotal`) equal to 0 can be selected. The records that are repeatedly requested also increase during the run. For those records, the instances of access (`requestsYear`) are set using a probability distribution, and the accumulation of instances of access (`requestsTotal`) is updated.

At run time, the agents follow three steps that take place annually:

1. The number of agents that are requested for the first time is updated.

2. If agents are not digitally available, they accumulate instances of access (requests) according to the assigned probability distribution.

3. If agents are not digitally available, they can be selected to be digitised.

Annually, the number of agents that are digitised is chosen according to the number of requests for digitisation. That happens in two steps:

1. The number of requests (R_s and R_b) is converted to the proportion of the collection, which is used to calculate the proportion of agents to be digitised.

2. The agents are then randomly chosen to be digitised if they are in state `inUse` and depending on the accumulated instances of access (`requestsTotal`).

Similarly to Equation 5.1 in the SD model, the last step is the calculation of the decrease in requests in the reading room (R_r) according to:

$$\frac{\partial R_r}{\partial t} = -R_r M_d \tag{5.9}$$

but in ABM, R_r is not modified by the percentage of requested records that are digitised but by the accumulated requests for records that become `digitallyAvailable`. Therefore, M_d, 'digitisation match' (see Section 5.2.3.3), is defined as follows:

$$M_d = \frac{\sum_{kd=m}^{n} \texttt{requestsTotal}_{kd}}{\sum_{k=m}^{n} \texttt{requestsTotal}_{k}} \tag{5.10}$$

where k are the agents requested at least once in the reading room (`requested = true`), and, from this group, kd are those agents that have become `digitallyAvailable` in that year.

The model was built using the software AnyLogic 8.6.

5.5.2 Experiments

5.5.2.1 Experiments description

Like the SD experiments, the ABM approach models the decrease in requests (R_r) from 2006 to 2018 using the SAA data set as input values. However, compared to the transparency and straightforwardness of the SD model, the ABM requires more input values in several steps to define the collection, as shown by the following description of the experiment:

1. The collection size is 1,000,000 archival records.

2. The number of agents is 4,000, except for the experiments where the importance of the number of agents is tested. In those experiments, we run the same experiment with a different number of agents each time: 1,000, 2,000, 4,000 and 6,000.

3. The percentage of the collection that has been requested increases every year. At the end of the run, 15% of the collection (agents) has been requested, and ca. 55% of the requested agents has been requested only once (`requestsTotal = 1`).
   The probability of an agent being accessed once is assigned with a binomial random distribution. At initialisation, the probability parameter, p, is 0.009, equivalent to an average of 378 agents (standard deviation 4.9) in ten repeated runs. During the run, the probability parameter is changed to p of 0.002. These two probabilities are chosen to obtain the number of records requested as found in the SAA data set before 2006 (first year of the run) and in 2018 (last year of the run). Only agents with no instances of access (`requestsTotal = 0 `) can be selected.
   The agents that will accumulate more than one instance of access (`popular = TRUE`) are selected with a probability of 0.05 at initialisation. During the run, the increase of this group is 0.2% of the whole population per year.

4. We model the initial requests of the popular agents, those accumulating access requests. The best fit to the data was obtained with an exponential distribution with a rate of $\lambda = 0.3$ at initiation. We only apply this distribution to accessed items, which means that the minimum value is 1.
   The data show that the popularity of items changes every year. However, the accumulation of instances of access tends to be gradual rather than sudden. There are many possible ways to model this. One method that offers a good fit to the data is to consider that the number of accesses of each item (`requestsYear`) remains the same or increases by a randomly distributed amount per year. The best fit is obtained with a normal distribution of $mean = -1$ and standard deviation $sigma = 1$ to model the annual instances of access. Because the accumulation of instances of access (`requestsTotal`) cannot decrease, only positive values are added to the accumulation of instances of access (`requestsTotal`). Investigating the dynamics of the popularity change, however, was not an objective of this research.

5. The annual number of agents digitised within the SoD and the block digitisation programme (R_s and R_b) and the match percentage of SoD and block

digitisation (M_s and M_b) are based on the SAA data set.

6. Agents are selected for digitisation if they are in state `InUse` and depending on the instances of use (`requestsTotal` $= 1$ and `requestsTotal` > 1). The proportion of these two groups is based on the SAA data.

Despite the effort to reproduce the use of the collections as closely to the actual data as possible, some level of information is lost in the ABM model compared to the actual data. In this experiment with 4000 agents, one agent is equivalent to 250 archival records. The conversion of input values to agents will lead to some loss of information (e.g. the last part of the long tail in Figure 5.2) or when the input value (e.g. digitised records) is smaller than 250 archival records. In addition, as explained in point 4, the cumulative frequency distribution of requests seen in Figure 5.2 was not obtained with a yearly Pareto distribution seen in the actual data, but an initial exponential frequency distribution annually modified by a normal distribution that gave the best fit.

Therefore, to test whether we were reproducing the same data of the SAA, two output values were used during the experiments as a control: the percentage of records requested and digitised and the instances of access digitised (21% and 31% respectively, according to the SAA data).

5.5.2.2 Results and discussion

In the first experiment, we explored the effect of the number of agents on the variability of the output in repeated runs. To explore this aspect, each computational experiment was repeated 10 times. Figure 5.23 shows that a higher variability can be expected by a lower number of agents, but from a certain number of agents, 4,000 in this case, the expected variability remains similar. It is also worth noting that only the models containing more than 4,000 agents correctly reproduced the expected outcome of the number of access requests in the reading room.

Figure 5.24 shows the output of the 10 runs with 4,000 agents. At the end of the run, the largest value within the range of the deviations from the mean is $\pm 4\%$. As typical in ABM, there are several elements with stochastic behaviour in this model, mostly due to the use of probability distributions. For example, at the beginning of the run, the initial number of agents requested once is selected according to a binomial distribution, and the probability parameter is p of 0.04. With a population of 4,000 agents, 160 agents should be selected. When the same experiment is repeated, we observe that, due to the low value of the probability,

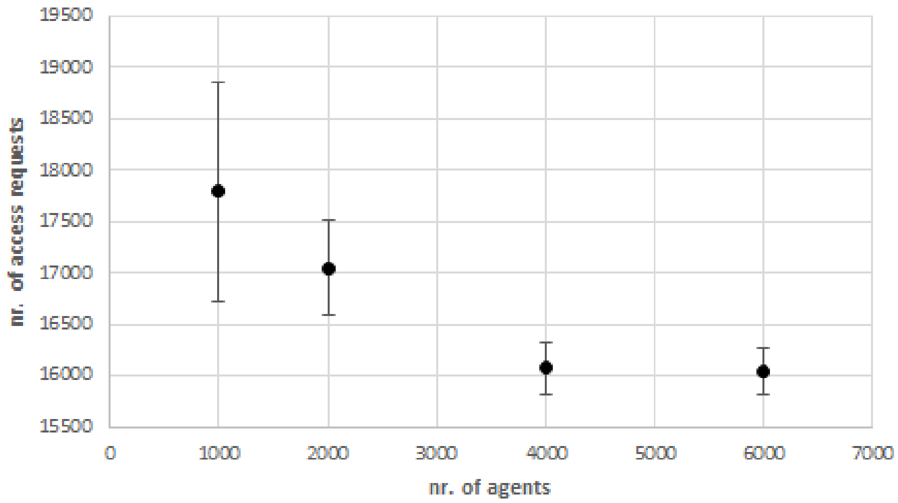


FIGURE 5.23: Comparison of the ABM output (number of access requests in the reading room (R_r) at the end of the run) depending on the number of agents. Error bars indicate the standard deviation.

the number of initial agents requested once varies from 113 (0.028) to 183 (0.045). Similarly, the instances of use for popular agents are set using probability distributions that are updated annually. At the end of repeated runs, the sum of `requestsTotal` varies from 1529 to 1714. In addition, the agents that meet certain criteria are selected randomly for digitisation. The percentage of digitised requests varies from 29.0% to 33.5%. Therefore, multiple runs are desirable.

Figure 5.25 shows the mean value of the output of the repeated runs. The ABM experiments confirm that the accumulated instances of access of the frequently requested records are an important element to reproduce the results of the reference data set. Whereas the curve seen in the SD model failed to reproduce the shape of the SAA data set through the years, the ABM model seems to succeed in reproducing both the mild decrease in requests at the beginning of the run and the sharper decline in requests from 2015.

It is important to notice that the model distinguishes between those agents requested once and those requested more than once. Choosing the right proportion of agents within each of these two groups is key to obtaining accurate results. The proportion of the agents requested once is based on the data analysis of the SAA data set. However, in the model, the selection of the agents requested more than once occurs randomly, and no further criteria of the instances of access (`requestsTotal`) are applied.

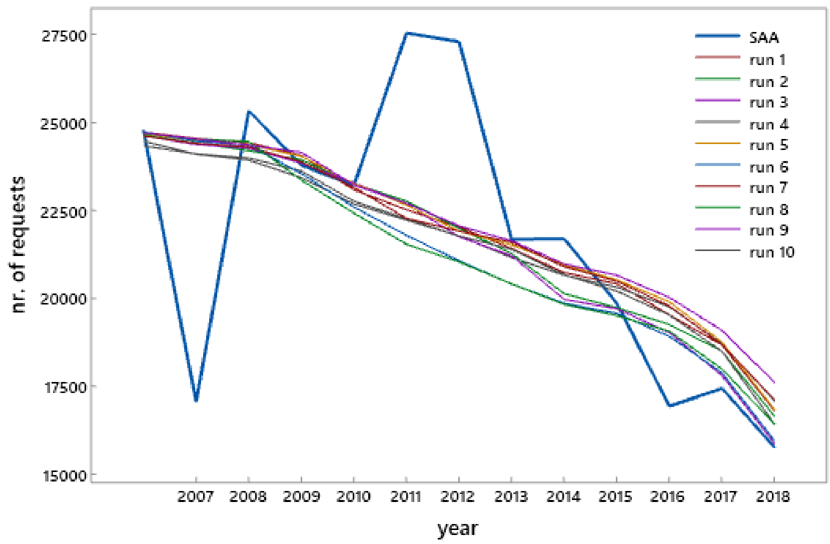


FIGURE 5.24: Comparison of the decrease in access requests according to the SAA data set and the output output of 10 repeated runs of the ABM simulation with 4,000 agents.

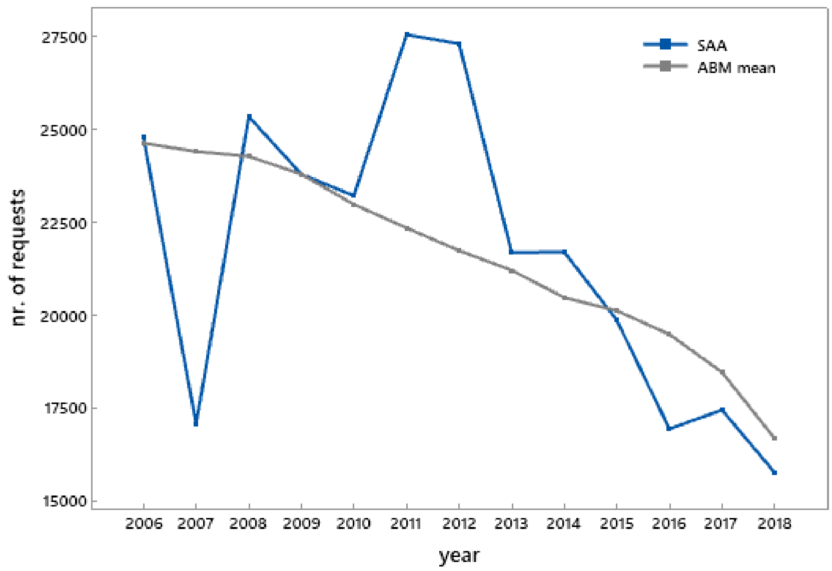


FIGURE 5.25: Annual number of requests in the reading room according to ABM simulation showing the mean value of the repeated experiment shown in Figure 5.24 and the SAA data as a reference.

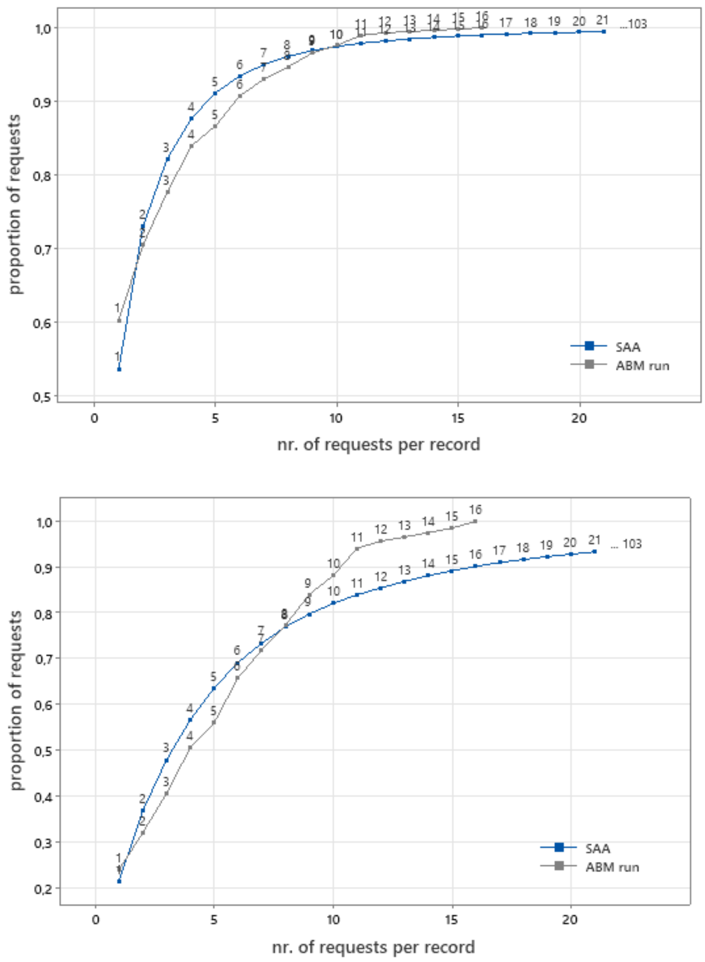


FIGURE 5.26: Comparison of the ABM simulation output (Fig. 5.24) and the SAA data set, regarding the number of records/agents requested (*top*) and the total number of requests in both cases (*bottom*).

Every year, the instances of use are chosen for the popular agents according to a probability distribution. In the case of the SAA data set, the annual distribution and the accumulated distribution after 20 years is a Pareto distribution. The Pareto distribution was obtained at the end of the experiment as a result of an exponential distribution at initialisation and an annual normal distribution. During the run, the agents accumulated only one or two instances of use per year to obtain a relatively high percentage of agents requested twice. As shown in Figure 5.26, the modelling of the accessed records and the accumulation of requests are in line with the SAA data set.

At the end of the period, the fraction of accumulated requests that become available in the SAA data set, M_r (Eq. 5.10), equals 0.319. In the repeated experiments, this value is between 0.255 and 0.319 due to the stochastic elements. The closer

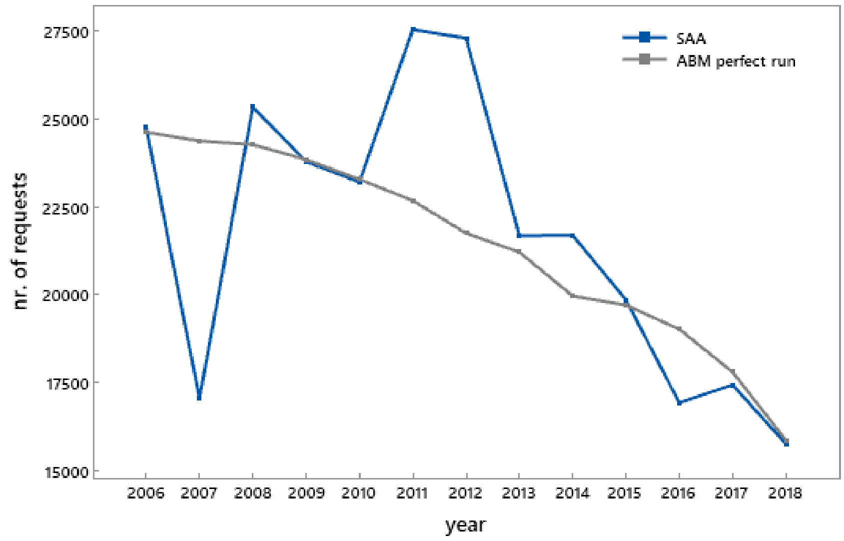


FIGURE 5.27: Perfect fit of the decrease in requests in an ABM simulation with the same percentage of M_r at the end of the run as in the SAA data set.

the fraction value in the experiment is to the value in the SAA data set, the better fit to the actual data, as in the case of the run shown in Figure 5.27.

As any distribution can be introduced in the model, we explored the impact of modifying the frequency distribution to model the accumulation of requests through the years in the following example. That is, we used an exponential distribution at initialisation and during the run. Figure 5.28, *left*, shows that the obtained distribution does not resemble the SAA distribution in that case.

Nevertheless, and quite unexpectedly, when this exponential distribution is used in the model, then a perfect fit with the reference data set can still be obtained (Fig. 5.28, *right*). The difference is that the perfect fit is obtained if the M_r value at the end of the run is between 0.27 and 0.28, instead of 0.319, as in the SAA data set.

These computational experiments confirm that the accumulated requests for the popular records are an important element to reproduce the results of the reference data set. Having the right proportion of popular records is essential to obtain accurate results.

Hence, in the next experiments, we investigated whether similar results could

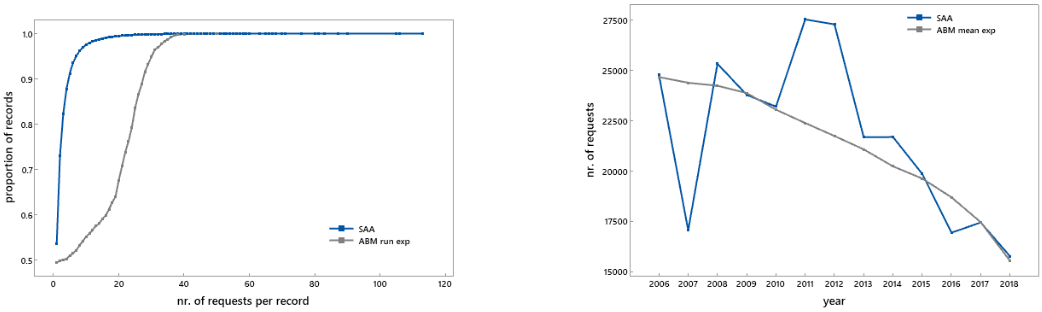


FIGURE 5.28: (*Left*) Comparison of the cumulative distribution (proportion) of the numbers of records requested in the SAA reading room and the experiment with exponential distribution for the number of requests per year. (*Right*) Comparison of the ABM experiment and the SAA data showing the number of requests in the reading room.

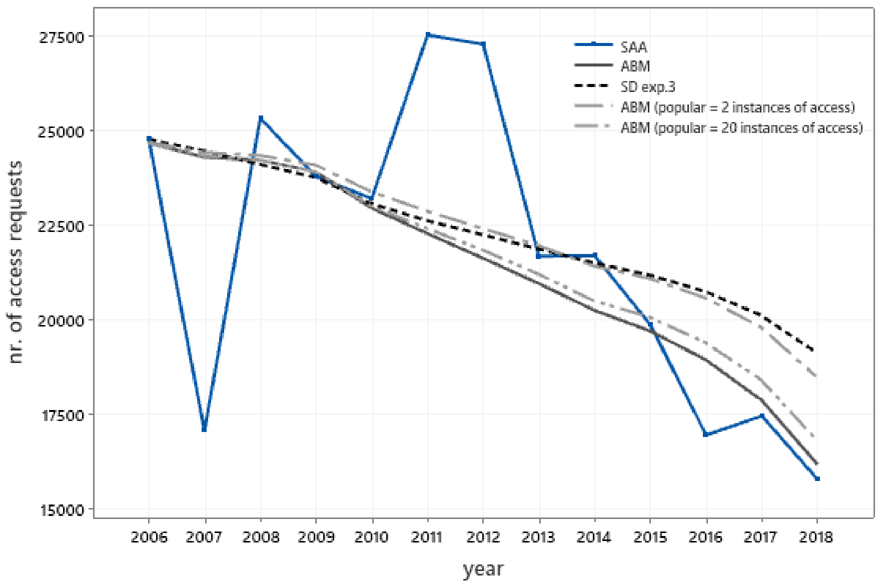


FIGURE 5.29: Annual number of requests in the reading room according to experiments using different distributions for the number of requests per year.

be obtained if, rather than modelling the number of access of each agent according to a distribution, a fixed value of the instances of access is assigned to the repeatedly accessed agents (`popular = TRUE`). This fixed value could be seen as an average of instances of access over time, and it would make the ABM model more simple, with no need for distributions. Two possibilities were tested: in one experiment, the agents accumulated two instances of access every year as long as they were not digitally available, and in the second experiment, they accumulated 20 instances of access. Interestingly, Figure 5.29 shows that when two instances of access are assigned, similar results are obtained to those seen in the SD model, where no distinction was made between accessed once and repeatedly accessed. However, when agents are assigned with 20 instances of access, then the output is close to the one where the instances of access were modelled using a frequency distribution. It seems that assigning 20 instances of access works as the average of the values resulting from the Pareto distribution, and as a result, a better fit is obtained.

5.5.2.3 Limitations of the model

One limitation of the model is that converting the input data to agents might lead to the loss of information. Let us analyse this issue in more detail by converting the distribution of records to agents according to the number of times they have been accessed during the studied period of time at the SAA, assuming that an agent is 250 records. We observe that the difference is mainly in the long tail of the frequency distribution of the requests that would be lost, specifically, records that have been requested between 15 and 113 times, which account for ca. 10% of all requests. Table 5.8 and Figure 5.30 show the conversion to agents would result in the loss of the tail and that the contribution of the requests of each category (e.g. records requested, once, twice, ..., 15 times) to the total of requests will be different than the reference data set. However, this difference seems to have little effect on obtaining satisfactory results.

The ABM allows us to build a model with a great level of detail, which appears to lead to a model closer to the system that we are trying to reproduce. More detailed input data are required to develop this level of detail compared to the SD approach. During the development of this ABM model, we could experience the greatest pitfall of the ABM approach, namely the loss of transparency. Due to the large amount of detail in the data, it becomes very difficult to verify and analyse how outputs are generated. For example, it is not straightforward to understand why satisfactory results are obtained even when a distribution is used that does

TABLE 5.8: Conversion of number of records to agents in the ABM model.

instances use	nr. records	nr. agents	agents requestsTotal
1	96978	388	388
2	34984	117	234
3	16663	67	201
4	9914	40	160
5	6259	25	125
6	4163	17	102
7	2763	11	77
8	1998	8	64
9	1418	6	54
10	1040	4	40
11	767	3	33
12	569	2	24
13	494	2	24
14	400	2	14
15	325	1	15

not reproduce the actual data set.

These experiments have shown two main pitfalls of the model that are related to the opacity of the ABM approach due to:

- a clear risk of overfitting, as many combinations of parameters can lead to the same results (as seen when a good fit might be obtained with two different frequency distributions as input value);
- the many stochastic elements present in the model that make it difficult for the user to understand how values change during the run (therefore, creating a need for control variables).

Therefore, the proposed ABM model is for now not suitable to be used directly by end users at this stage and should be seen as a research tool.

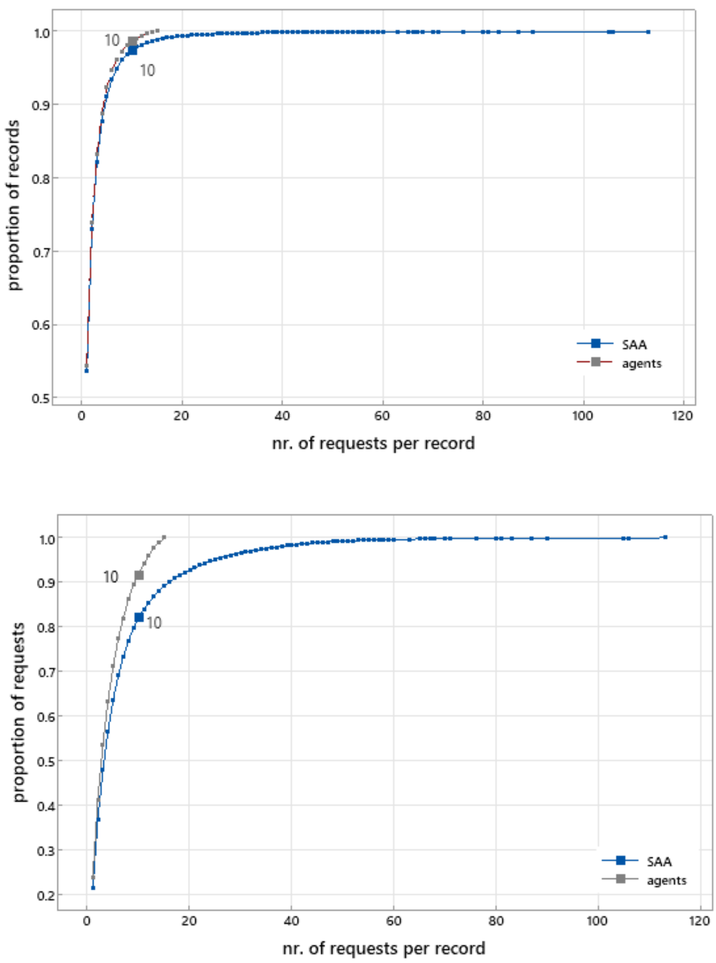


FIGURE 5.30: Comparison of the conversion of number to records to agents (*top*) and the number of requests in both cases (*bottom*).

5.6 Verifying and validating the simulation models

In the previous sections, the experiments were designed to developing understanding of how the model works, as well as whether the computerised model works as intended. These tests can be seen as part of the validation phase in the development of a model, and are essential to build credibility of the computerised model. The validation step, which is composed of verification and validation, is an essential step in the development of any simulation model. Whereas verification is formally described as 'ensuring that the computer program of the computerized model and its implementation are correct', validation relates to 'substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model' (Sargent; 2020, p. 16). Put it in a more simple way, '[m]odel verification deals with building the model right' (Balci; 1994, p. 123), whereas '[m]odel validation deals with building the right model' (Balci; 1994, p. 121).

A list as well as a taxonomy of traditional verification methods of the computerised model is given by Whitner and Balci (1989), Balci (1994), Roungas et al. (2017). Verification and validation can be conducted following an informal or formal approach. In this thesis no statistical methods (Sargent et al.; 2016; Kleijnen; 1995) have been used to validate the models, but several tests have been conducted as part of the verification and validation phase. Regarding the verification of the model, the following tests were conducted during the development of the computerised models (access and preservation sub-models):

- checking syntax analysis: AnyLogic supports the checking of errors found during code generation and compilation. During model execution, runtime errors are also given, like computational errors in expressions (e.g. division by zero) as well as simulation errors due to logical errors of model execution (e.g. a statechart is unable to exit a branch because all exiting transitions are closed);
- consistency checking for units (e.g. requests per year for the variables in the access part);
- verification of intermediate and final simulation output by means of manual calculation: several intermediate and final simulation results were calculated manually, and the results were compared with outputs of the simulation program (e.g. percentage of the collection becoming digitally available, or, in the next chapter, percentage of the collection which underwent a deacidification treatment);
- verification of intermediate simulation output by means of graphs: several outputs of the simulation program were monitored during the run using graphs (e.g. accumulation of requests (see Fig. 5.26);
- sensitivity analyses: the model was tested using extreme values (e.g. zero digitisation requests);
- comparison to other models: although it was not possible for the access part of the model, since this was the first attempt to develop a mathematical model on the topic, in the next chapter when preservation is modelled, the results of the simulation result will be compared with those of existing models.

Regarding the validation of the model, the traditional methods can be grouped in two main approaches: comparison of model output to data and exploration of the model behaviour.

- Comparison to historical data: the same input data is used to compare model behaviour with system behaviour, as it has been shown in the experiments in the previous sections;
- parameter variability-sensitivity analysis: although this type of analysis was not conducted in this part of the model, in the next chapter sensitivity analysis will be conducted to develop understanding of the preservation part of the model.

For the validation of the access part of the model a comparison of the simulation results with historical data is made, as this is the most approachable method that we have to validate the model. Therefore, in this section the data sets of the Noord-hollands Archief (NHA) and the Utrechtsarchief (UA) were used for further validation of the SD and ABM models developed in the previous sections. The data sets comprehend about 10 years of automated data on reader requests. Although these institutions have been using automated systems for a longer period to keep records of the reading room history, it is not unusual that when a new system is introduced, the older data are not converted to the new system.

5.6.1 Noord-hollands Archief data set

5.6.1.1 Data analysis

The data set contains the history of requests from 2008 to 2019. During this period, the number of requests in the reading room remained mostly stable (Fig. 5.31). However, from 2015, three of the four years show a slight decrease in requests, which might be seen as the beginning of a downward trend. Because the NHA started the digitisation of this collection in 2015, it might be tempting to assume that the slight decrease in accessed records in recent years is related to the beginning of digitisation. Based on the analysis and model results presented in the next section, we explore whether evidence can be found for the impact of digitisation on the number of requested records.

A first analysis of the NHA data set shows that this data set follows the characteristics seen in the SAA data set. For instance, Figure 5.32, *top*, shows the frequency distribution of instances of use at the NHA. In the period of eleven years, 73,968 records have been accessed in the reading room. Like the SAA, only 1%

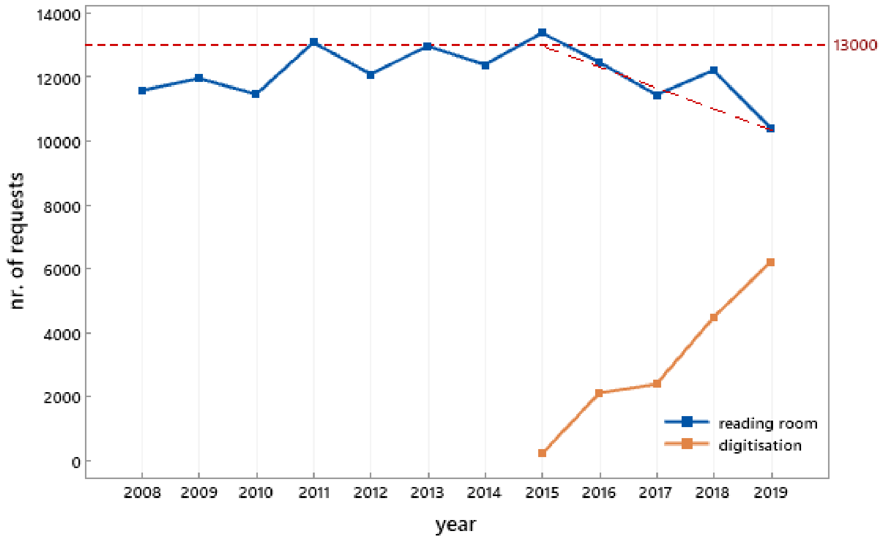


FIGURE 5.31: Number of requests in the NHA reading room between 2008 and 2019 and number of records that have been digitised annually since 2015. The dashed red line indicates the slight downward trend in the number of requests since 2015.

of these records have been requested on average more than once per year during this time. We observe the same Pareto distribution, but in this case, with a slightly higher percentage of records requested once and a shorter tail. These differences are probably due to the fact that this data set is the result of the accumulation of requests during 11 years, whereas the SAA data set shows the cumulative distribution of 20 years. The longer the period under study, the higher the chance that records requested once are accessed for a second time and that the frequently accessed records continue being accessed and accumulating requests.

Between 2015 and 2019, 15,490 records were digitised (5,696 records within SoD against 9,794 records within block digitisation). So far, 6% of the records accessed in the reading room has also been digitised (Fig. 5.33). This means that one-third of the digitised records has also been requested in the reading room, which is similar to the percentage in the SAA. Confirming what we observed in the SAA data set, SoD requests are more in line with the reader requests in the reading room: whereas 51% of the SoD request are records that have ever been accessed in the reading room, only 28% of block digitisation include records requested in the reading room. In addition, SoD makes the highest contribution to the 'digitisation match' (M_r) (Fig. 5.32, *bottom*).

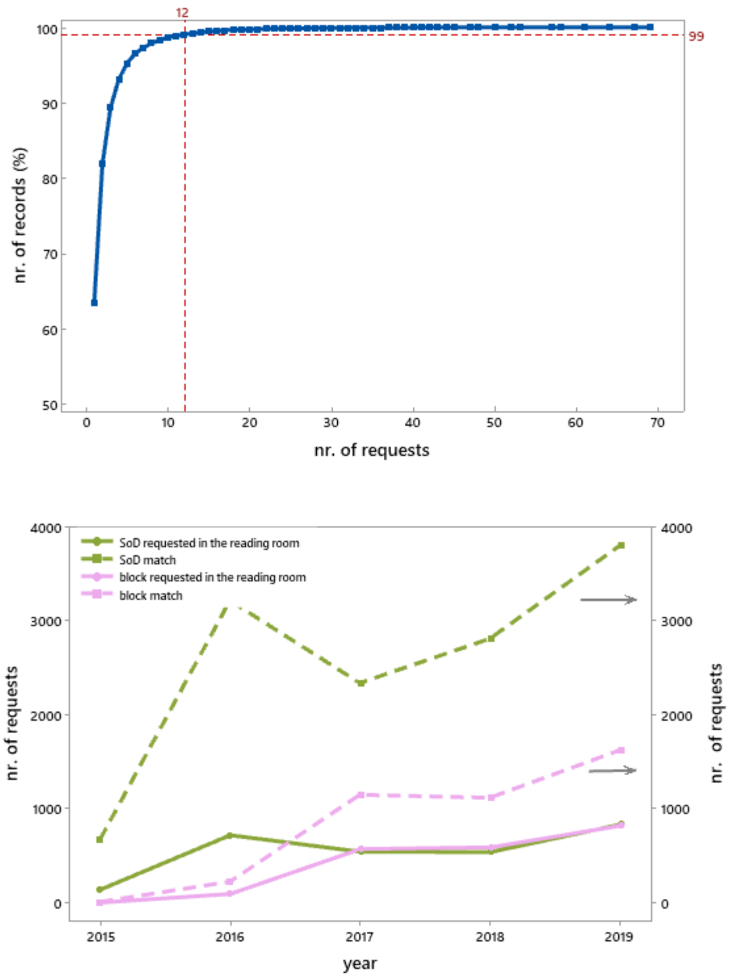


FIGURE 5.32: (*Top*) Cumulative distribution (percentage) of the numbers of records that have been requested in the NHA reading room during a period of 11 years. (*Bottom*) Number of records requested in the reading room that have been digitised (solid lines), and sum of requests history of this group (accessed and digitised) (dashed lines), according to the two digitisation programs.

5.6.1.2 Experiments

The aim of the experiment is to explore whether the slight decrease observed in the number of requests is related to digitisation by comparing the results of the two different approaches, SD and ABM. The two models are built using the NHA data set as input data. The output that we are interested in is the annual number of requests (R_r).However, in the case of ABM, it is important to evaluate whether the model is reproducing other output values correctly. One example of a control output is the frequency distribution of the records obtained at the end of the run compared to the frequency distribution of the actual data set (Fig. 5.32, *top*). Other control outputs are summarised in Table 5.9. These control outputs are used to

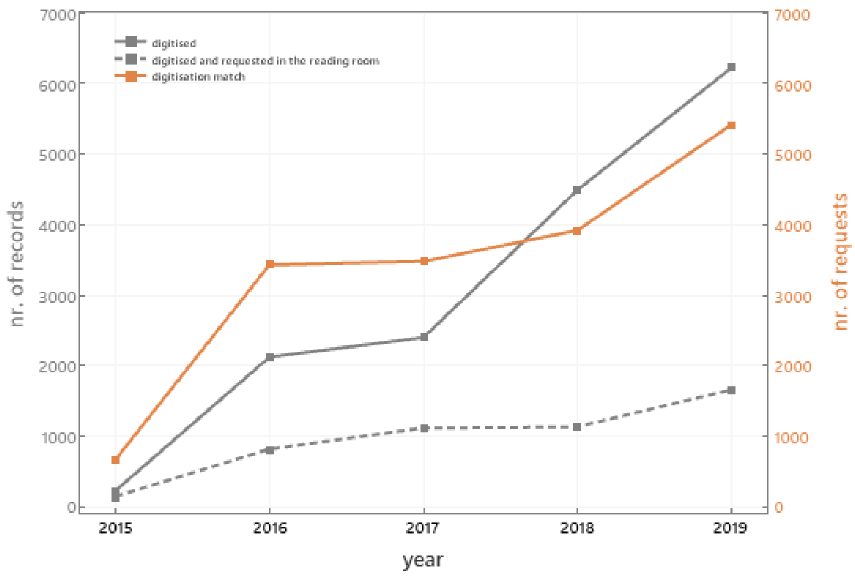


FIGURE 5.33: Number of records digitised per year (grey line); number of records requested in the reading room that have been digitised (grey dashed line); and sum of requests history of this group (accessed and digitised), referred as 'digitisation match' (orange line), at the NHA.

calibrate the model. The main input used to calibrate the model are the fractions of records requested for the first time (P) and frequently requested (`popular`). The model is further calibrated by adjusting the number of records per agent until the values of the control output are met.

The decrease in requests is calculated from a certain year, in this case, 2016. Although digitisation started at the NH in 2015, only 230 records were digitised in that year, whereas in subsequent years, the production was more than 2000 records digitised per year. In addition, the number of requests in 2016 seems to be closer to the average requests in previous years, whereas in 2015, the number of requests was exceptionally high, and 2015 could be considered an outlier (Fig. 5.31). Therefore, 2016 is used as reference point to compare the decrease

TABLE 5.9: Comparison of control outputs between the NHA data set and the outputs of the ABM model.

Output	**NHA data set**	**Experiment**
Records requested and digitised	6.5%	7.0%
Requests digitised	12%	11%

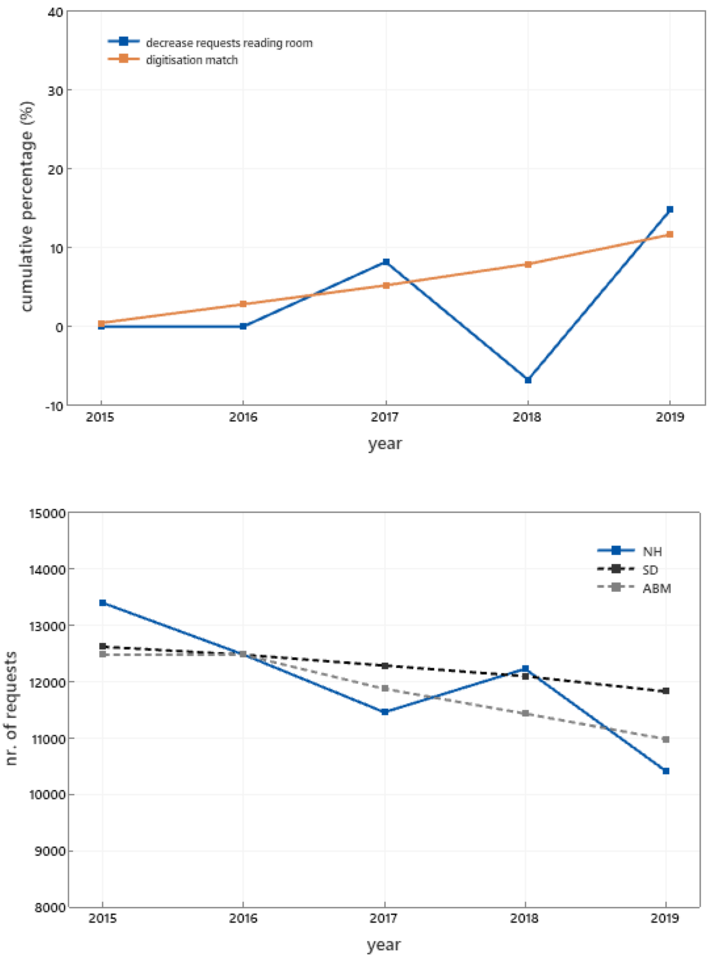


FIGURE 5.34: (*Top*) 'Digitisation match' given as a cumulative percentage (orange line) and percentage decrease in requests compared to 2016 at the NHA. (*Bottom*) Comparison of SD and ABM experiments and the NHA data showing the number of requests in the reading room.

in requests with the 'digitisation match' (Fig. 5.34, *top*) and number of requests at initialisation of the experiments (Fig. 5.34, *bottom*).

Although the data points are quite limited, similar to the SAA experiment, the resulting R_r in the ABM model seems to be closer to the actual data set than the resulting R_r in the SD model (Fig. 5.34, *bottom*). This result confirms the importance of introducing the number of requests per record in the analysis. In addition, the fit between experiments and actual data supports the hypothesis that the slight decrease in annual requests in the reading room since 2016 might be the result of providing digital access to the collections.

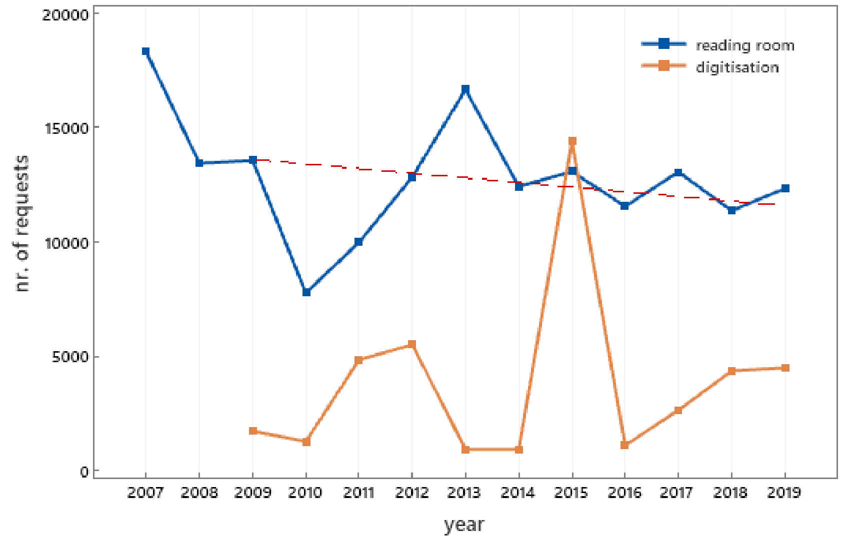


FIGURE 5.35: Number of requests in the UA reading room between 2007 and 2019 and number of records that have been digitised annually since 2009. The dashed red line indicates the slight downward trend in the number of requests since 2015.

5.6.2 Utrechtsarchief data set

5.6.2.1 Data analysis

The UA data set includes data from 2007 to 2019. After several digitisation pilot projects, the UA started to digitise their collections in 2009. In a period of ten years, 42,296 records have been digitised (Fig. 5.35). During this period, the general trend of the number of requests in the reading room has been slightly downward but less pronounced than the decrease seen in the NHA. This is an interesting observation because the UA has been digitising more records and for a longer period than the NHA. Figure 5.36 shows that large digitisation projects have been conducted, but with a low 'digitisation match'.

The frequency distribution for the accumulated number of requested records shows a similar Pareto distribution as the one seen in the SAA and the NHA, where more than the half of the records (65% of the requested records) have been accessed only once during the study period (Fig. 5.37, *top*). However, this percentage is slighter higher than the percentage of records accessed only once in the NHA. Furthermore, the tail of the UA curve is shorter, and 99% of the records have been requested less than 0.8 times per year, which is slightly less than the 1.1 seen at the NHA. These statistics indicate that the UA collections tend to be

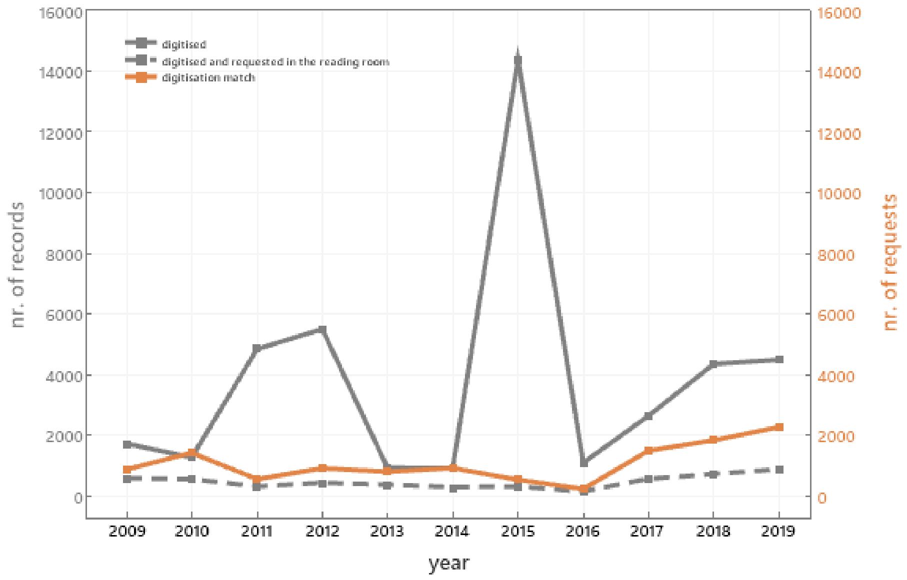
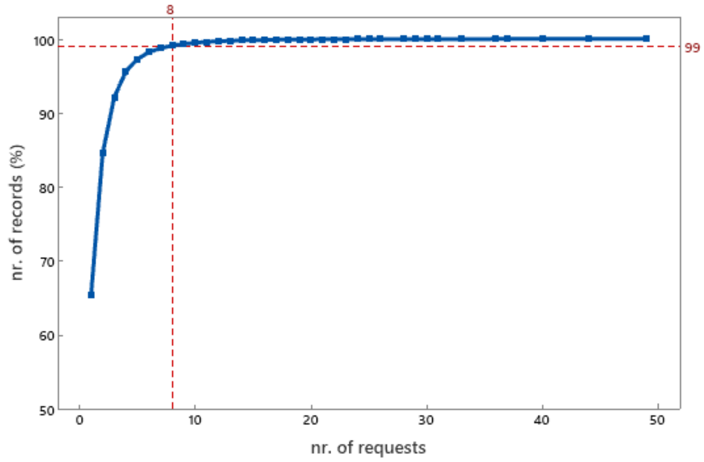


FIGURE 5.36: Number of records digitised per year (grey line), number of records requested in the reading room that have been digitised (grey dashed line), and sum of request history of this group (accessed and digitised), referred as 'digitisation match' (orange line) at the UA.

TABLE 5.10: Statistics on the use of the collections at the NHA and the UA.

Variable	**NHA**	**UA**
No. records requested in reading room(total)	73,968	96,542
No. records digitised (total)	15,490	42,296
No. SoD records within records digitised	5,696 (36%)	5,218 (12%)
No. records digitised within records requested	4,856 (6.5%)	4,845 (5.0%)
No. reading room requests (total)	145,495	166,268
No. digitised requests within reading room requests	16,936 (11.5%)	12,007 (7.2%)
Match SoD (records)	51%	38%
Match Block (records)	28%	17%



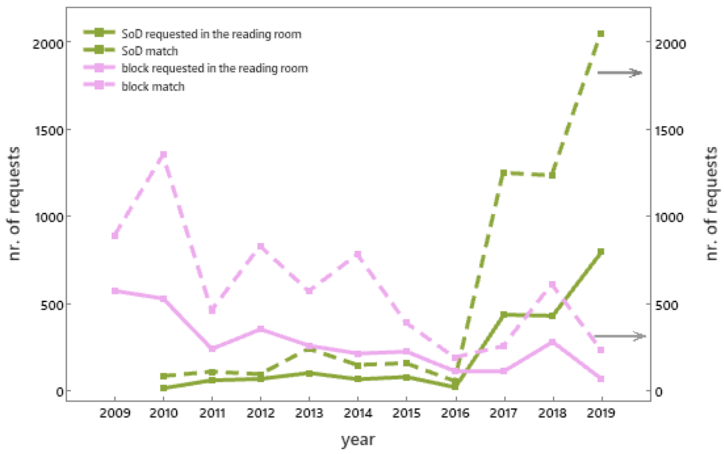


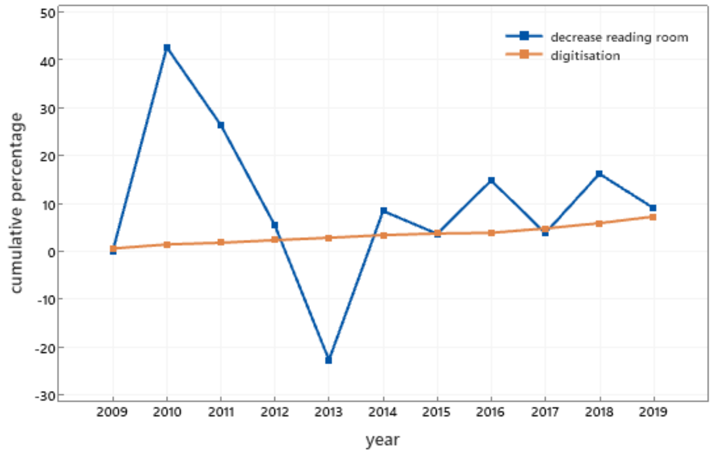
FIGURE 5.37: (*Top*) Cumulative percentage of the numbers of records that have been requested in the UA reading room during a period of 12 years. (*Bottom*) Number of records requested in the reading room that have been digitised (lines), and sum of request history of this group (accessed and digitised) (dashed lines), according to the two digitisation programs at the UA.

slightly less repeatedly accessed than the NHA. Further statistics of the NHA and the UA are summarised in Table 5.10. The statistics on the use of the collections at the two institutions show that the NHA has a higher percentage of digitised records requested in the reading room. The NHA has also been digitising records that were frequently accessed in the reading room. Part of this higher percentage can be explained by the higher number of records digitised within the SoD programme at the NHA. Both data sets confirm the observation made in the SAA data set, where SoD is better targeting the records that have been requested in the reading room. In the case of the SAA and the NHA, this match (M_a) is constant throughout the period. The UA data show a more irregular match (Fig. 5.37, *bottom*), but even so, the average match of the SoD (M_s) is higher than the block match (M_b). However, this lower percentage of M_s and M_b observed at the UA might be the result of the relatively low percentage of frequently requested records at the UA compared to the NHA: collections with a well-identified group of frequently requested records might expect a higher impact of digitisation than collections with a less well-defined group of popular items.

Regardless of this observation, the difference in the M_r between the two archives is clear. Let us take the values in 2017 as an example. Both data sets have been accumulating requests for almost 10 years, and approximately 2500 records have been digitised in both archives on that year. Whereas at the NHA, 1130 of the records have also been accessed in the reading room, only 553 records of the UA belong to this group. Moreover, whereas the group of requested and digitised is 1570 requests at the UA, it is 3923 requests at the NHA.

5.6.2.2 Experiments

To build the SD and ABM models, we follow the same approach as in the examples of the SAA and the NHA. Figure 5.38, *bottom*, confirms that the results of ABM are slightly closer to the actual data set than the SD model. By including the number of requests by records in the ABM model, the model reproduce the 'match percentage' (M_r) in the actual data set (Fig. 5.38, *top*). As seen in the case of the NHA, the results of the ABM model seem to indicate that the slight decrease in the number of reading room requests can be explained by the digitisation of the collections.



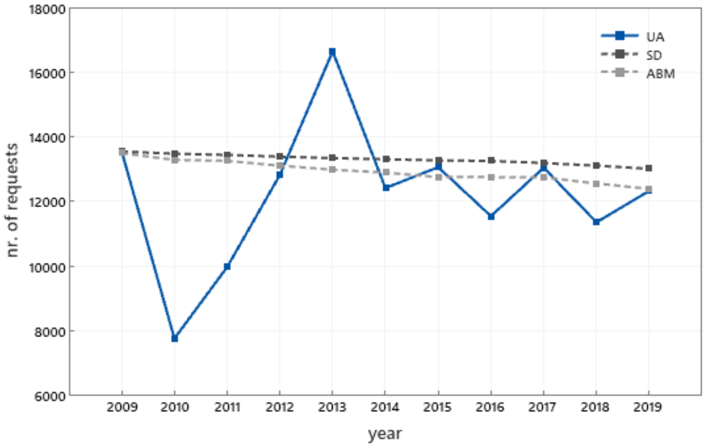


FIGURE 5.38: (*Top*) 'Digitisation match' given as a cumulative percentage (orange line) and percentage decrease in requests compared to 2009 at the UA. (*Bottom*) Comparison of SD and ABM experiment and the UA data showing the number of requests in the reading room.

5.7 Conclusions

In this chapter, we proposed two modelling approaches, SD and ABM, to investigate the effect of digitisation on the level of demand for archival documents in the reading rooms of institutions. Based on the data analysis conducted on the actual usage data of the SAA, we identified the main hypothesis that might explain the observed decrease in the number of access requests, namely the impact of digitisation projects. This analysis formed the basis of the development of the proposed simulation models.

The study of the history of accesses data set showed that data analysis is a crucial source of information as both input data for the model and to inform institutions about how the collections are used so they can understand the impact of digitisation strategies. As expected, the longer the period of data collection, the clearer the trends over time, particularly if the data were kept before and after digitisation programmes started. Data sets with a period of more than 10 years were rarely found during this study. Nevertheless, 10 years of data were sufficient to identify trends and to be used as input data for the model.

The data analysis showed evidence to support the hypothesis that the digitisation of the original collections will eventually result in a decrease in access requests in the reading room. How quickly digital access will overtake the traditional physical access depends on how digitisation strategies target those archival records that have been (frequently) requested by the readers. In this regard, we found that SoD performs better than the projects initiated by the institution where subsets of whole collections are selected for digitisation.

The relationship between digitisation of the collections and the use of the original collections was further investigated using simulation. The SD approach showed that making those archival documents digitally available that had not been requested before has an almost negligible effect on the observed decrease in the access requests in the simulated period of almost 15 years. The observed decrease in access requests can be mostly explained by the increase in the requested-at-least-once records becoming digitally available. However, the results of the simulation indicated that the results of the SD models slightly underestimated the decrease in access requests in the reading room. The SD model does not consider the popularity of the archival records, understood as the number of access requests per record. Therefore, it can be expected that the SD approach will produce more accurate results for those collections where the group of repeatedly requested records accounts for a relatively small percentage of the total requests.

More accurate results of the decrease in access requests were obtained when the agent-based simulation approach was applied. In this model, agents represented archival documents with their own characteristics, for instance, the number of access requests per year. The results of the ABM confirmed that the digitisation of frequently requested records will have a more noticeable impact on the reduction of access requests than records that have only been requested once in the past. A frequency distribution was used to model the instances of access. However, further investigation showed that it is not essential to reproduce the Pareto distribution of the number of access requests observed in the actual collection to obtain satisfactory results. The ABM experiments pointed out that differentiating between groups of records requested once and groups of records with a high level of demand is also sufficient to model the use of the collections in the reading room. In view of these results, we expect that if the differentiation between these two is included in the SD model, the accuracy of the models could be improved because the level of detail of ABM seems not to be essential.

A clear disadvantage of the ABM is that a great level of detail is needed to model the agents individually, notably increasing the complexity of the model, resembling a black box. In contrast, the SD model is transparent, resulting in an approach that is more intuitive to build and to use.

In the two presented models, the number of requests is only affected by digitisation. In such a case, we are assuming that other factors remain unchanged, such as the preference of the readers to use SoD instead of visiting the reading room. Modelling the behaviour of the readers is a crucial part of the model that has been omitted for now, but it should be taken into account in further research.

The hypotheses formulated in the first part of this chapter were tested by analysing the data of two additional archival institutions. The data of the case studies supported the formulated hypotheses, but as in those institutions only a small decrease in requests has been observed in the last decade, more data is needed for further validation of the models.

Regarding the modelling of the collections, the level of detail (number of access requests per record) seems to be sufficient to obtain satisfactory results. Another layer of information is present in the real world, in addition to the records; records are grouped within archives or collections. This distinction in groups is not included in the model, and records are randomly selected for digitisation depending on whether they have been frequently requested or not, without making a further distinction. This micromodelling seems to be relevant only if we would be interested in learning more about the effect of digitisation of a certain archive or collection. A very extensive data analysis of each archive and collection is required to include this level of detail in the model, which would further increase the complexity of the model. This kind of analysis does not need to be included in the model, but it could be conducted as part of the data analysis of the data set.

This study is the first to use simulation approaches, SD and ABM, to model one of the main processes in archives and libraries, providing access to the collections. The proposed model informs archival institutions how digitisation is affecting traditional access to the collections. In Chapter 7, we will show that the model can be used to explore the impact of different digitisation strategies and to investigate when a decrease in access requests can be expected. This information will support the development of digitisation strategies that aim to enhance access to the collections.

This study has provided valuable information on the frequency of use of the collections. In the next chapter, we will link this information to the preservation of the collections because physical use of the collections leads to wear and tear of the most vulnerable records.

References

Andro, M. (2018). *Digital libraries and crowd sourcing*, John Wiley & Sons, Hoboken, NJ, USA.

ARA (2019). Survey of Visitors to UK Archives 2018, *Technical report*, ARA National Surveys Group.
URL: *https://www.archives.org.uk/images/Public_Services_Quality_Grp/Survey_of_Visitors_to_UK_Archives_2018_-_National_Headline_Report_.pdf*

ARA (n.d.). Archive Surveys.
URL: *https://www.archives.org.uk/what-we-do/archive-surveys.html*

Astle, P. J. and Muir, A. (2002). Digitization and preservation in public libraries and archives, *Journal of Librarianship and Information Science* **34**(2): 67–76.

Ayris, P., Wheatley, P., Aitken, B., Hole, B., McCann, P., Peach, C. and Lin, L. (2010). The LIFE3 Project: Bringing digital preservation to LIFE, *Technical report*.
URL: *http://www.life.ac.uk/3/docs/life3_report.pdf*

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study, *Annals of Operations Research* **53**: 121–173.

Bantin, J. and Agne, L. (2010). Digitizing for value: A user-based strategy for university archives, *Journal of Archival Organization* **8**(3-4): 244–250.

Brancolini, K. R. (2000). Selecting research collections for digitization: Applying the Harvard model, *Library Trends* **48**(4): 783–799.

Brown, L. (2003). Useful or useless use statistics? A summary of conference presentations on usage data from the 22nd Annual Charleston Conference, Issues in Book and Serial Acquisition, *Serials Review* **29**(2): 145–150.

Campagnolo, A. (ed.) (2020). *Book conservation and digitization: The challenges of dialogue and collaboration*, ARC Humanities Press.

Chapman, J. and Yakel, E. (2012). Data-driven management and interoperable metrics for special collections and archives user services, *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* **13**(2): 129–151.

COUNTER (2021). COUNTER Code of Practice Release 5.0.2.
URL: *https://cop5.projectcounter.org/en/5.0.2/*

Coyle, K. (2006). Mass digitization of books, *Journal of Academic Librarianship* **32**(6): 641–645.

Dalgleish, P. (2011). The thorniest area: making collections accessible online while respecting individual and community sensitivities, *Archives and Manuscripts* **39**(1): 67–84.

Dobreva, M., O'Dwyer, A. and Konstantelos, L. (2011). User needs in digitization, *in* L. Hugues (ed.), *Evaluating and Measuring the Value, use and Impact of Digital Collections*, Facet, chapter 6.

Duff, W. M., Yakel, E., Tibbo, H. R., Cherry, J. M., Mckay, A., Krause, M. G. and Sheffield, R. (2010). The development, testing, and evaluation of the archival metrics toolkits, *The American Archivist* **73**: 569–599.

Erway, R. and Schaffner, J. (2017). *Shifting gears. Gearing up to get into the flow*, 2nd edn, OCLC Research, Dublin, OH.

European Commission (n.d.). Shaping Europe's digital future. Digital cultural heritage.
URL: *https://digital-strategy.ec.europa.eu/en/policies/cultural-heritage*

Hirtle, P. B. (2002). The impact of digitization on special collections in libraries, *Libraries & Culture* **27**(1).

Holtman, M. (2017). Digitisation based on industrial principles. A new vision for the quality management of mass digitisation projects, *Technical report*, Amsterdam City Archives, Amsterdam.

Hughes, L. (2003). Why digitize? The costs and benefits of digitization, *in* L. Hughes (ed.), *Digitizing Collections. Strategic Issues for the Information Manager*, Facet Publishing, London, chapter 1.

Hughes, M. (2016). A long-term study of collection use based on detailed Library of Congress classification. A statistical tool for collection management decisions, *Collection Management* **41**(3): 152–167.

IFLA (2002). Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives, *Technical report*, International Federation of Library Associations and Institutions (IFLA) and International Council on Archives (ICA).
URL: *https://www.ifla.org/files/assets/preservation-and-conservation /publications/digitization-projects-guidelines.pdf*

Jeurgens, C. (2013). The scent of the digital archive: Dilemmas with archive digitisation, *BMGN - Low Countries Historical Review* **128**(4): 30–54.

Jones, D. (2008). Mass digitization of historical records, *Serials* **21**(2).

Kelly, E. J. (2014). Assessment of digitized library and archives materials: A literature eview, *Journal of Web Librarianship* **8**(4): 384–403.

Kemp, J. (2016). How digitisation integrates in the world of archives preservation, *Journal of the Institute of Conservation* **39**(1): 57–63.

Kleijnen, J. P. (1995). Verification and validation of simulations models, *European Journal of Operational Research* **82**: 145–162.

Lejeune, B. (2007). The effects of online catalogues in London and other museums: A Study of an alternative way of access, *Papers from the Institute of Archaeology* **1**: 79–97.

Lindsay, H. (2003). Preservation microfilming and digitization at London Metropolitan Archives: Surveying and conservation preparation prior to image capture, *Paper Conservator: Journal of the Institute of Paper Conservation* **27**: 47–57.

Ling, T. and McLean, A. (2004). Taking it to the people: Why the National Archives of Australia embraced digitisation on demand, *Australian Academic and Research Libraries* **35**(1): 2–15.

Macpherson, P. (2010). Building a better horse and buggy: The privileging of access in reading rooms over online access, *Archives and Manuscripts* **38**(2): 61–78.

Maroso, A. L. (2003). Digitization as a preservation practice: Look before you leap, *Serials Librarian* **45**(2): 27–37.

Martell, C. (2008). The absent user. Physical use of academic library collections and services continues to decline 1995-2006, *Journal of Academic Librarianship* **34**(5): 400–407.

Mcavena, K. (2017). Information seeking: Taking a closer look at archival user studies, *Master thesis*, Universiteit Leiden.

McCausland, S. (2011). A future without mediation? Online access, archivists, and the future of archival research, *Australian Academic and Research Libraries* **42**(4): 309–319.

Meyer, E. T., Eccles, K., Thelwall, M. and Madsen, C. (2009). Usage and impact study of JISC-funded. Phase 1 Digitisation projects & the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR), *Technical report*, University of Oxford.
URL: *http://repository.jisc.ac.uk/6997/1/TIDSR_FinalReport_20July2009.pdf*

Miller, L. K. (2013). All text considered: A perspective on mass digitizing and archival processing, *Source: The American Archivist* **76**(2): 521–541.

Mills, A. (2015). User impact on selection, digitization, and the development of digital special collections, *New Review of Academic Librarianship* **21**(2): 160–169.

Morrisey, L. (2010). Data-driven decision making in electronic collection development, *Journal of Library Administration* **50**(3): 283–290.

NARA (2018). 2018-2022 Strategic Plan, *Technical report*, National Archives and Records Administration (NARA).
URL: *https://www.archives.gov/files/about/plans-reports/strategic-plan/2018/strategic-plan-2018-2022.pdf*

Nauta, G. J., van den Heuvel, W. and Teunisse, S. (2017). Europeana DSI 2-Access to digital resources of European heritage deliverable D4.4. Report on ENUMERATE Core Survey 4, *Technical report*, DEN Foundation.

O'Brien Miller, L. and Tedone, M. (2011). Archives conservation discussion group 2011. Digitization and its effect on conservation treatment decisions. How has wide-spread digitizing of collections changed our approach to treatment?, *The Book and Paper Group Annual* **30**: 115–127.

Oliver, G. (2011). The digital archive, *in* L. Hugues (ed.), *Evaluating and Measuring the Value, Use and Impact of Digital Collections*, Facet, chapter 4.

Ooghe, B. and Moreels, D. (2009). Analysing selection for digitisation: Current practices and common incentives Analysing Selection for Digitisation Current Practices and Common Incentives, *D-Lib Magazine* **15**(9/10).

Parkinson, N. (2005). Digitisation in the UK– the case for a UK Framework, *Technical report*, JISC.
URL: *www.jisc.ac.uk/digitisation*

Pickford, C. (2002). Archives: A statistical overview, *Cultural Trends* **12**(48): 1–36.

Poole, N. (2010). The cost of digitising Europe's cultural heritage, *Technical report*, Collection Trust.
URL: *https://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf*

Prescott, A. and Hughes, L. (2018). Why do we digitize? The case for slow digitization, *Archive Journal*.
URL: *https://www.archivejournal.net/essays/why-do-we-digitize-the-case-for-slow-digitization/#footnote_20_7962*

Rhee, H. L. (2015). Reflections on archival user studies, *Reference and User Services Quarterly* **54**(4): 29–42.

Richardson, G. P. (2011). Problems with causal-loop diagrams, *System Dynamics Review* **27**(3): 219–243.

Rikowski, R. (2011). Digital libraries and digitisation. An overvieuw and critique, *in* R. Rikowski (ed.), *Digitisation Perspectives*, Sense Publishers, Rotterdam, chapter 2, pp. 21–42.

Riley-Reid, T. D. (2015). The hidden cost of digitization – things to consider, *Collection Building* **34**(3): 89–93.

Roungas, B., Meijer, S. and Verbraeck, A. (2017). A framework for simulation validation and verification method selection, *in* A. Ramezani and M. Williams, E. Bauer (eds), *Proceedings of the 9th International Conference on Advances in System Simulation, SIMUL 2017*, pp. 35–40.

Sargent, R. G. (2020). Verification and validation of simulation models: an advanced tutorial, *Proceedings of the 2020 Winter Simulation Conference*, pp. 16–29.

Sargent, R. G., Goldsman, D. M. and Yaacoub, T. (2016). A tutorial on the operational validation of simulation models, *Proceedings of the 2016 Winter Simulation Conference*, pp. 163–177.

Schaffner, J., Snyder, F. and Supple, S. (2011). *Scan and deliver: Managing user-initiated digitization in special collections and archives*, OCLC Research, Dublin, OH.

Shepherd, E. (2009). *Archives and archivists in 20th century England*, Ashgate.

Tanner, S. and Deegan, M. (2011). Inspiring research, inspiring scholarship. The value and benefits of digitised resources for learning, teaching, research and enjoyment, *Technical report*, JISC and KDCL, London.

URL: *https://www.kdl.kcl.ac.uk/fileadmin/documents/Inspiring_Research_Inspiring_Scholarship_2011_SimonTanner.pdf*

Terras, M. M. (2011). The rise of digitization. An Overview, *in* R. Rikowski (ed.), *Digitisation Perspectives*, Sense Publishers, Rotterdam, chapter 1, pp. 3–20.

Thylstrup Bonde, N. (2018). *The Politics of mass digitization*, The MIT Press, Cambridge, MA.

Tucker, S. (2007). Doors opening wider: Library and archival services to family history, *Archivaria* **62**: 128–158.

Tyler, D. C., Falci, C., Melvin, J. C., Epp, M. L. and Kreps, A. M. (2013). Patron-driven acquisition and circulation at an academic library. Interaction effects and circulation performance of print books acquired via librarians' orders, approval plans, and patrons' interlibrary loan requests, *Collection Management* **38**(1): 3–32.

Valtysson, B. (2017). From policy to platform: the digitization of danish cultural heritage, *International Journal of Cultural Policy* **23**(5): 545–561.

VanSnick, S. and Ntanos, K. (2018). On digitisation as a preservation measure, *Studies in Conservation* **63**(sup1): 282–287.

Whitner, R. and Balci, O. (1989). Guidelines for selecting and using simulation model verification techniques, *in* E. MacNair, K. Musselman and P. Heidelberger (eds), *Proceedings of the 1989 Winter Simulation Conference*, pp. 559–568.

Yuan, W., Van Ballegooie, M. and Robertson, J. L. (2018). Ebooks versus print books: Format preferences in an academic library, *Collection Management* **43**(1): 28–48.

Chapter 6

Preservation

'An old Chinese proverb says, 'Men grow old, pearls grow yellow, there is no cure for it'. It does not matter whether you completely agree with that; however, in all of nature this is true, and it is true of papers too.'

Ding 2009, p. 949

The development of the preservation part of the Archival Preservation Management (APM) model is at the core of this chapter. We first discuss the dose–response function and other models developed to predict the chemical and mechanical degradation of paper collections. The aim of the review is to justify the choices made in this part of the model, leading to the preference for the agent-based modelling (ABM) approach for the APM-preservation model.[1]

As is explained in detail in the course of this chapter, the preservation part of the APM model has been developed according to the diagram shown in Figure 6.1. First, the chemical degradation of the paper, understood as a loss of degree of polymerisation (DP), is modelled, and the effects of preservation measures, such as changing the environmental conditions or deacidification, are determined. Then, the relationship between the chemical condition of the paper and the risk of mechanical degradation, which is seen as the accumulation of tears, is established. As the wear and tear related to the use of the collections are strong, the accumulation of wear and tear is linked to the access part discussed in the previous chapter.

[1]Parts of this chapter (Section 6.1.4, 6.4.1, 6.5, 6.7, including figures in Section 6.5.2.3, 6.7.1, 6.7.2, 6.7.3 and Figure 6.17) have been published in Cristina Duran-Casablancas, Matija Strlič, Gabriëlle Beentjes, Gerrit de Bruin, Jaap van der Burg & Josep Grau-Bové (2021) A Comparison of Preservation Management Strategies for Paper Collections, Studies in Conservation, 66:1, 23-31.

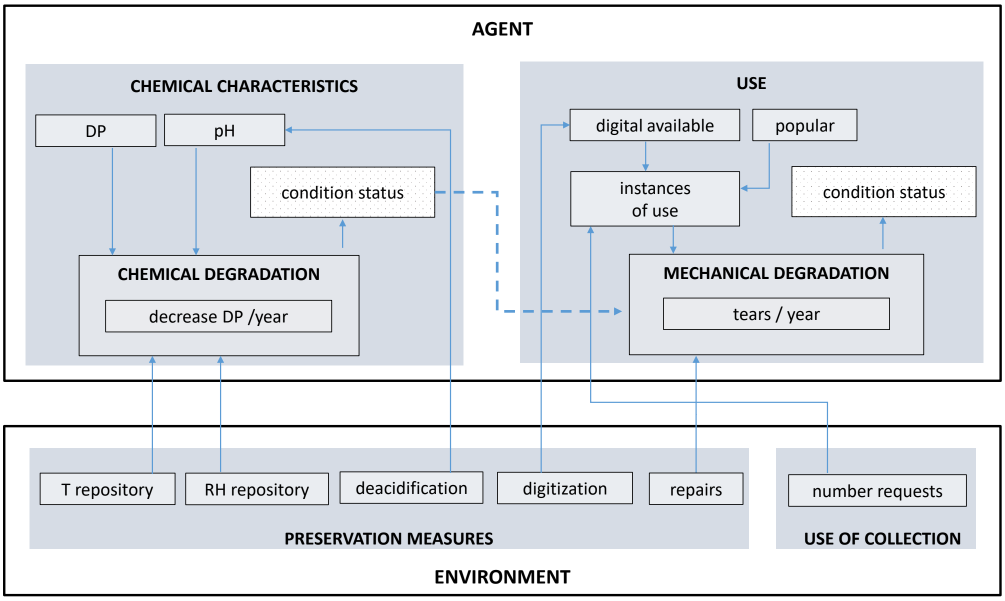


FIGURE 6.1: Diagram of the preservation part of the APM model.

Following the description of the APM model, we briefly discuss the sources of uncertainty in the APM-preservation model, particularly related to the pH variable. Finally, to demonstrate the application of the APM-preservation model as a virtual laboratory, we examine the impact of deacidification treatments and several storage condition strategies on the lifetime of the collections.

6.1 Modelling chemical degradation in paper collections

In Chapter 2, we presented tools that support decision-makers when assessing and evaluating the impact of storage conditions on the (chemical) degradation of the collections. We also briefly introduced the models underlying these tools. In this chapter, we explore the equations of cellulose degradation on which these tools are based and the studies in which different preservation strategies have been compared. The aim of this introductory section is not to critically review the literature on the kinetics of cellulose degradation but to put in context the reviewed tools by examining the equations used in each approach.

6.1.1 The kinetics of cellulose degradation

The causes of paper degradation are biological, physical and chemical. Within chemical degradation, there are two main mechanisms by which cellulose deteriorates: acid-catalysed hydrolysis and oxidation (Daniels; 2006). Acid-catalysed hydrolysis has been identified as the most important mechanism in cellulose degradation (Whitmore and Bogaard; 1994; Zou et al.; 1996). This is particularly the case for collections dating from the 19th century, when low-quality wood fibres and alum-rosin sizing were introduced into the paper manufacturing industry, resulting in acidic pH values of paper. Rosin sizing was added to the pulp as internal sizing from the 1840s until the 1980s (Barrow; 1974). To fix the hydrophobic rosin molecules on the surface of the fibres, a mordant, such as alum, was added. Aluminium sulphate ($Al_2(SO_4)_3 \cdot 18H_2O$), also known as papermaker's alum, is the most efficient in a pH of 4–5.5 (Roberts; 1996; Gess and Rodriguez; 2005; Hubbe; 2006). During this process, sulphuric acid is produced as a by-product, which together with aluminium sulphate are the main causes of acidity of paper. However, other sources of acidity in the paper have also been identified; these include residual chemicals from pulping and bleaching processes, oxidation of lignin to acidic products, transfer of acids from storage materials, and pollutants (e.g. such as sulphur dioxide and nitrogen dioxide (Strlič and Kolar; 2005; Kyujin; 2013)) present in the storage room.

In the presence of an acid, the glycoside bond is hydrolysed, which causes the cellulose polymer to split into two smaller units (Strlič and Kolar; 2005). Cellulose is also susceptible to oxidation. During oxidative degradation, carbonyl and carboxyl groups are formed in the weak parts of cellulose (hydroxyl groups), causing secondary reactions and eventually chain scissions (Ding; 2009).

For decades, research has been conducted into the kinetics of cellulose degradation to not only determine the long-term performance of paper but also predict paper permanence. Ding and Wang (2008) describe the three distinct steps when analysing the kinetics of cellulose degradation. The first step is to define the degradation variable, understood as the formation of by-products, loss in mechanical integrity (e.g. tensile strength (TS)) or change in chemical structure (e.g. DP). The second step is to describe the equation of the degradation variable 'that produces consistent relationships between degradation and material response and allows representation of degradation evolution rates' (Ding and Wang; 2008, p. 208). In the third step, these relationships are applied to predict the degree of degradation of cellulose at ambient conditions and its end of useful life.

Historically, the rate of paper degradation has been characterised in terms of its dependence on the DP of paper, since the main degradation mechanisms in cellulose, hydrolysis and oxidation lead to chain scissions. The most widely used equation to study the kinetics of cellulose is the one proposed by Ekenstam in 1936:

$$\ln\left(1 - \frac{1}{\mathrm{DP}_0}\right) - \ln\left(1 - \frac{1}{\mathrm{DP}}\right) = k \cdot t \tag{6.1}$$

where a random first-order chain scission reaction is assumed. However, the order of the reactions has been a topic of discussion (Calvini; 2012) due to the simplified Ekenstam equation used in modern literature:

$$\frac{1}{\mathrm{DP}} - \frac{1}{\mathrm{DP}_0} = k \cdot t \tag{6.2}$$

where k is the degradation rate coefficient (year^{-1}), t is the ageing time expressed in years and DP is the number-average degree of polymerisation of cellulose at time t and zero.

A detailed explanation of the meaning of the Ekenstam equation can be found in Strlič and Kolar (2005, p. 32), Calvini (2012), Ding and Wang (2008). However, we must put this equation into context by analysing some of its implications. According to Eq 6.2, the degradation of cellulose is approximated by plotting reciprocal DP vs. time. When the Ekenstam equation is applied, a rectilinear plot is obtained where the slope of the line should equal k, the rate constant. However, if the DP value is high and all bonds are equal, then the DP change follows the first-order rate law where the rate is proportional to the number of remaining unbroken bonds (Strlič and Kolar; 2005, p. 32). Therefore, when the direct plot of the DP of experimental data (and not the reciprocal) is plotted as a function of time, a rectangular hyperbola is obtained (Fig. 6.2). The limitations of the applicability of the Ekenstam equation have been pointed out by Emsley and Stevens (1994). For instance, the equation is only applicable for large initial values of DP and when the end-peeling reactions play a minor role in the degradation process. This is in part because the decrease of accessible bound in the amorphous areas during a chain scission is not taken into account in the Ekenstam equation. Hence, in this

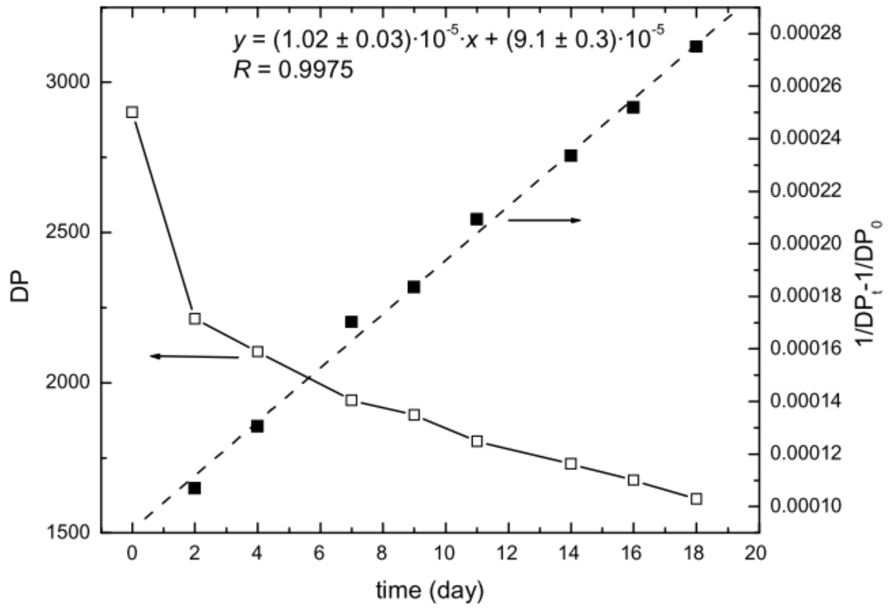


FIGURE 6.2: Changes in DP of Whatman filter paper sample with DP_0 of 2900, aged at 90 an 65%, and the same plot transformed according to the Ekenstam equation (Strlič and Kolar; 2005, p. 34).

equation, degradation occurs according to the rate constant *k*. To address this discrepancy, Emsley (1994) proposed an adapted model, where the decrease of *k* over time is included in a first-order manner, which provides a model closer to the results of experimental studies that have shown that four degradation phases can be distinguished with differing rates of chain-scission (Strlič et al.; 2001).

Following the same line, the Ekenstam equation implies that the value set as the minimum DP is 1. The Ekenstam equation does not take into account that when the DP value approaches the number of monomers in a typical crystalline region, the degradation reactions will strongly slow down. This value of DP is referred to as the levelling-off degree of polymerization (LOPD), and it has been included in the model proposed by Calvini (2005) as the minimum value that DP will reach.

Despite the limitations of the simplified Ekenstam equation, Emsley and Stevens (1994) have shown that the equation applies to a wide range of experimental data. Therefore, the Ekenstam equation has been widely applied in studies to describe acidic degradation, light ageing and accelerated thermal ageing (Ding and Wang; 2008). According to Emsley and Stevens (1994), the linear relationship is applicable until the DP value of ca. 200, which is also the critical value associated with the end of life of cellulose. Therefore, although the limitations of the Ekenstam equation are well known, the equation is applied to calculate the time remaining before the DP drops to the value regarded as the end of the useful life according to the following equation:

$$t = \frac{1}{k}\left(\frac{1}{\mathrm{DP}_t} - \frac{1}{\mathrm{DP}_0}\right) \tag{6.3}$$

In addition to the Ekenstam equation, another important equation in studies that predict the degradation of cellulose is the one proposed by S. Arrhenius in 1889. The Arrhenius equation links the degradation rate k to T as follows:

$$k = Ae^{-E_a/RT} \tag{6.4}$$

where k is the degradation rate coefficient (s^{-1}), A is the frequency factor (s^{-1}), Ea is the activation energy ($\mathrm{J\,mol^{-1}}$), R is the universal gas constant ($8.314\ \mathrm{J\,mol^{-1}\,K^{-1}}$) and T is the temperature (K).

The Arrhenius equation can be rearranged to plot the linear relationship between $\ln(k)$ and the inverse T. The activation energy (E_a) determines the slope of the plot.

$$\ln(k) = -(E_a/R)(1/T) + \ln A \tag{6.5}$$

The Arrhenius equation is used in accelerated ageing tests to determine how T affects the reaction rate. In such tests, the RH is kept constant. To include the effect of RH on the deterioration rate, several series of Arrhenius need to be repeated at different humidity values. If the activation energy of the material of interest is known, then the Arrhenius equation allows us to compare the effect of environmental conditions on the rate of deterioration.

While the activation energy in the Arrhenius equation indicates the sensitivity of the degradation rate to temperature changes, the parameter A, the frequency factor or the pre-exponential factor, represents all other experimental parameters such as humidity, acidity, exposure to pollutants and light, and physical structure of paper (Menart et al.; 2011). In Zou et al.'s study (1996), A was decomposed by including the concentration of hydrogen ions and the moisture content in the paper.

$$A = A_n + A_W[H_2O] + A_{WH}[H_2O][H^+] \tag{6.6}$$

where $[H^+] = 10^{-pH}$. Note that A includes hydrolytic reactions (A_W), acid-catalysed reactions (A_{WH}), and non-hydrolytic reactions such as oxidation (A_n).

Combining Eqs. 6.4 and 6.6, k can be defined as follows:

$$k = (A_n + A_W[H_2O] + A_{WH}[H_2O][H^+])^{-E_a/RT} \tag{6.7}$$

$[H_2O]$ is the water content in paper. An important point of discussion in the literature has been the assumed linear dependence of water content on RH in the environment, which might be true only in a limited range of RH values (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and De Bruin; 2015b). Strlič et al. proposed using the Paltakari and Karlsson equation to express the water content in paper as follows:

$$[H_2O] = \left(\frac{\ln(1-\mathrm{RH})}{1.67T - 285.655}\right)^{\frac{1}{2.491-0.012T}} \tag{6.8}$$

If Eq. 6.8 is included in Eq. 6.7, then $k = f(T, \mathrm{RH}, \mathrm{pH})$. Based on the experimental data collected from 121 paper degradation experiments, Strlič et al.(2015b) found that the best fit, a linear combination of the factors and not necessarily directly derived from Eq. 6.7, was obtained using the following equation:

$$\ln(k) = a_0 + a_1[H_2O] + a_2\ln[H^+] - \frac{a_3}{T} \tag{6.9}$$

Combining Eqs. 6.9 and 6.8, the dose–response function for paper is then defined as follows:

$$\begin{aligned}\ln(k) =& 36.981 + 36.72\left(\frac{\ln(1-\mathrm{RH})}{1.67T - 285.655}\right)^{\frac{1}{2.491-0.012T}} \\ &+ 0.244\ln\left(10^{-\mathrm{pH}}\right) - \frac{14300}{(T+273.15)}\end{aligned} \tag{6.10}$$

where k is the rate constant (year $^{-1}$), RH is expressed as a ratio, T (°C) and pH is that of paper.

Having defined k, the number of years until paper reaches the end of useful life, which is understood as a critical DP value, can be calculated by applying the Ekenstam equation as seen in Eq. 6.3.

6.1.2 Studies on the comparison of preservation measures

The equations of Ekenstam and Arrhenius and their adaptations have been widely used in studies predicting the extension of the useful life of paper. The main difference in these studies is the level of detail of how k is defined. Whereas some of the studies apply the general approach of the Arrhenius equation (using the activation energy value), others attempted to define k in more detail, by including the most important factors such as the characteristics of paper (e.g. pH).

For example, using the Arrhenius equation and the concept of activation energy, Michalski (2003) explored the general role of "double the life for each five-degree drop, more than double the life for each halving of relative humidity". Vinther Hansen and Vest (2008) applied the lifetime prolongation reported in the ASHRAE Handbook to estimate the lifetime of collections. In these approaches, the slope of the line on the Arrhenius plot (degradation rate) was determined by the specific activation energy of the material. However, to simplify the calculations, average activation energy is chosen, which is considered representative of the material under study (e.g. using just one activation energy for paper materials, without differentiating between paper types). The tools discussed in Chapter 2, the eClimateNotebook and Physics of Monuments, follow this approach.

The Arrhenius approach, in combination with the Ekenstam approach, has been applied in studies conducted by Balazik et al. (2007) and in the PaperTreat Project, co-funded by the European Commission's 6th Framework Programme (Kolar; 2009). Based on experimental data, these studies reported how not only environmental conditions but also other preservation measures, such as deacidification, can extend the useful life of paper. In a recent study by Tétreault et al. (2019), a more elaborate model was presented based on the Arrhenius and Ekenstam approaches, as well as the models proposed by Calvini (2005) and Ding and Wang (2008), where the kinetics of cellulose degradation are described in more detail by including exogenous factors (e.g. temperature, RH and gaseous pollutants) as well as endogenous factors such as the increase in acidity and the decrease in moisture content during paper degradation. The aim of Tétreault's study was not only to compare different environmental scenarios but also to describe the decay curve of cellulose.

While the results of these experimental studies are specific to the analysed paper, the dose–response function of paper degradation proposed by Strlič et al. (2015b) has been developed to support the management of collections, by calculating an object's remaining time until it becomes unfit. For example, Table 6.1 shows how many years it takes to reach the critical DP of 300, depending on the initial pH and DP at 18 °C and 50% RH. According to the model, the pH of paper is the most important factor that explains the degradation rate of paper. When examining the stability of paper, the type of fibre is considered less relevant than the manufacturing processes, such as rosin-sizing and certain pulping and bleaching processes, that introduce acids in the paper (Gurnagul et al.; 1993). However, it is clear that even with a low pH, if the initial DP is high, as in the case of rag paper, then it would still take centuries to reach the critical DP of 300.

TABLE 6.1: Number of years to reach the critical DP of 300, depending on the initial pH and DP at 18 °C and 50% RH, according to the dose–response function for acid-catalysed degradation. The red font indicates the time horizon of 500 years. The values in grey are the result of combining DP and pH values that are not normally expected in actual paper.

pH	**DP**							
	400	**500**	**600**	**700**	**800**	**900**	**1000**	**...3000**
3	56	88	110	126	137	146	154	197
3.5	73	116	145	166	181	193	203	261
4	97	154	192	219	240	256	269	345
4.5	128	203	254	290	317	338	355	456
5	167	268	335	382	418	446	469	603
5.5	223	356	445	508	556	593	622	800
6	295	471	589	673	736	785	824	1059
6.5	390	624	779	891	974	1039	1091	1402
7	516	826	1032	1179	1289	1375	1444	1856
7.5	684	1093	1366	1561	1707	1821	1912	2458
8	905	1447	1809	2067	2261	2411	2532	3255
8.5	1198	1916	2395	2737	2994	9193	3353	4310

However, how does the dose–response function of paper relate to the results reported in other studies? In the next table, we compare the dose–response function with three other studies. The first is the study by Vinther Hansen and Vest (2008).

In their study, the lifetime of collections was explored following the lifetime prolongation reported in the ASHRAE Handbook (2007), which is a direct application of the Arrhenius equation. The second is the one by Balazic et al. (2007), in which the Ekenstam and the Arrhenius equations were applied to extrapolate the results of artificial ageing of paper samples, to obtain data on the effects of deacidification and cool and cold storage on the useful life of paper. The third study is Tétreault's study (2019), which proposed a more elaborate model including more parameters. According to the Tétreault model, the calculated decay curves of cellulose are not linear since k is not constant due to different factors (e.g. the increase in pH during degradation). These studies report the factor of how useful time of paper (pH 6) is extended, compared to a baseline, usually in storage conditions of 20-21 °C and 50% RH, by changing the environmental conditions. We compared the results of these three studies with the output of the dose–response function, using the same input values as reported in the literature. Table 6.2 shows that the factors reported in the literature are similar to the values predicted by the dose–response function. Only Tetreault's study reported a slightly more prominent difference. This might be explained by the fact that Tétrault's model took into account the change in variables, such as pH and moisture content over time, whereas, in the remaining models, these variables are assumed to be constant.

In the PaperTreat Project, co-funded by the European Commission's 6th Framework Programme, the effect of deacidification, as well as the actual pH of the paper, was also explored to predict the extension of useful life of paper (Kolar; 2009). This study showed that the factor of stabilisation is expected to be higher for very acidic papers. PaperTreat reported a factor of 3–6 for very acidic paper (pH value of 3.5) and 1.5 to 2.5 for paper with a pH value of 5.5. These factors are comparable to the ones calculated according to the dose–response function (factor 2 to 7) if a pH value of 7 is assumed after deacidification.

In Vinther Hansen and Vest's study (2008), based on the ASHRAE Handbook, the life expectancy of an acidic collection is calculated. Since the collection is well-described, we can reproduce the data set and simulate the degradation of the collection for 200 years, the period suggested by Vinther Hansen. In this case, where predictions are made for collections, the outputs of the two models are also in line, as indicated in the Table 6.3.

The results of these analyses suggest that the difference in the factors between

TABLE 6.2: Predicted extension of useful time of paper (factor) under different storage conditions according to Balazic (2007), Vinther Hansen (2008), and Tétreault (2019) compared to the dose–response function for acid-catalysed degradation (initial DP 1500, pH 6) developed by Strlič et al. (2015b)

Scenario		**Vinther Hansen (1)**	**Balazic (2)**	**Tétreault (3)**	**Model**	
T **(°C)**	**RH (%)**	**Factor**	**Factor**	**Factor**	**DP decrease /year**	**Factor**
27	65	-	-	3	7.2	3.8
21	50	-	-	-	1.9	baseline (3)
20	50	-	-	-	1.7	baseline (1-2)
21	15	-	-	2	0.5	3.5
18	40	2	-	-	0.9	1.9
15	50	-	1.9 ± 0.8	-	0.8	2.0
12	45	4	-	-	0.5	3.7
10	40	-	-	5.5	0.3	6.7
5	50	-	8 ± 4	-	0.2	8.9
5	30	20	-	-	0.1	18.9

the studies is small and that they are equally valid to explore the relationship between the environmental conditions and paper degradation. However, the advantages of using the dose–response function are clear:

- the dose–response function has been formulated and, therefore, the calculations are straightforward.
- The dose–response function allows for calculating the remaining useful years of paper. Although it is important to note that although a quantitative output is given, the output should be interpreted as an indication, similar to the report on factors. However, reporting in years instead of factors provides more context on the expected prolongation of the useful life of paper.
- The inputs (DP and pH) are perhaps more familiar to the end-users than the concept of activation energy; this allows us to explore the difference

TABLE 6.3: Predicted extension of the useful life of the collections at the Royal Library Denmark, according to different preservation scenarios, suggested by Vinther Hansen and Vest (2008), and the dose–response function for acid-catalysed degradation developed by Strlič et al. (2015b).

Scenario		**Royal Library Denmark**			**Model**		
T (°C)	**RH (%)**	**Unfit for purpose (%)**	**Difference baseline scenario (%)**	**Factor**	**Unfit for purpose (%)**	**Difference baseline scenario (%)**	**Factor**
18	50	80	-	-	71.5	-	-
16	45	60	20	1.3	52.5	19.0	1.4
12	45	35	45	2.3	33.0	38.5	2.2
5	30	18	62	4.4	15.2	56.3	4.7
deacidification		40	40	2.0	34.6	36.9	2.1

between different papers.

- By including pH as a parameter, the effect of deacidification can be also explored.

6.1.3 Pollutants

As seen in the previous sections, temperature and RH are the main exogenous parameters included in any model to predict the chemical degradation of paper. Pollutants have been extensively studied, but only a few models include the presence of pollutants as part of the model.

A literature review on the effect of pollutants on cellulose degradation is beyond the scope of this research. An extended exploration into pollutants, particularly indoor pollution, in archival collections is offered by Menart (2013). However, we would like to review some recent studies that have introduced pollutants as a parameter for the prediction of the life of paper. Menart (2011) questions some previous studies that investigated the negative effects of pollutants through high concentrations of pollutants not found in realistic storage conditions. In her study, Menart (2014) calculated the lifetime of paper exposed to realistic pollutant concentrations (nitrogen dioxide (NO_2) and acetic acid (CH_3COOH) and compared the results with those of a control experiment, where it was assumed that paper will degrade even in the absence of pollutants. She concluded that at realistic conditions of 100 ppb acetic acid and 10 ppb NO_2, there was no effect on most of the papers. Paper for which a small effect has been observed would still reach lifetimes of several millennia.

In view of these results, Di Pietro et al. (2016) suggest reconsidering the chemical air filtration in archives and libraries, from a cost-benefit perspective. The limited negative impact of acetic acid on paper degradation has been further stressed by this research group (Ligterink and Di Pietro; 2018). Ligterink and Di Pietro developed a model consisting of three parts: (1) an equilibrium model between the acetic acid in the air and the acid absorbed by the paper, (2) the pH-shift of the paper due to the absorbed acetic acid, and (3) the de-polymerization given the calculated pH-shift. Note that this model proposes to introduce a pH-shift as part of the dose–response function. The authors concluded that since the predicted pH change of the paper can only be noticed for paper with initial pH of 6–8.5 and not in a lower range, it is uncertain whether acetic acid should be included in the model given its limited effect on the degradation rate of acidic papers.

In the model of paper degradation mentioned earlier, Tétreault et al. (2019) also introduced exogenous pollutants and volatile organic compounds as model parameters. Based on the results of their study (Tétreault et al.; 2013), which demonstrated that NO_2, formic acid and acetic acid are reactive pollutants in cellulose degradation, Tétrault developed different scenarios to model the impact of pollutants. The experiments showed that the effect of 30 ppb of NO_2 would be equivalent to that of 50 ppb of formic acid, whereas acetic acid is about 100 times less reactive. However, as these results were not directly compared to a control experiment in the absence of pollutants, it is difficult to evaluate the effect of pollutants based on this model.

6.1.4 pH of paper and deacidification

As discussed in section 6.1.1, pH is a key parameter in the paper degradation model. However, the measurement of the pH of a paper is not generally easy. pH is measured in solutions, which means that water needs to be added to the paper to measure the concentrations of hydrogen ions. Therefore, the amount of water influences the pH level. In an evaluation of nine methods to measure the pH of paper, Saverwyns et al. (2002) concluded that only the extraction techniques and the surface electrode yielded accurate results. Saverwyns also reported that '[t]he errors on the measurements, which are a measure of the precision, are significantly higher with the surface electrode as compared to extracts' (Saverwyns et al.; 2002, p. 630). The results of this study are consistent with a previous finding (Brandis; 1993) that surface methods are less repeatable and return lower values than extraction methods.

According to Saverwyns et al. (2002), the difference in pH value obtained using cold extraction or with surface electrodes can be explained because little dilution occurs in the surface electrode method. Furthermore, Joel et al. (1972) showed that the added matter to the paper (e.g. size, buffer) dominates the pH measurement. For instance, a difference between surface pH and cold extraction of 1–1.5 pH units has been reported in several studies (Strlič et al.; 2004; Ahn et al.; 2011), particularly in the case of papers loaded with alkaline fillers or gelatin sized papers. For the gelatin (surface) sizing, the surface will be more acidic due to the presence of alum (potassium aluminium sulphate), which is added to the gelatin to reduce the risk of biodeterioration and control the gelatin's viscosity. Strlič and Kolar pointed out that 'in most gelatin surface sized papers, the surface will be more acidic than the bulk (up to a few pH units), so that on the basis of a surface pH measurement, a wrong decision regarding the need for deacidification can be taken' (Strlič and Kolar; 2005, p. 27).

Another uncertainty regarding the pH of paper is the change in the acidity level of cellulose during ageing. The degradation of paper is an auto-catalysed process, since 'pH is not only the cause but also the result of paper degradation' (Strlič et al.; 2020, p. 8296). In the case of deacidified paper, the alkaline reserve aims to neutralise the acids that are potentially formed during ageing. Although observed in artificial ageing (Bansa; 1998; Ramin et al.; 2009; Hanus et al.; 2006), the change in pH has only been reported in a few studies of naturally aged papers (Porck; 2006; Ahn et al.; 2011). However, the extent of re-acidification in naturally aged samples is difficult to determine due to the expected differences in the instrumentation, the analytical procedure and electrode calibration and users, between the two points of measurement. Barański et al. (2005) explored the pH values of paper as a function of ageing time for a model softwood cellulose paper containing 0.8‰, using the Ekenstam equation. The experiments showed an increase of acidity during artificial ageing at quite severe conditions (90°C and high Al content). However, in less severe conditions, the drop occurred after two and half days ($\approx$ 0.3 units) and remained almost constant over time.

6.1.4.1 Mass-deacidification

Deacidification aims to neutralise the acids and introduce an alkaline reserve to guarantee the sustainability of the treatment. In general, documents written after 1850 are usually the target group for deacidification because they are potentially produced on acidic paper (Hanus and Mináriková; 2002). Within this group, some institutions, such as the Swiss National Library, have identified three groups depending on the urgency of treatment: first collections dating between 1930 and 1985, those printed before 1900 and up to 1930, and those from 1985, where individual items are selected (Grossenbacher; 2006).

Besides the age of paper production, the most important selection criterion is the acidity level (pH < 7) of the paper to be treated (Pilette; 2004). However, to narrow this group and prioritise documents at a greater risk, a lower pH value is chosen as the upper limit for the selection. For example, Metamorfoze, the Dutch National Programme for the Preservation of the Paper Heritage, advises selecting acidic paper with pH less than 6 (Beentjes; 2016). At the British Library, although deacidification is carried out on an individual basis with no mass-deacidification programme in place, their selection criterion used to be a pH less than 7, but since 1999, only the material is deacidified if the pH is below 5.5 (Shenton; 2006). A high lignin content, examined with a phloroglucinol spot test, has also been suggested as a selection criterion (Porck; 2006).

Another selection criterion is whether the paper is too brittle, and, therefore, too fragile to be handled. Paper that has reached this state will not benefit from the deacidification treatment, as long as this treatment is not performed alongside a strengthening treatment for the cellulose fibres, which remains a topic of research. Deacidification is a preventive, rather than a curative, measure. Therefore, for some institutions, an important selection criterion is that deacidification takes place when paper is still strong and flexible (Grossenbacher; 2006).

Two major processes for mass-deacidification can be identified, according to how the alkaline reserve is supplied: as a solution or a suspension (dispersion) in non-aqueous, inert solvents. An overview of the existing mass deacidification systems is given in Blüher and Großenbacher (2006), Baty et al. (2010), Zervos and Alexopoulou (2015). The quality control after treatment has been a topic of discussion since the development of the deacidification treatment. Quality control standards recommend a pH between 7 and 9 after treatment, and the current DIN recommendation is an alkaline reserve of 0.5-2.0% of $MgCO_3$ (Cedzová et al.;

2006; Grossenbacher; 2006; Andres et al.; 2008; Baty et al.; 2010). A previous study showed that better performance of deacidification treatment can be achieved with homogeneous solutions compared to dispersion variants, due to the better penetration of alkaline reagents in cellulose fibres (Potthast and Kyujin; 2017).

In a study conducted by Ahn et al. (2011), the pHs of deacidified books with Papersave and CSC Book Saver® method from two different libraries were measured. The results showed that 84% and 73% of these samples met a pH value of 7. The efficiency of the deacidification treatment depends on the material composition of the paper and its state of preservation, e.g. pH value, at the time when the deacidification was conducted. Another factor that plays a role is the date of the records. It has been observed that among books published from 1960 onward, the deacidification treatment was more successful compared with books published before 1960.

6.2 Modelling mechanical degradation in paper collections

6.2.1 Wear and tear function

As part of the Collections Demography Project, in addition to the dose–response function for chemical degradation, a wear-out function was developed to model the accumulation of mechanical degradation during handling (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and De Bruin; 2015a). The wear-out function calculates the number of missing pieces per handling as a function of the DP of paper. The equation describing this function for large missing pieces (LMP) is:

$$ln(no_{LMP}/handling) = -(0.01050 \pm (0.00097)^{\cdot} DP + (5.02 \pm 0.51) \qquad (6.11)$$

The function for the accumulation of tears can also be derived from the same data set of the Collection Demography Project. In the regression model, we included only samples with DP < 800, when the accumulation of tears is a function of the instances of handling, and we excluded samples with a higher DP, since experimental results have shown that the accumulation of wear and tear occurs randomly for paper with a DP > 800 (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and De Bruin; 2015a). Bound and unbound samples ($N = 27$) were included in the same function, and the mean number of tears per challenge, which is equivalent to 10 instances of use, was used as input. Figure 6.3 shows the regression of the mean tears per challenge versus DP.

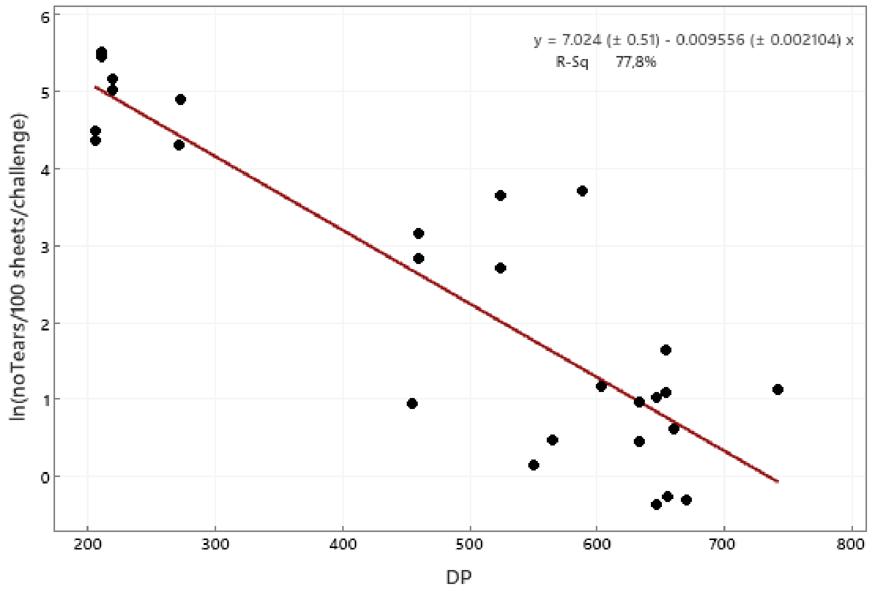


FIGURE 6.3: Accumulation of tears per challenge (equivalent to 10 instances of handling) depending on the DP of paper, per 100 sheets.

Therefore, the equation describing the function for tears can be described as follows:

$$ln(no_{Tears}/challenge) = -(0.0096 \pm (0.002)^{\cdot} DP + (7.0236 \pm 0.51) \qquad (6.12)$$

Using a logarithm equation, the range between the data is reduced, resulting in a slightly better regression fit compared to the equation when the data are not transformed ($R^2 = 0.78$ versus $R^2 = 0.71$). However, it is important to remember that by using a logarithm we are assuming that the accumulation of tears as a function of the actual DP of the paper is exponential (Fig. 6.4). In the case of the accumulation of tears, this assumption is reasonable since tears occur more easily when they are already present in the paper.

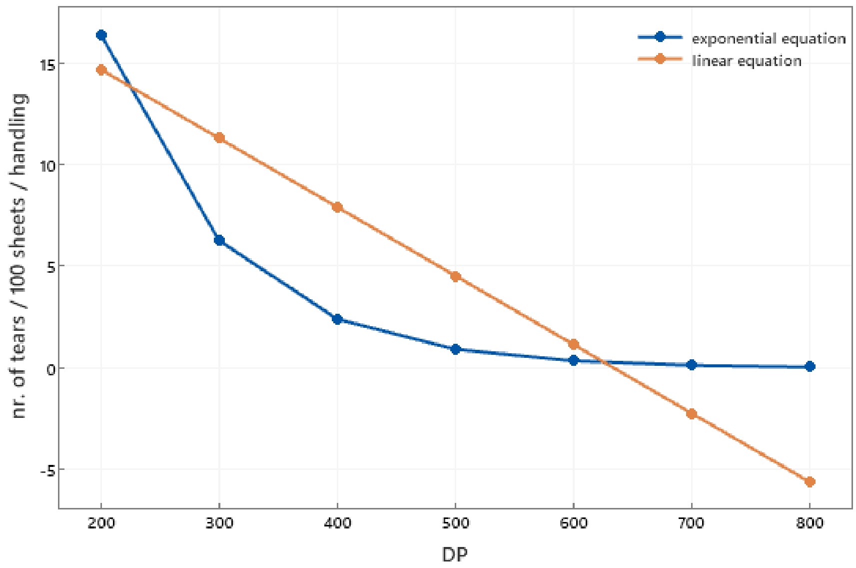


FIGURE 6.4: Number of tears per handling per 100 sheets depending on the DP of paper, according to a logarithmic or linear equation.

6.2.2 Survey data

The use of survey data to provide evidence for the patterns of decay in archival collections has been discussed in an epidemiological study conducted at Amsterdam City Archives (SAA) (Duran-Casablancas et al.; 2019). In an epidemiological study, the prevalence of a particular condition among a group of individuals, as well as the main factors related to that condition, are the topics of study. The observational study conducted at the SAA indicated the frequency of use in the reading room as the main factor in the accumulation of wear and tear. It was also observed that other factors, such as the thickness of inventory number documents and age, help predict mechanical failure more accurately. As discussed by Duran et al. (2019), factors prior to use (design and manufacture) and factors during use (maintenance and usage) need to be taken into account when analysing the causes of failure. In the case of archival collections, protection, understood as the type of housing, is considered an important factor to reduce mechanical degradation. For example, it was observed that boxes perform slightly better as a protective measure compared to (open) folders, when documents are requested in the reading room, as well as for records containing more than 100 sheets (Duran-Casablancas et al.; 2019). However, although these differences are statistically significant, a wear and tear function has not been developed yet, where the thickness or type of protection is included.

In an SAA cross-sectional study that we conducted in 2018 (Duran-Casablancas et al.; 2019), a small percentage of papers containing lignin and/or low pH values were present in the sample. Therefore, in this PhD project, we conducted another epidemiological study to investigate the relationship between mechanical degradation and use in collections with lignin-containing papers. To reduce the heterogeneity of the target population, the study was conducted with one archive of the Dutch National Archives (NA). The selected archive (52.5 metres and 3612 inventory numbers), dating from 1944 to 1950, contained mostly acidic groundwood paper. Instead of a cross-sectional study, we opted for a case-control study design. In these studies, the cases are selected because they have already developed a symptom, and then they are compared with the controls, without such symptoms. One of the advantages of these studies is the small sample size required compared to cross-sectional studies (Daly and Bourke; 2007). For this study, we selected 19 cases and 20 controls ($n = 39$). The division between case and controls was made according to the percentage of pages with missing pieces in one inventory number ($>$ 5% pages with missing pieces).

We found that the cases, with a higher percentage of failed pages, had been, on average, more frequently accessed than the control cases. However, these results were not conclusive, as the observed difference in use was not statistically significant ($t = 1.79$, $df = 26$, p 0.085). The observed difference in wear and tear could be related to other factors such as insufficient storage protection in the past.

It was also observed that according to the principal component analysis (PCA), failure in the cases was indeed mostly correlated to the number of accesses. However, for the controls, failure was correlated to not only the number of requests but also a low tensile strength (TS) and high lignin content. These results suggest that when the use is low, papers with certain chemical characteristics (eg. low TS, high lignin) will be the first in accumulated wear and tear. However, with high use, in combination with paper in poor condition, the number of requests is the dominant factor, suggesting that those with tears will accumulate even more tears, as shown in the difference in the percentage of failed pages for the two groups (Fig. 6.5).

Although these results help to formulate hypotheses, further research is needed before this type of data can be used in mathematical models. During this case-control study, the limitations encountered in previous observational studies were further confirmed:

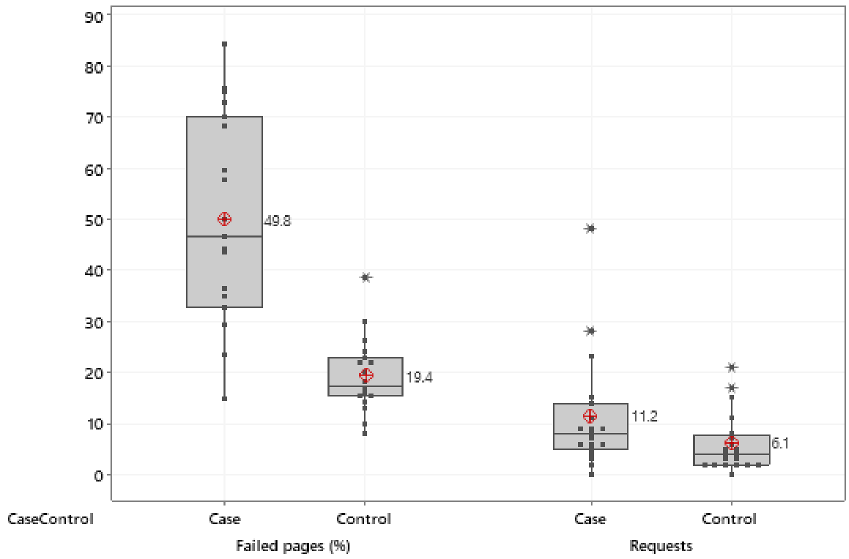


FIGURE 6.5: Boxplots of the percentage of failed pages and the number of requests for the two groups in the case–control study conducted on an archive containing acidic paper.

1. Data on the history of use is only known for a limited number of years (usually 10 to 20 years), which gives information for only a few years of the total use and maintenance since the creation of the records.

2. Inventory numbers usually refer to a stack of paper. However, the stack is often not homogeneous. Different types of papers and sizes can be found in one single inventory number. Therefore, it is difficult to conclude the relationship between damage and characteristics of the paper, when different types of paper can be found in the same stack.

3. Besides the number of instances of access and the chemical characteristics of the paper, other factors, such as the type of protection and the thickness of the stack, may contribute to wear and tear. Further research is needed to quantify the effect of these factors on the accumulation of wear and tear.

6.3 APM-Preservation model boundaries

Based on the findings in the literature review and the observational study conducted in the previous sections, we can now select the variables that will be included in the model. Here follows a summary of these choices (Fig. 6.6.

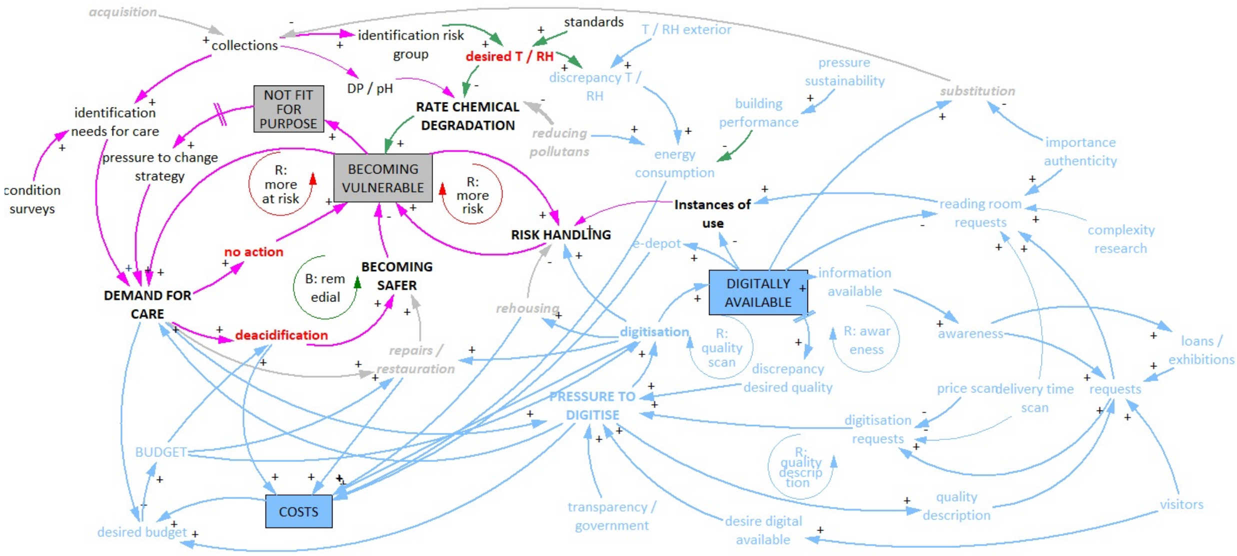


FIGURE 6.6: Causal loop diagram showing the variables included in the preservation part of the APM model. In red, conservation measures included in the model. In grey, the conservation measures that have been omitted. In blue the variables that are part of the access and cost part of the model.

Regarding which simulation approach (e.g. level of aggregation) is the most appropriate for this part of the model, in view of the aforementioned results, as well as the tools presented in Chapter 2, we learned that two main approaches are used: either the collections are represented by one single paper (e.g. all the collection has a pH value of 6) or a more realistic approach is suggested by representing the collection as a group of objects with different pH values. The dose–response function has shown the importance of pH, not only for the rate of degradation but also to study the effect of deacidification. Hence, by approaching the collections as one single type of object (with a certain pH) considerable information is lost. In the case of an SD approach, collections could be divided into three subgroups depending on their vulnerability, as a 'compartmental models', and using an 'ageing chain' structure, to capture some of the individual characteristics of the collections. However, this SD aggregate approach becomes more difficult when other variables are added, which are the result of certain combinations (e.g. wear and tear and DP and number of accesses).

Since these combinations of properties are key in the dose–response function, ABM appears to be the most adequate approach, as it allows the level of detail required to model the diversity of characteristics of every agent. In addition, using ABM in this part of the model, each parameter, e.g. the selection of the agents to be treated, the level of treatment and the time span of the treatment, can be adjusted at initialization of the run and, owing to the software used in this project, at any given time step during the run. These features will help explore the effect of different conservation measures during the lifetime of collections.

To summarise, regarding the modelling of the chemical degradation of paper collections:

1. The part of the model regarding the chemical condition of the collections will follow the dose–response function for acid-catalysed degradation (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and De Bruin; 2015b) since this model is well-defined, includes not only temperature and RH but also the pH of paper as a variable, and the results are comparable to those of other models.

2. Pollutants will not be included in the model for now, since the difference in output, between including pollutants or not, appears to be small.

3. The pH will not change over time since the change in the acidity of paper during natural ageing has not been quantified so far.

4. Change in environmental conditions (temperature and/or RH) and deacidification will be explored as potential preservation measures.

5. pH and DP of the paper will be the selection criterion for deacidification.

6. The effect of deacidification will be modelled as a frequency distribution of the pH of the paper after treatment. In this model, the relationship in the literature between the date of the paper and the effectiveness of the treatment will not be included.

Regarding the modelling of the mechanical degradation of paper collections:

1. Wear and tear, understood as the accumulation of tears over time, will be modelled as a function of the number of accesses and the DP of paper at the moment of access.

2. Results of epidemiological studies (e.g. effect of housing, the thickness of the stack of paper) will not be included, as no robust data are yet available on the effect of the (physical) characteristics of the collections.

3. The model will report the risk of collections of accumulating wear and tear, but the remedial preservation measure of tear mending will not be included.

Since the wear and tear function reports the number of tears per 100 sheets, the thickness of the stacks should be included too. This level of detail will not be included for now.

6.4 APM-Preservation model description

6.4.1 Chemical condition

Following the ABM approach, the first step in the model is to define the individual characteristics of the agents. An initial DP, pH value and a state (good, fair and critical) are attributed to each agent: good condition if the DP value is higher than 800, fair condition if the DP value is between 301 and 800, and critical condition if the DP has reached the critical value of 300. This classification follows the characterisation proposed by Strlič et al. (2015a) of fitness–for–use, where papers with a DP less than 300 are defined as unfit for any purpose due to their poor mechanical properties. During a run, the condition of the agents may change depending on their actual DP. A Boolean variable (`deacidified`) states whether the agents have been deacidified. By default, the agents are not deacidified, to compare the effect of deacidification to other preservation strategies.

As the focus of this study lies in the management of the chemical degradation of paper collections, the model includes two conservation options: changing the environmental conditions and deacidification (as discussed in section 6.1.1 and 6.1.4.1). The environment is controlled by modifying the values of temperature and humidity. Deacidification is controlled by setting three parameters: the number of metres deacidified per year and the pH and DP values used to select the agents to be deacidified.

The impact of the inputs on the behaviour of the agents is modelled using the dose–response function developed for degradation of cellulose (see Eq.6.10 in section 6.1.1) that calculates the remaining years until a DP value of 300 has been reached. To simulate the decrease in DP over the years, the rate of decrease is calculated by dividing the difference between the initial DP and critical DP of 300 by the predicted remaining years. This value is subtracted yearly from the actual DP of the agent during the duration of the run.

To model the effect of deacidification, when an agent is selected for this treatment, the pH value of the agent returns to a value between 7 and 9.5. As shown in a study by Ahn et al. (2011), depending on the applied deacidification method, different frequency distributions of the surface pH after treatment are obtained. In the section Experiments (section 6.6.2), we test different distributions to study the uncertainty associated with the effectiveness of the deacidification treatment and whether a theoretical uniform distribution for the pH value after deacidification is good enough for the model.

The output of the model is the annual percentage of the collection in a certain state, as well as the percentage of collection deacidified. The simulation has a run time of 500 years. This time horizon was chosen in agreement with the preservation horizon of 500 years reported in studies that investigated stakeholder attitudes (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and Bruin; 2015). Another time horizon could be chosen if so desired. In addition, in every time step, the input values can be varied. This enables selecting the duration of a certain measure and the time of implementation.

6.4.2 Mechanical condition

The wear and tear function in section 6.2.1 calculates the number of tears per handling, depending on the actual DP value of the paper. So to model the mechanical condition, we need to link two parts of the model: one that models the decrease in DP (section 6.4.1) and one that models the instances of use in the reading room in the access part of the model. In Chapter 5 we saw that the instances of use (`requestsTotal`) depends on whether the agent is frequently requested (`popular`) and still in state `InUse`. These instances of use (`requestsTotal`) are the input of the wear and tear function, together with the actual DP.

Although the number of tears is a quantitative output, this output should be interpreted qualitatively due to the following aspects:

1. According to the experimental design, the number of tears is calculated for 100 sheets and during ten instances of use. However, the number of sheets in one inventory model is not a parameter in the APM model.

2. The number of tears is calculated per agent. However, as already discussed in the access part of the model (Chapter 5), the instances of use are modelled using probability distributions, which do not necessarily correspond with the actual instances of handling in the reading room.

3. One agent represents a certain number of metres or inventory numbers depending on the unit. However, this number is an approximation, which is mainly chosen to meet the right proportion of records that have been requested and/or digitised within a collection.

Hence, the output of this part of the model is not the number of tears present in the collection, but the number of metres with a certain risk to accumulate mechanical degradation as a function of the chemical characteristics of the paper (DP) and the frequency of use in the reading room. Regarding wear and tear, agents can undergo four states, depending on the number of tears that they accumulate. By default, the agents are in good state (no tears), and the state may change depending on the number of tears (fair if the number of tears is between 25 and 49; poor for the number of tears between 50 and 74; and critical for more than 75 tears). Another threshold could be chosen if so desired.

The time horizon of this part of the model is 100 years, since the frequency of use is also modelled for a time span of 100 years in the access part of the model.

6.4.3 Layout of the APM-preservation model

Besides the APM model where access, preservation and cost come together, we also built each part of the model as a single model. The aim of the APM-preservation model is to explore the effect of deacidification and storage conditions on the chemical degradation of the collections. Figure 6.7 shows the dashboard of this model built using the software AnyLogic 8.6.

This model includes pre-settled collections, which will be described in the next section. At the beginning of the run, one type of collection is chosen. Other inputs can be modified at the beginning and during the run (see Fig. 6.7). Herewith, different preservation strategies can be easily explored. The main output of the model is the chemical condition of the collection over the time span of 500 years, as a percentage of the collection in critical, fair and good condition. Two histograms visualise how the pH and DP values of the collection change over time. The histogram of the pH values is particularly useful to get a sense of the extent of this measure. The percentage of the collections deacidified is also reported. The costs of the deacidification will be discussed in the next chapter. However, it can be observed that the costs of the deacidification are included in the inputs, as the available budget per year for deacidification, or as calculated output as annual and total costs of deacidification.

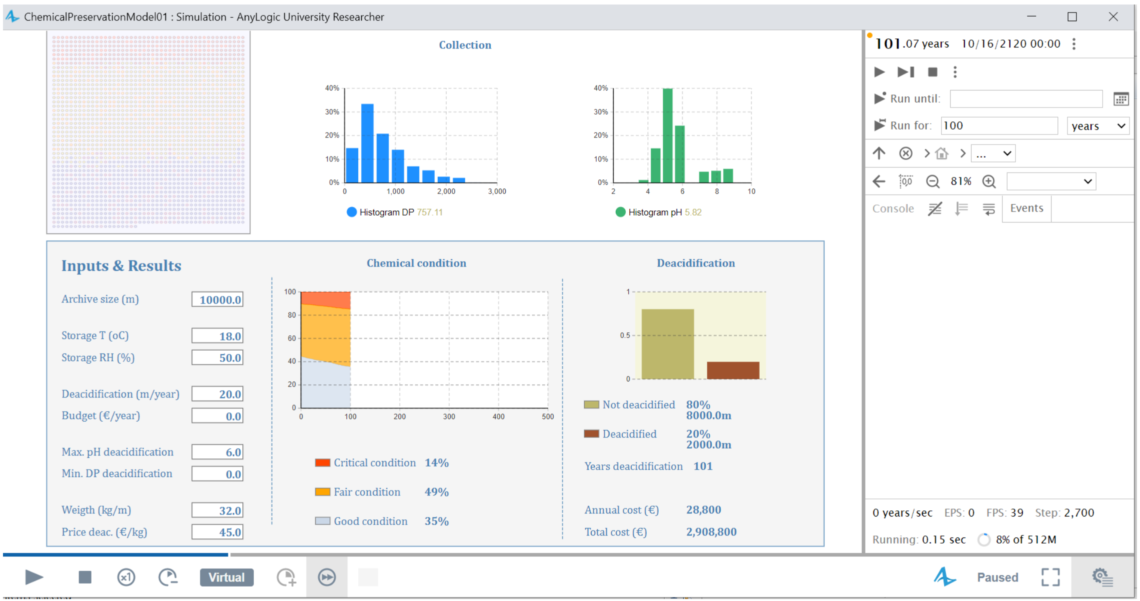


FIGURE 6.7: Screenshot of dashboard of the (chemical) preservation part of the APM model.

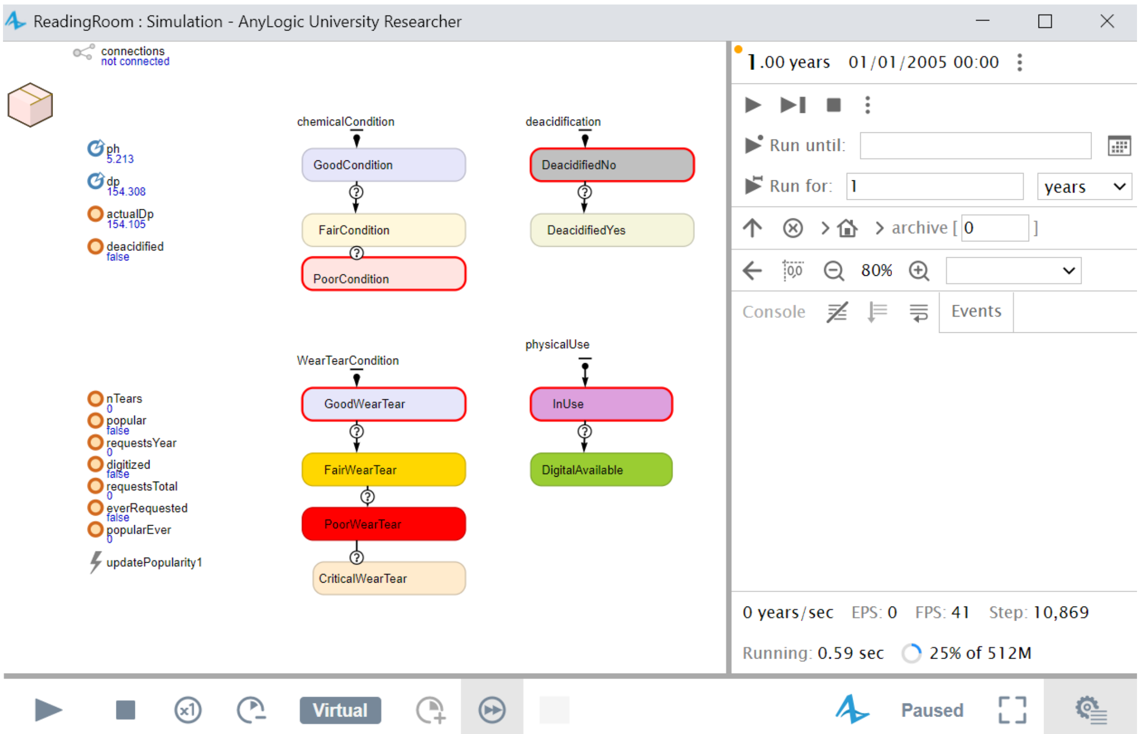


FIGURE 6.8: Screenshoot of one of the agents in the APM model.

Figure 6.8 shows the variables that define an agent. In the APM-preservation, only pH and DP, as well as the state chart of the chemical condition and deacidification, are part of the model. Regarding the variables to model wear and tear, the variables are the same as the ones stated in the previous chapter on the impact of digitisation on the access requests (see Chapter 5). However, now, to model wear and tear, the two parts of the model (access and chemical preservation) are linked in a single model. Therefore, variables have been added to count the number of tears, as well as a state chart, to report the mechanical condition of the whole collection.

A concept of the dashboard, as well as the description for the whole APM model, can be found in Chapter 8 and Appendix G.

6.5 Model data set generation

To explore the impact of preservation measures on different types of collections, we created model data sets, representing archival and library collections. To build the archival collections, the SurveNIR historic reference paper collection (665 samples (Strlič, Grossi, Dillon, Bell, Fouseki, Brimblecombe, Menart, Ntanos, Lindsay, Thickett, France and De Bruin; 2015b) and data from 431 papers of the collections of the Amsterdam City Archives (Duran-Casablancas et al.; 2017) and the Dutch National Archives (section 6.2.2) were collected using the SurveNIR instrument (Lichtblau e.K., Germany).

Three main model data sets were generated to perform the tests. The data sets represent three types of collections: acidic, modern and mixed collections. In this section, we discuss how the data sets were generated, the reference data sets used and the characteristics of these model data sets.

6.5.1 Reference data sets

In an epidemiological study conducted at the collections of the Amsterdam City Archives (Duran-Casablancas et al.; 2019), 431 paper samples, dating from the 17^{th} to the 20^{th} century, were analysed using the SurveNIR instrument (Lichtblau e.K., Germany). SurveNIR identified rag papers, bleached pulp papers and groundwood papers. To increase the number of papers dating after 1850, the results of a survey conducted at the Dutch National Archives (section 6.2.2) were also added to the data set. The resulting reference paper collection consisted of

541 papers, of which 34% were rag paper, 42% were bleached pulp paper and 24% were groundwood paper. Further, 58% of the samples had a pH value less than 6.

A limitation of the SurveNIR tool is that it does not provide DP values for groundwood paper. In Chapter 8, this limitation is explained in more detail, and a regression model to estimate the DP value using other variables provided by the SurveNIR is presented. For the experiments presented in this chapter, we chose a more simple approach. Based on the general correlation between DP and pH (Strlič et al.; 2020), we created a probability distribution for the DP value of the groundwood paper using the SurveNIR data set. The SurveNIR data set is a historical reference data set, which has been used to develop the SurveNIR instrument. For a description and analysis of the data set, see Strlič et al. (2020). To generate the DP values for groundwood paper, the values for acidic paper from the SurveNIR data set were used. Different probability distributions were tested for samples with a pH value $<$ or $\geq$5. The best fit to the data set was found with a lognormal probability distribution:
(1) pH $<$ 5: Loc: 5.99; Scale: 0.57; Thresh: 167 ($N = 66$; $p >$0.250)
(2) pH $\geq$5: Loc: 6.6; Scale: 0.48; Thresh: 47,04 ($N = 147$; $p>$0.250)

These probability distributions were used to generate the missing DP of groundwood papers (Fig. 6.9).

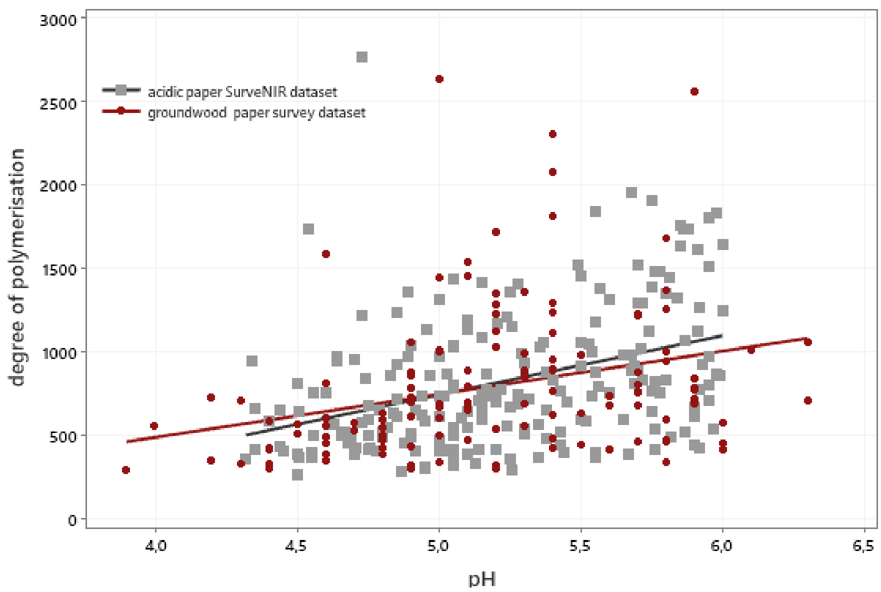


FIGURE 6.9: pH and DP of SurveNIR acidic papers and pH and generated DP of groundwood papers of the survey data set with regression line.

6.5.2 Model data sets

The reference data set of the Dutch archive collections, including the samples with the derived pH values for groundwood paper, was then used to generate the model data set. Three main types of collections were defined to represent collections with a different percentage of acidic paper in order to model collections with different life expectancy (see Table 6.1):

1. Acidic collections, with mainly papers with low DP and pH values (mostly groundwood paper).
2. Modern collections, dating from 1850, containing groundwood paper and bleached pulp paper.
3. Mixed collections, containing rag, bleached pulp and groundwood paper.

Rag paper is generally considered as a good-quality of paper, due to the quality of the raw materials: rag paper is made of linen and cotton fibers with gelatin as internal sizing. The pH value of rag paper is around 7 and has a DP of 1500. Rag paper was the only type of paper until around 1850 when wood fibers started to be used for papermaking. Compared to the rag paper, the raw materials and additives of groundwood paper, also called mechanical wood pulp, produced a low-quality acidic paper (see section 6.1.1). The papermaking fibers of the bleached pulp paper also derived from wood, but in these papers (most of) the lignin has been chemically removed from the fibers, resulting in a type of paper with a higher permanence.

6.5.2.1 Degree of polymerisation values generation

Based on the distributions of the Dutch archive collections, three different distributions were applied to generate the DP values for the model data set:

1. Acidic collections: Lognormal distribution – Loc: 6.6; Scale: 0.7
2. Modern collections: Weibull distribution – Shape: 2.2; Scale: 1300
3. Mixed collections: Lognormal distribution – Loc: 6.6; Scale: 0.7 and Weibull distribution – Shape: 6; Scale 1900

For the mixed collections, two distributions were used to generate the rag papers and modern papers separately.

Figure 6.10 and 6.11 show the histogram of DP values of the reference data set

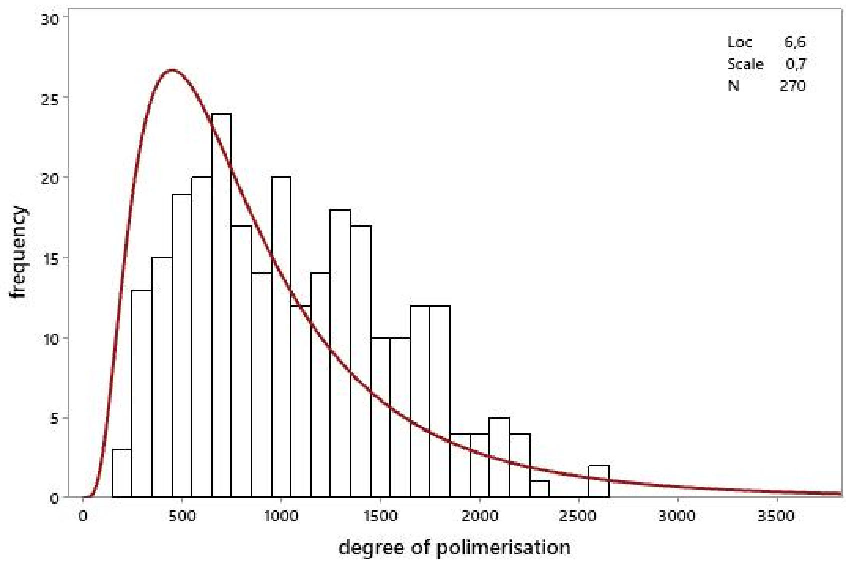


FIGURE 6.10: DP values of the Dutch archives reference data set (bins) with distribution (curve) used to generate the DP values of the model data set representing acidic collections.

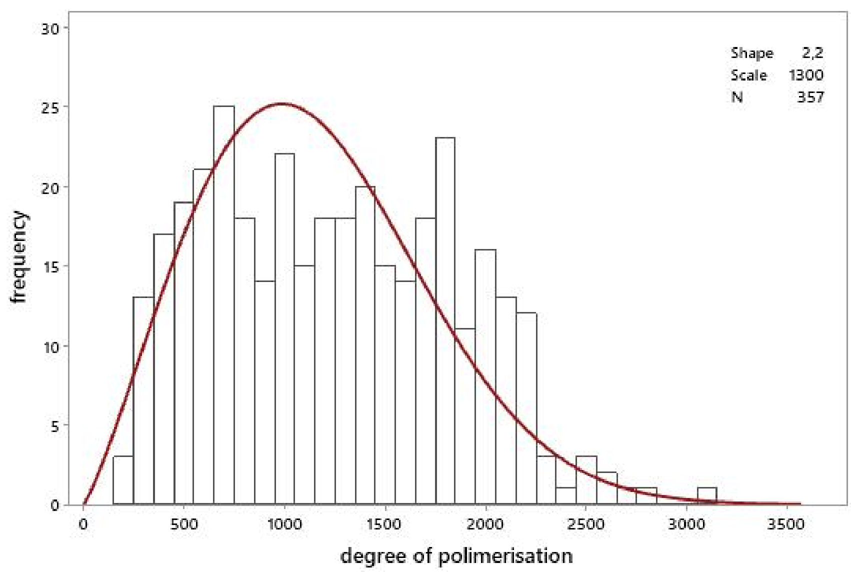


FIGURE 6.11: DP values of the Dutch archives reference data set (bins) with distribution (curve) used to generate the DP values of the model data set representing modern collections.

TABLE 6.4: Input values of the normal distribution to generate pH values depending on DP value.

DP	mean	SD
< 1000	5.150	0.5169
$1000 \leq$ DP > 1500	5.704	0.6466
$1500 \leq$ DP $>$2000	6.167	0.6209
≥ 2000	6.34	0.5907

of the Dutch archive collections and the probability distribution used to generate the model data set. The probability distribution was slightly different from the actual data, to intensify some of the characteristics of the type of collections (e.g. the expected low mean value for the DP for acidic collections).

6.5.2.2 pH values generation

The pH values were then created with a randomly generated data function following a normal distribution. Based on the observed pH data in the Dutch archives reference data set, different probability distributions were used for different ranges in the DP value for the modern and mixed collections (see Table 6.4).

To generate the acidic collection, two probabilities were applied:

- DP < 1000: Normal distribution – Mean: 5.072; SD: 0.4779
- DP > 1000: Weibull distribution – Shape: 25; Scale: 5.7

6.5.2.3 Generated model data set

The three generated data sets can be described as follows (Fig. 6.12):

1. Acidic archival collection (C.1): The first data set represents a collection of mostly groundwood paper, solely including papers with a pH value less than 6 (31% of them with a pH value less than 5). Consequently, the DP frequency distribution is positively skewed, with a mode of approximately 700 and a mean DP of 859.
2. Modern archival collection (C.2): In this data set, the DP frequency distribution is slightly positively skewed, and the mode value is approximately 1000 to represent the chemical characteristics of groundwood and bleached

pulp papers dating after 1850. Compared to a mixed collection, the percentage of acidic paper is higher (71% with a pH value less than 6 and 22% less than 5).

3. Mixed archival collection (C.3): This data set represents a collection containing rag, bleached pulp and groundwood paper. This results in a bimodal DP frequency distribution, with a first wave representing acidic paper with a mode value of approximately 600 and a second wave for rag paper and good quality modern paper with a higher mode value of approximately 1800. Regarding the pH, 61% of the collection had a pH value less than 6, and 17% had a pH value less than 5.

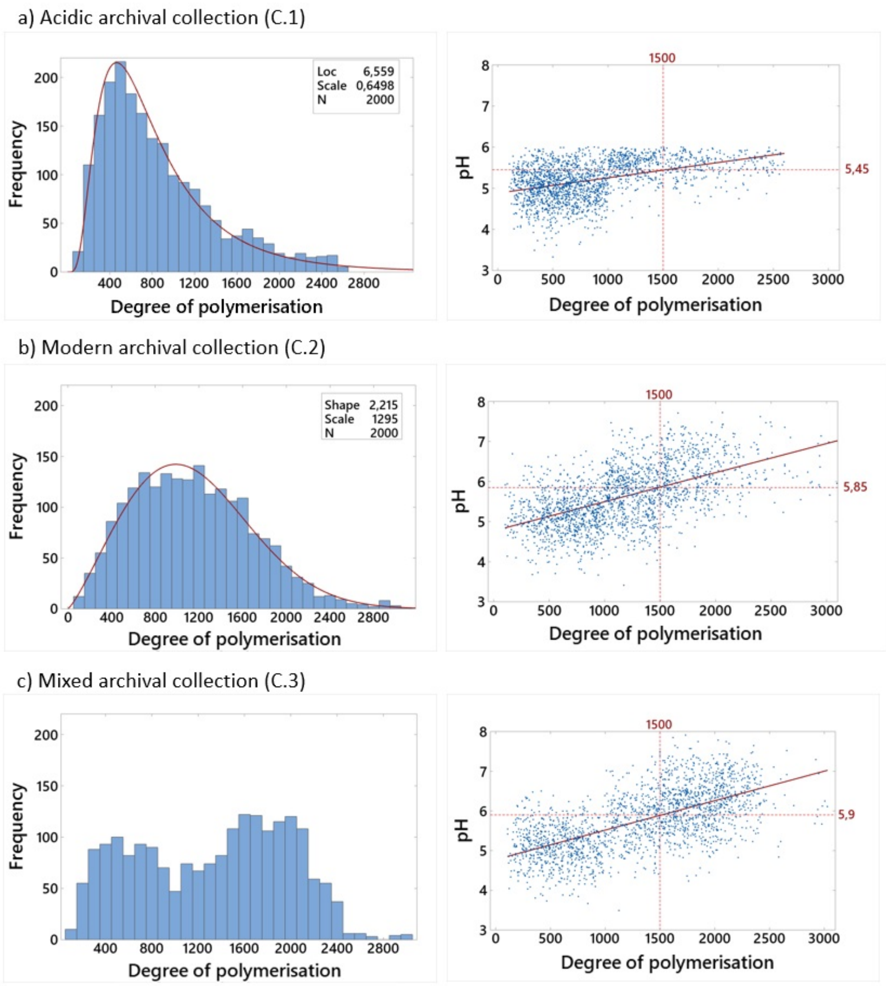


FIGURE 6.12: Generated data sets for a) acidic (C.1), b) modern (C.2) and c) mixed (C.3) collections. The graphs show the DP frequency distribution and the individual values of DP and pH of a population of 2.000 agents.

In addition to the acidic collection described above, which represents archival collections, two more data sets were generated to characterise acidic library collections dating from 1850. In some of these collections, up to 90% of the printed books were acidic, with low folding endurance (Bajzíková et al.; 2008; Vinther Hansen and Vest; 2008). Figure 6.13 shows the DP frequency distribution of a library collection (L.1), with a mean DP of 444 and a second library collection (L.2) with an even lower mean DP of 398. The acidic archival collection (C.1) shown in Figure 6.12 is also included as a comparison.

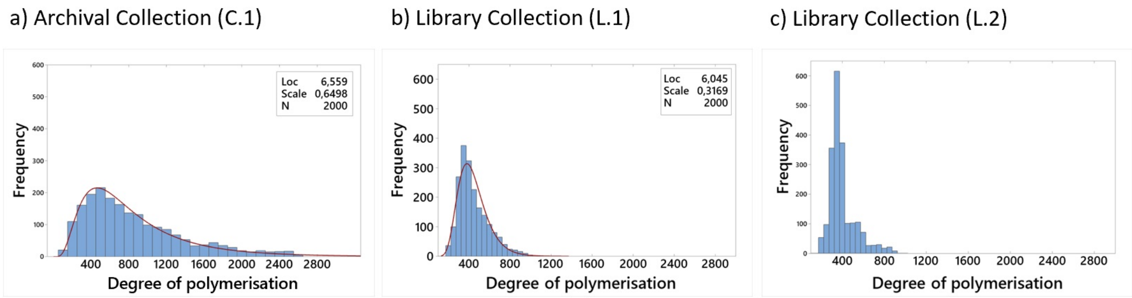


FIGURE 6.13: Generated data sets for three acidic collections with different DP frequency distributions of a population of 2.000 agents.

6.6 Exploring uncertainty

As seen in Chapter 2 , different types of uncertainties have been defined that should be accounted for in decision making. When dealing with mathematical models, two major sources of uncertainty pertain to the model parameters (*parametric uncertainty*) and the model structure (*structural uncertainty*) (Bilcke et al.; 2011). While parametric uncertainty relates to the values of the input parameters, structural uncertainty arises from the assumptions inherent in the model structure. The causes of uncertainties can be found in the chaotic behaviour due to non-linearity or sensitivity to initial conditions, the stochasticity and variability present in the model (*aleatory uncertainty*) and lack of knowledge of the model (*epistemic uncertainty*) (Oberkampf et al.; 2002). An inherent part of modelling is the examination and reporting of an outcome's uncertainty. For example, as in the case of health care decisions models (Briggs et al.; 2012), stochastic/aleatory uncertainty may refer to the random variability in outcomes between identical individuals or the variability in input values due to the heterogeneity of individuals' characteristics. This uncertainty is usually represented as probability distributions that can take on values in a known range. Some uncertainties, such as aleatory uncertainty, can be tested, measured and reported. Although it is clear that this uncertainty is just one of the uncertainty types present in models, the next section will only focus on a few examples of aleatory uncertainty, as example of how sensitivity analysis can be applied to explore and mitigate uncertainty.

In the APM model, the aleatory uncertainty is present due to not only the variability in the input values of the agents with their characteristics (e.g. pH value or the number of access) but also the aleatory combination of different behaviours and past experiences of the agents (pH value *and* number of accessed *and* digitised). Due to the stochastic features, the outcome of the two experiments may be quite different. The difference in the experiments depends on the extent of the stochastic variation in the model and the number of agents in the model.

In the next sections, we explore the importance of the number of agents on the stochastic uncertainty of the results and the parametric uncertainty, taking the example of pH values as an input parameter to achieve confidence in the results.

6.6.1 Number of agents and stochasticity

The first step in ABM is to build a population with agents. The number of agents needs to be sufficient to reproduce the dynamic behaviour of the model. However, the number of agents, together with the complexity of the model, will determine the run time. To establish the correct number of agents for the APM model, we take the model built for the SAA as an example. The first experiment was designed to compare the run time depending on the number of agents. As indicated in Figure 6.14, this relationship is exponential.

In Chapter 5 (Fig. 5.23), we explored the impact of the number of agents on the variability of the output in repeated runs, concerning the access part of the APM model. From that experiment, we concluded that above 4,000 agents, the expected variability remains similar and that models containing more than 4,000 agents reproduce the expected outcome accurately.

However, it is important to note that the expected variability may differ for each output. To demonstrate this point, we compared the number of metres of the collection in critical, poor and fair mechanical condition after a time horizon of 100 years. Contrary to the results of the access part of the model, the output for wear and tear are notably similar, regardless of the number of agents (Fig. 6.15).

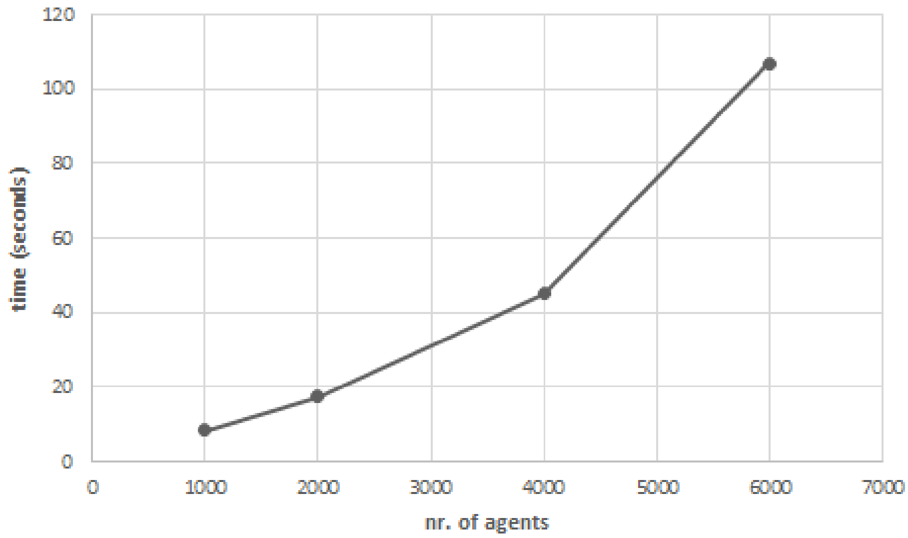


FIGURE 6.14: Run time of an experiment with a time horizon of 100 years depending on the number of agents in the model.

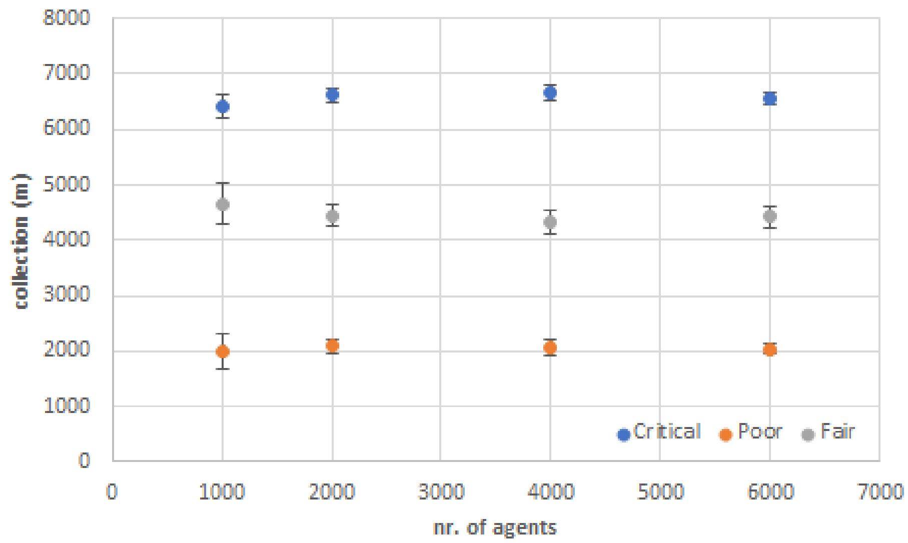


FIGURE 6.15: Output of the number of metres of the SAA collections in critical, poor and fair condition after a time horizon of 100 years generated by models with different number of agents. Error bars indicate the standard deviation.

Only the results of the model with 1,000 agents have a slightly higher standard deviation than the other models.

6.6.2 Parametric uncertainty

Parametric uncertainty is associated with each input parameter of the dose–response function of the chemical degradation and the wear and tear function. A good example to explore parametric uncertainty is pH, as it is an essential input to model the acid-catalysed degradation of the paper. First, we assessed the sensitivity of the dose–response function to pH. Then we used the model to explore the uncertainty of the frequency distributions of pH values after deacidification.

For the experiments presented in the next sections, we created a population of 2,000 agents. This implies that in the case of a population representing a collection of 50,000 linear metres, an agent is equivalent to 25 metres of archival records and that the minimal amount of metres to be selected for deacidification per year is 25 metres, or in the case of a collection of 15,000 linear metres, one agent is 7.5 metres.

6.6.2.1 Sensitivity dose–response function to pH

As seen in section 6.1.4, depending on the measurement method, a slightly different pH value can be obtained from the same paper sample. In addition, the instrumental error should be also taken into account when evaluating the uncertainty associated with pH values. For example, in this study, pH values were collected using the SurveNIR instrument. The reported SurveNIR instrumental errors are 0.71 and 0.63 for the pH of rag paper and modern paper, respectively (Brown et al.; 2020).

To assess the impact of the uncertainty associated with pH, we performed a sensitivity analysis. We designed three experiments. The parameters were kept the same across all three experiments, except for the pH values, which were increased and decreased by 0.5 from the baseline value, based on existing literature (see section 6.1.4). The experiments were conducted on the acidic collection (C.1). Figure 6.16 shows the percentage of the collection in critical condition (DP below 300) for the three experiments. The results showed that whereas the difference from the baseline is ±5 after 200 years, the difference increases up to ±14 at the end of the run.

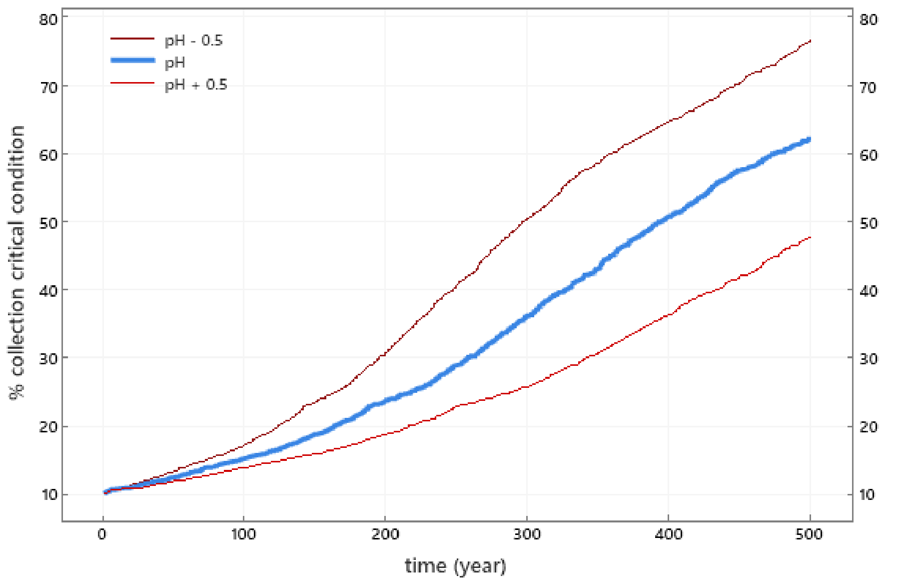


FIGURE 6.16: Percentage of the collection in critical condition (DP below 300) depending on pH input values to explore pH uncertainty.

We also conducted the same experiment on the mixed collection (C.3). The maximum difference between the runs was ca. ±10. One of the main reasons that might explain the difference observed over time, as well as between the types of collections, is the exponential effect of the pH on the DP loss (Fig. 6.17). A 0.5 difference in the lower region (for example 4 and 4.5) will result in a difference of expected DP loss of 100 in 100 years. Meanwhile, in the higher region of the pH (for instance 6.5 and 7), the difference is 26. The level of uncertainty depends on the input values (heteroscedastic uncertainty): a higher uncertainty is expected in collections with a higher percentage of acidic papers. Unfortunately, for us, the group with a higher uncertainty is also the group that we are more interested in due to the urgency of preservation actions.

This sensitivity analysis stressed Briggs' observation that 'the value of a model-based analysis lies not simply in its ability to generate a precise point estimate for a specific outcome but also in the systematic examination and responsible reporting of uncertainty surrounding this outcome and the ultimate decision being addressed' (Briggs et al.; 2012, p.835). In our case, the aim of the model is not to predict but to compare the impact of different preservation scenarios. Therefore, although we are reporting on the outcome numerically, it needs to be stressed that the predicted estimates should only be for qualitative analysis to identify parameters with a higher impact on the output.

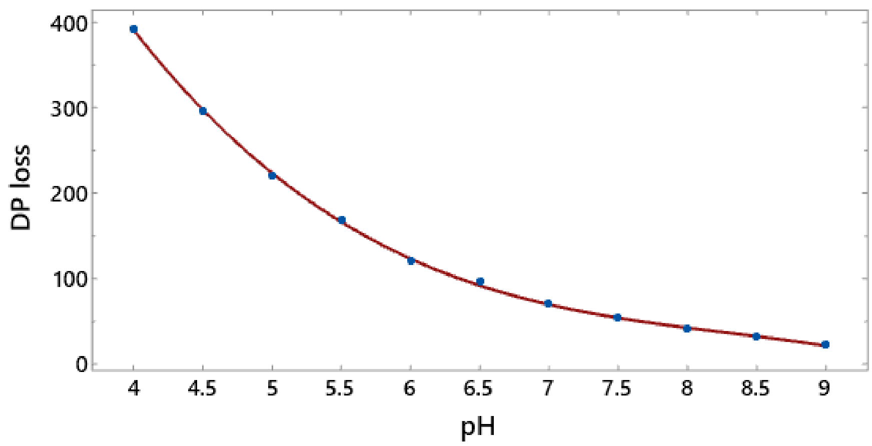


FIGURE 6.17: Expected DP loss in 100 years depending on the pH value of the record, assuming storage conditions of 18 °C and 50% relative humidity, according to the dose–response function developed for acidic catalysed degradation of cellulose (initial DP 1000) developed by Strlič et al. (2015b).

6.6.2.2 Distribution of pH after deacidification

An interesting feature of the software used is that any probability distribution can be created. In the following experiment, we tested two frequency distributions to investigate whether the uncertainty regarding the pH after deacidification should be taken into account. After deacidification, treatment is expected so that the pH value of the treated paper will be higher than 7. However, as discussed in section 6.1.4, in practice, different frequency distributions have been found, depending on the type of paper and method. To test the impact of the pH values after deacidification, we conducted two experiments: one using the uniform distribution with values between 7 and 9; and the other with a customised probability distribution based on the reported pH values after deacidification by Ahn et al. (2011). The experiment was conducted on the archival acidic collection (C.1), and 250 metres were deacidified per year. This resulted in the deacidification of the entire collection after 250 years. The results were evaluated after a time horizon of 500 years.

Figure 6.18 shows the distribution of pH values after deacidification that resulted from the two probability distributions. The same figure also shows the frequency distribution of DP values after 500 years. We observed that the distribution as the mean DP value is the same between the two probability distributions. From this experiment, we can conclude that the probability distribution of the pH after deacidification has a limited impact on the variability of the simulation-generated output.

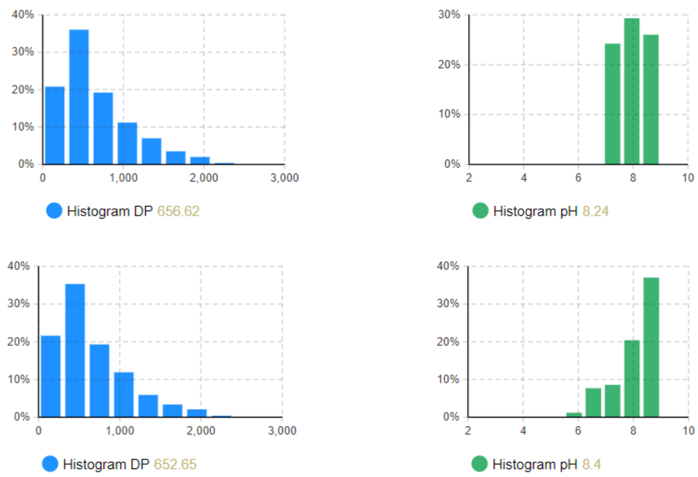


FIGURE 6.18: Frequency distribution of the DP values after 500 years, as result of the application of a uniform (*top*) customised distribution of pH values (*bottom*)

In view of these results, the uniform distribution will be used as default to model deacidification actions in the model.

6.7 Exploring preservation management scenarios

In this section, we explore how the model can be applied and how the findings can be used to inform preservation management decisions. First, we analysed the importance of the selection criteria for deacidification, regarding the pH and DP values for the selection. After identifying the best selection criteria for deacidification, we explored how similar preservation outcomes can be achieved by deacidifying or improving the storage environmental conditions. Lastly, we assessed the long-term effect of delaying the decision to implement a certain preservation strategy.

6.7.1 Effect of selection criteria for deacidification treatment

First, we analysed the importance of the selection criteria for deacidification. Although paper with a pH less than 7 is appropriate for deacidification, institutions opt to prioritise papers with a lower pH value, with 5.5 as the upper limit (Porck; 2006; Shenton; 2006). Similarly, some institutions only select records for the deacidification treatment when paper is still strong and flexible enough and exclude brittle paper that will not benefit from the treatment, as long as strengthening is not part of the treatment (Grossenbacher; 2006).

In the first experiment, we explored the long-term effect of a deacidification strategy that prioritises the most acidic records. In the case of acidic and modern collections, the rate of the treatment, 100 m per year for 100 years, resulted in the deacidification of 20% of the collections. In the case of mixed collections, the treatment could be solely implemented on 17% of the collection, as this is the total percentage of the collection with a pH less than 5.

TABLE 6.5: Comparison of the effect of a deacidification treatment on the percentage of the collection unfit for purpose after 500 years when different selection criteria for deacidification were applied. The last column shows the difference in the end results between the baseline and experiment.

Type collection	Experiment	pH	DP <300	Collection deacidified (%)	Unfit for purpose (%)	Difference Baseline-Exp. (%)
C.1: Acidic	Baseline 1	< 6	included	20	52.9	0
	Exp. 1.1	< 5	included	20	46.9	6
	Exp. 1.2	< 6	excluded	20	51.9	1
	Exp. 1.3	< 5	excluded	20	44.9	8
C.2: Modern	Baseline 1	< 6	included	20	29.1	0
	Exp. 1.1	< 5	included	20	20.5	9
	Exp. 1.2	< 6	excluded	20	28.7	0.4
	Exp. 1.3	< 5	excluded	20	19.6	9.5
C.3: Mixed	Baseline 1	< 6	included	17	25.2	0
	Exp. 1.1	< 5	included	15	19.7	5.5
	Exp. 1.2	< 6	excluded	17	24.6	0.6
	Exp. 1.3	< 5	excluded	15	19.6	5.6

The results indicated that, for all three types of collections, lowering the pH value to 5 as a selection criterion resulted in a lower percentage of collections in critical conditions in the long term (Table 6.5). Modern collections seem to benefit the most from this strategy. When the DP value was used as a selection criterion, the results indicated that excluding items that are too brittle has little impact on the overall effectiveness of the treatment. In summary, in these scenarios, using pH as a selection criterion for deacidification is a better strategy than using the DP.

The experiments showed that prioritising the most acidic records, with a pH value less than 5, had a noticeable positive effect on the effectiveness of the treatment. In this study, effectiveness was measured in terms of preventing records from reaching the critical DP of 300. Naturally, by narrowing the records eligible for deacidification, the probability of selecting records that would benefit the most increases.

Although the benefits of a comprehensive selection are clear, such a detailed selection can be extremely time-consuming, particularly when inventory numbers are a composite of different types of paper. Institutions must, therefore, balance out the process of selection against the chance of treating part of the collections with a lower priority, leading to higher deacidification costs.

6.7.2 Effect of deacidification vs storage conditions

In the following series of experiments, preventive preservation measures related to the storage conditions were compared to the conservation measure of deacidification of the collections. In the first experiment, we tested the best results that can be achieved if it were feasible to deacidify the whole collection within one year and with no delay. The tests were conducted on the three acidic collections, C.1, L.1 and L.2 (Table 6.6). At the beginning of the run (time 0), the percentage of the collection with a DP less than 300 was 10% for each of the collections. In the case of the acidic collection with a higher DP average (C.1), if deacidification was conducted, then just 14% of the collection would reach the state unfit for purpose. The same results were obtained if the collection is kept in storage conditions of 5 °C and 45% RH, which were also equivalent to the results when the storage conditions were 8 °C and 40% RH. However, in the case of the other two acidic collections, we observed that the lower the DP average of the collection, the larger the differences in effectiveness between the preservation strategies. In these two cases (L.1 and L.2), even if the whole collection was deacidified, more than 25% of the collection reached a DP less than 300 after 500 years. For these types of collections, to achieve a better level of protection, cold storage (storage temperature below 5 °C) seems to be the most effective option.
When considering archive collections as a whole, it would not be feasible to implement deacidification on such a scale, within a short time, or to change to cold storage. Therefore, in the following experiment, we explored more realistic scenarios: a small change in the repository conditions, assuming 18 °C and 50% RH

TABLE 6.6: Comparison of strategies for three acidic collections after a time horizon of 500 years. The last column shows the difference in the end results between the baseline and the experiment.

Type collection	Experiment	Experiment description	Unfit for purpose (%)	Difference Baseline-Exp. (%)
C.1	Baseline 2	18 °C – 50%	62.5	0
	Exp. 2.1	16 °C – 40%	31.2	31.3
	Exp. 2.2	5 °C – 40%	12.5	50.0
	Exp. 2.3	deacidification	14.4	48.1
L.1	Baseline 2	18 °C – 50%	94.7	0
	Exp. 2.1	16 °C – 40%	68.2	26.5
	Exp. 2.2	5 °C – 40%	20.1	74.6
	Exp. 2.3	deacidification	28.9	65.8
L.2	Baseline 2	18 °C – 50%	97.3	0
	Exp. 2.1	16 °C – 40%	80.2	17.1
	Exp. 2.2	5 °C – 40%	25.4	71.9
	Exp. 2.3	deacidification	39.4	57.9

as baseline or deacidification of a part of the collection. The results showed visible differences at an early stage of the experiment conducted on acidic collections (C.1) (Fig. 6.19). In the case of the other two collections (C.2 and C.3), the differences became visible after 100 years. The largest impact on the preservation of the collection was observed when temperature and RH were lowered, particularly in the case of acidic collections (C.1) (Table 6.7). In the case of mixed collections (C.3), similar results were obtained between the deacidification of 15% of the collection and lowering the temperature and RH to 16 °C and 45%.

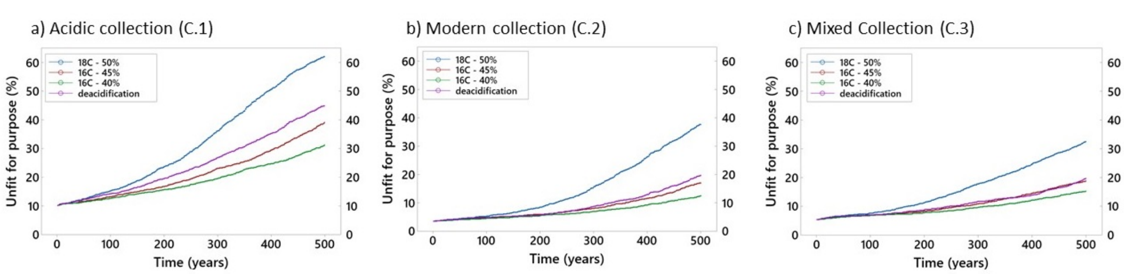


FIGURE 6.19: Results of the four preservation strategies on (a) acidic, (b) modern and (c) mixed collection, in the time horizon of 500 years.

TABLE 6.7: The preservation strategies presented in Figure 6.19 are compared at the end of the time horizon of 500 years. The last column shows the difference in the end results between the baseline and the experiment.

Type collection	Experiment	Experiment description	Unfit for purpose (%)	Difference Baseline-Exp. (%)
C1: Acidic	Baseline 3	18 °C – 50%	62.5	0
	Exp. 3.1	16 °C – 45%	39.1	23.4
	Exp. 3.2	16 °C – 40%	31.2	31.3
	Exp. 3.3	20% deacidification	44.9	17.6
C2: Modern	Baseline 3	18 °C – 50%	37.8	0
	Exp. 3.1	16 °C – 45%	16.9	20.9
	Exp. 3.2	16 °C – 40%	12.4	25.4
	Exp. 3.3	20% deacidification	19.6	18.2
C3: Mixed	Baseline 3	18 °C – 50%	32.5	0
	Exp. 3.1	16 °C – 45%	18.7	13.8
	Exp. 3.2	16 °C– 40%	15.6	16.9
	Exp. 3.3	15% deacidification	19.6	12.9

These experiments illustrate that similar results can be obtained by lowering temperature and RH and by deacidifying a certain percentage of the collection, depending on the type of the collection, i.e. the percentage of acidic paper within a collection. In the next experiment, we calculated how many metres need to be deacidified during a period of 100 years to obtain the same results as when the storage conditions of 16 °C and 40% RH are met. The experiments showed that it is not feasible to achieve these results by deacidifying records with a pH < 5 exclusively. For example, in the case of the acidic archival collection (C.1), if only records with a pH less than 5 were deacidified (37% of the collection) then 38% of the collection would reach a DP less than 300. To achieve results consistent with those through storage conditions of 16 °C and 40%RH (ca. 31% in critical condition), records with a pH value up to 6 need to be included. Table 6.8 shows the percentage of the collection during 100 years that needs to be deacidified to achieve the same results as storage conditions of 16 °C and 40% RH.

TABLE 6.8: Number of metres that need to be deacidified per year during a period of 100 years to achieve similar results on the preservation of the collection of 50.000 metres compared to storage conditions of 16°C and 40% RH.

Type collection	Deacidification (m/year)	Collection deacidified (%)
C.1: Acidic	350	70
C.2: Modern	300	60
C.3: Mixed	225	45

We also explored various less demanding scenarios, for example decreasing the temperature to 16 °C and the RH to 40%. All types of collections would benefit from a small decrease in the temperature and RH in the repositories, but when dealing with acidic collections, differences in preservation were already visible within the time span of the next 100 years. The improvement was most noticeable for acidic collections, with a mean DP of 800. However, we observed that when dealing with collections with a mean DP less than 500, a small decrease in temperature and RH had limited effects, and more rigorous measures, such as cold storage, were the only option to keep the percentage of brittle paper lower than 25% of the collection.

Compared to deacidification, changing the environmental conditions seemed to be more effective, as the whole collection benefited from this preservation measure. Moreover, when the concern is the preservation of the collections as a whole, deacidification might turn into a highly expensive option. For example, institutions holding small-size mixed collections have the option to deacidify a relatively small part of their collection (ca. 15%) in 100 years, achieving similar results as when changing the environmental conditions to 16 °C and 45% RH, if the selection criterion for deacidification is pH less than 5, and records with DP less than 300 are excluded. If the selection criterion is pH less than 6, then the percentage to be deacidified increases to 35%. To obtain the same level of effectiveness as storage conditions of 16 °C and 40% RH, the percentage of collection that needs to be treated increases from 45% for mixed collections to 70% for acidic collections. Table 6.9 shows an overview of the simulations comparing the effect of different environmental conditions and deacidification when records with a pH less than 6 are deacidified.

TABLE 6.9: Comparison of environmental strategies versus deacidification for acidic collections (L.1 and C.1) and modern and mixed collections (C.2 and C.3) after a time horizon of 500 years. First column shows different environmental strategies. Second column shows the percentage of the collection that should be deacidified to get the same results as the environmental strategies. Records with a pH value less than 6 were selected for deacidification. The duration of the deacidification treatment was 100 years.

Library collection (L.1)		
Storage conditions	Collection deacidified (%)	Unfit for purpose (%)
18 °C – 50%	-	95
16 °C – 45%	30	77
16 °C – 40%	45	68
10 °C – 45%	100	39
5 °C – 40%	-	20
Acidic archival collection (C.1)		
Storage conditions	Collection deacidified (%)	Unfit for purpose (%)
18 °C – 50%	-	62
16 °C – 45%	50	39
16 °C – 40%	70	31
10 °C – 45%	100	17
5 °C – 40%	-	13
Modern archival collection (C.2)		
Storage conditions	Collection deacidified (%)	Unfit for purpose (%)
18 °C – 50%	-	38
16 °C – 45%	45	17
16 °C – 40%	60	12
10 °C – 45%	71	6
5 °C – 40%	-	4
Mixed archival collection (C.3)		
Storage conditions	Collection deacidified (%)	Unfit for purpose (%)
18 °C – 50%	-	33
16 °C – 45%	35	19
16 °C – 40%	45	16
10 °C – 45%	-	8
5 °C – 40%	-	7

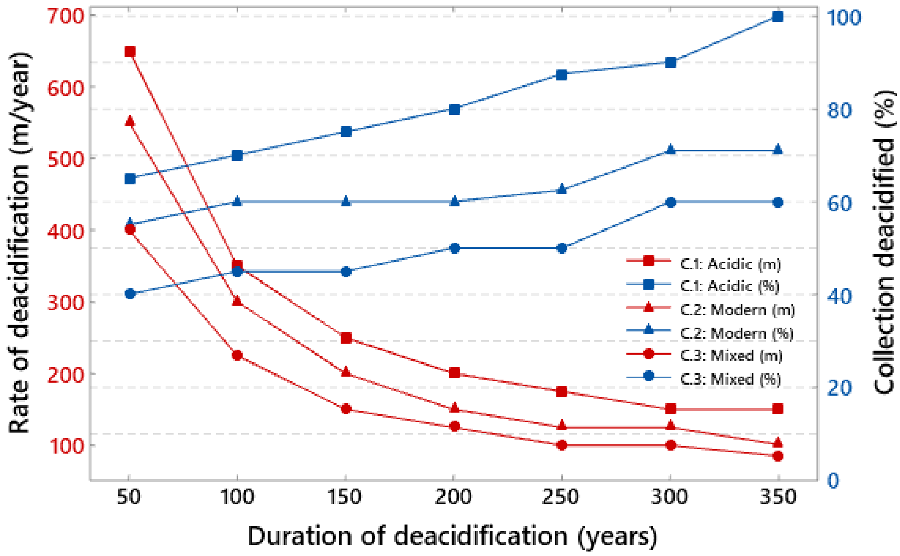


FIGURE 6.20: A comparison of strategies for deacidification with the same effectiveness as storage conditions at 16 °C and 40% RH. The strategy is described by the number of metres that need to be deacidified per year (y left axis) and for how long (x-axis). The resulting percentage of the deacidified collection after the treatment is also shown (y-right axis).

6.7.3 Effect of delaying deacidification

In the prior experiment (Table 6.9), several metres were deacidified during a period of 100 years to achieve similar results as those achieved through storage conditions of 16 °C and 40% RH. However, similar results can also be achieved by following other strategies. Fig. 6.20 shows the number of metres that need to be deacidified in each of the archival collections (C.1 – C.3) if another time span for the treatment is chosen, to obtain the same results as stated in Table 6.9. The longer the time span, the lower the number of metres that need to be deacidified per year. However, as the collection degrades over time and the part that has not been treated yet degrades even faster, slowing the treatment rate will result in a higher percentage of the collection requiring treatment to obtain the same results.

In the last experiment, we explored the consequences of delaying the decision to deacidify. As an example (Fig. 6.21), we delayed in steps of 50 years the strategy of deacidifying the collections within 50 years (first strategy in Fig. 6.20). The effect of the delay in implementing a preservation measure is most noticeable in acidic collections, where the percentage of metres unfit for purpose increases linearly with the delay. The effect is also clear in the case of modern collections, but

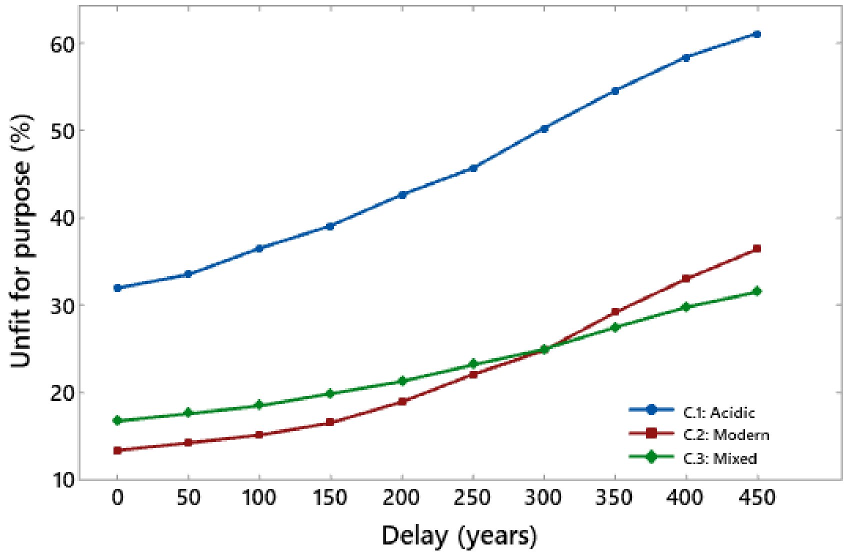


FIGURE 6.21: Effect of the strategy of deacidifying part of the collection in 50 years (delay 0 years) and the effect of the same strategy if this implementation is delayed in the steps of 50 years.

rather than being linear, the consequences are more severe further in the future. The results were similar to the delayed implementation of lowering the temperature and RH in the repositories.

Preservation planning is generally driven by the available budget within the operational planning of 3–5 years (Bell et al.; 2018). However, to save financial resources in the long-term, institutions that choose deacidification as a conservation strategy need to keep in mind that the most efficient strategy is to deacidify as much and as soon as possible. Through simulations, we could examine the effect of preservation measures over time, while degradation of the collection takes place. We calculated that if an institution opts for a lower rate of deacidification, then deacidification treatment needs to be carried out for a longer period to obtain the same results, and consequently, a higher percentage of the collection will be deacidified at a higher cost.

6.8 Conclusions

In this chapter, we described the part of the model that simulates the effect of preservation measures on the lifetime of paper collections. The model consists of two interrelated sub-parts: (1) the modelling of chemical degradation of paper collections (2) acts as an input to model the accumulation of wear and tear.

The modelling of the chemical degradation has been extensively studied; therefore, existing models, such as the dose–response function, for acid-catalysed degradation could be easily integrated into the APM model. The resulting model is a user-friendly model that allows end-users to explore the effect of preservation measures such as deacidification and environmental storage conditions. Supported by the diverse possibilities offered by the software, the model can run simulations of what-if scenarios with a high level of detail, where each parameter regarding the selection of the agents to be treated, the treatment level and the treatment time span can be adjusted prior to the run or at any given point during the run. The results of the computational experiments are directly applicable to decision-making, thereby having the potential to influence the collection management policies within institutions.

Regarding the modelling of wear and tear, this area of research is in its early stages; therefore, it is less straightforward how it should be included in the APM model compared to the chemical degradation model, where assumptions and simplifications can be more clearly justified. For instance, the wear and tear function calculates the number of tears per 100 sheets, and for now, the number of uses calculated in the access part of the APM model is for an average number of records (as an agent is equivalent to a certain number of records). To better apply the model to real-world settings, the number of agents in the model should be increased (e.g. one agent equivalent to one inventory number), but that will considerably increase the run time of a computational experiment. In addition, the agents should include more detail levels (e.g. number of sheets per record). However, the question remains whether such levels of detail are needed to inform end-users on the benefits of certain preservation measures in the particular case of their collections. For now, the APM model is a simplified model where agents, representing a certain number of records, have the same amount of pages, and the DP and number of uses are randomly assigned. However, if desired, more variables could be easily added to the model to include a higher level of detail. One of these variables could be, for example, ‘date’. The date when the record was made could then be used to determine the probability of certain pH and DP values (e.g. records between 1850 and 1860 would have a higher probability of low DP and pH values) or the probability of being heavily requested (e.g. records dating from 1900 to 1950).

Although an ABM approach is used in this part of the model, the variability in the output regarding the presence of stochastic elements, and in relation to the numbers of agents, was limited. Therefore, for the preservation part of the model, it is not necessary to run the same experiment repeatedly. While the uncertainty related to the stochastic elements is small, this is not the case for the parametric uncertainty, particularly, the uncertainty related to the pH parameter. These findings are a reminder that these models should not be used as predictive tools, but as comparison options, due to the greater uncertainty as the time horizon increases. The model was also a valid tool to investigate whether parameter uncertainties matter. For instance, we tested the effect of different distributions to assign the pH value after deacidification treatments. The computational experiments showed that although there is a difference on the individual level of agents, the difference is not noticeable for the collection as a whole.

In this chapter, the challenges regarding the use of observational studies to gather data to eventually inform the (mathematical) models and/or validate the existing models have been addressed. It is difficult to justify the actual condition of the records when so little is known about the previous use of the collection during a significant period of time, and in view of our experience, the required sample size should be considerably increased in future studies. Nonetheless, over time, observational studies have the potential to become a rich source of information for models, which so far mostly rely on experimental data.

References

Ahn, K., Henniges, U., Blüher, A., Banik, G. and Potthast, A. (2011). Sustainability of mass deacidification. Part I: Concept, selection of sample books and pH-determination, *Restaurator* **32**(3): 193–222.

Andres, H., Blüher, A., Grossenbacher, G., Reist, M. and Wälchli, M. (2008). The Papersave Swiss-Process quality control and efficacy, *Restaurator* **28**: 3–28.

Bajzíková, M., Hanus, J. and Bukovský, V. (2008). Characterisation of the Slovak national library collections, *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, pp. 51–52.

Balazic, A., Habicht, S., Smodis, M., Kolar, J. and Strlic, M. (2007). Extending the useful life of paper. Evaluation of the effect of various preservation actions., *in* T. Padfield and K. Borchersen (eds), *Museum Microclimates*, National Museum of Denmark, pp. 39–41.
URL: *http://www.conservationphysics.org/mm/balazic/balazic.pdf*

Bansa, H. (1998). Aqueous deacidification with calcium or with magnesium?, *Restaurator* **19**: 1–40.

Barański, A., Łagan, M. and Łojewski, T. (2005). Acid-catalysed degradation, *in* M. Strlič and J. Kolar (eds), *Ageing and stabilisation of paper*, National and University Library, Ljubljana, chapter 6, pp. 85–100.

Barrow, W. (1974). *Permanence/Durability of the book—VII. Physical and chemical properties of book papers, 1507-1949.*, W.J. Barrow Research Laboratory Inc., Richmond, Virginia.

Baty, J. W., Maitland, C. L., Minter, W., Hubbe, M. A. and Jordan-Mowery, S. K. (2010). Deacidification for the conservation and preservation of paper-based works. A review., *BioResources* **5**(3): 1955–2023.

Beentjes, G. (2016). Ontzuren ... of niet.
URL: *https://www.metamorfoze.nl/nieuws-agenda/ontzuren-niet*

Bell, N., Cassar, M. and Strlič, M. (2018). Evidence for informed preservation planning and advocacy. A synoptic view, *Studies in Conservation* **63**(sup1): 8–14.

Bilcke, J., Beutels, P., Brisson, M. and Jit, M. (2011). Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: A practical guide, *Medical Decision Making* **31**(4): 675–692.

Blüher, A. and Großenbacher, G. (eds) (2006). *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern.

Brandis, L. (1993). A comparision of methods used for measuring the pH of paper., *AICCM BUlletin* **18**: 7–17.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J. and Paltiel, A. D. (2012). Model parameter estimation and uncertainty. A report of the ISPOR-SMDM modeling good research practices task force-6, *Value in Health* **15**(6): 835–842.

Brown, N., Coppola, F., Modelli, A., Amicucci, F., Lichtblau, D. and Strlič, M. (2020). Non-destructive collection survey of the historical Classense Library. Part I: Paper characterisation, *Heritage Science* **8**(1): 1–11.

Calvini, P. (2005). The influence of levelling-off degree of polymerisation on the kinetics of cellulose degradation, *Cellulose* **12**(4): 445–447.

Calvini, P. (2012). The role of the Ekenstam equation on the kinetics of cellulose hydrolytic degradation, *Cellulose* **19**: 313–318.

Cedzová, M., Gállová, I. and Katuščák, S. (2006). Patents for paper deacidification, *Restaurator* **27**(1): 35–45.

Daly, L. E. and Bourke, G. (2007). *Statestical analysis, interpretation and uses of medical statistics*, 5th edn, Blackwell Science Ltd, Oxford.

Daniels, V. (2006). Paper, *in* E. May and M. Jones (eds), *Conservation Science*, RSC Publishing, Cambridge, chapter 3.

Di Pietro, G., Ligterink, F., Porck, H. and de Bruin, G. (2016). Chemical air filtration in archives and libraries reconsidered, *Studies in Conservation* **61**(5): 245–254.

Ding, H. (2009). Degradation phenomenology and life modelling of paper., *in* S. I'Anson (ed.), *Advances in Pulp and Paper Research, Trans. of the XIVth Fund. Res. Symp. Oxford*, FRC, Manchester, pp. 947–1000.

Ding, H. and Wang, Z. D. (2008). On the degradation evolution equations of cellulose, *Cellulose* **15**(August 2007): 205–224.

Duran, C., Grau-Bové, J. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part I: exploring the concepts of reliability and epidemiology, *Heritage Science* **7**(1).

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2017). Data mining in collections: From epidemiology to demography, *ICOM-CC 18th Triennial Conference, 2017 Copenhagen*.

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part II: an epidemiological study, *Heritage Science* **7**(11).

Emsley, A. M. and Stevens, G. C. (1994). Kinetics and mechanisms of the low-temperature degradation of cellulose, *Cellulose* **1**(1): 26–56.

Gess, J. M. and Rodriguez, J. M. (eds) (2005). *The sizing of paper*, third edit edn, Tappi Press, Atlanta.

Grossenbacher, G. (2006). Paper deacidification as a measure for preserving originals, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern, pp. 7–20.

Gurnagul, N., Howard, R., Zou, X., Uesaka, T. and Page, D. (1993). The mechanical permanence of paper: a literature review, *J. Pulp Pap. Sci.* **19**: 160–166.

Hanus, J., Katuščák, S., Čeppan, M. and Bukovský, V. (2006). Research on paper deacidification in Slovakia, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, pp. 75–91.

Hanus, J. and Mináriková, J. (2002). Deacidification without equipment and money – dream or reality ?, *13th ICOM Triennial Meeting Rio de Janeiro Preprints*, pp. 603–608.

Hubbe, M. A. (2006). Paper's resistance to wetting. A Review of internal sizing chemicals and their effects, *BioResources* **2**(1): 106–145.

Joel, A., Indictor, N., Hanlan, J. and Baer, N. (1972). The measurement and significance of pH in paper conservation, *Bulletin of the American Group. International Institute for Conservation of Historic and Artistic Works* **12**(2): 119–125.

Kolar, J. (2009). SSPI-006584. PaperTreat. Evaluation of mass deacidification processes. 3rd report, *Technical report*.

Kyujin, A. (2013). *Sustainability of mass deacidification of library objects*, PhD thesis, Universität für Bodenkultur Wien.

Ligterink, F. and Di Pietro, G. (2018). The limited impact of acetic acid in archives and libraries, *Heritage Science* **6**(1).

Menart, E. (2013). *Indoor pollution in archival collections in the context of a changing environment*, PhD thesis, University College London.

Menart, E., De Bruin, G. and Strlič, M. (2011). Dose–response functions for historic paper, *Polymer Degradation and Stability* **96**(12): 2029–2039.

Menart, E., De Bruin, G. and Strlič, M. (2014). Effects of NO_2 and acetic acid on the stability of historic paper, *Cellulose* **21**: 3701–3713.

Michalski, S. (2003). Double the life for each five-degree drop, more than double the life for each halving of relative humidity, *ICOM committee for conservation, 13th triennial meeting Rio de Janeiro preprints vol. 1*.

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V. and Alvin, K. F. (2002). Error and uncertainty in modeling and simulation, *Reliability Engineering and System Safety* **75**(3): 333–357.

Pilette, R. (2004). Mass Deacidification : preservation option for libraries, *IFLA Journal* **30**(1): 31–36.

Porck, H. J. (2006). The Bookkeeper process and its application at the National Library of the Netherlands, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern, pp. 37–42.

Potthast, A. and Kyujin, A. (2017). Critical evaluation of approaches toward mass deacidification of paper by dispersed particles, *Cellulose* **24**: 323–332.

Ramin, M., Blüher, A., Reist, M. and Waelchli, M. (2009). Paper De-acidification, *Journal of Paper Conservation* **10**: 17–25.

Roberts, J. (1996). *The chemistry of paper*, The Royal Society of Chemistry, Cambridge.

Saverwyns, S., Sizaire, V. and Wouters, J. (2002). The acidity of paper: evaluation of methods to measure the pH of paper samples, *13th Triennial Meeting, Rio de Janeiro, 22-27 September 2002: preprints* pp. 628–634.

Shenton, H. (2006). Strategies for mass preservation treatment, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences - tomorrow's perspectives*, Swiss National Library, Bern, pp. 63–74.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and Bruin, G. D. (2015). Damage function for historic paper . Part I : Fitness for use, *Heritage Science* **3**(33).

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015a). Damage function for historic paper. Part II: Wear and tear, *Heritage Science* **3**(1): 36.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015b). Damage function for historic paper. Part III: Isochrones and demography of collections, *Heritage Science* **3**(1): 40.

Strlič, M. and Kolar, J. (eds) (2005). *Ageing and stabilisation of paper*, National and University Library, Ljubljana.

Strlič, M., Kolar, J., Koèar, D., Drnovšek, T., Šelih, V. S., Susiè, R. and Pihlar, B. (2004). What is the pH of alkaline paper?, *e-Preservation Science* **1**: 35–47.

Strlič, M., Kolar, J., Pihlar, B., Rychly, J. and Matisová-Ryvhlá, L. (2001). Initial degradation processes of cellulose at elevated temperatures revisited - chemiluminecence evidence, *Polymer degradation and stability* **72**: 157–162.

Strlič, M., Liu, Y., Lichtblau, D. A., De Bruin, G., Knight, B., Winther, T., Kralj Cigić, I. and Brereton, R. G. (2020). Development and mining of a database of historic European paper properties, *Cellulose* **27**(14): 8287–8299.

Tétreault, J., Bégin, P., Paris-Lacombe, S. and Dupont, A. L. (2019). Modelling considerations for the degradation of cellulosic paper, *Cellulose* **26**(3): 2013–2033.

Tétreault, J., Dupont, A. L., Bégin, P. and Paris, S. (2013). The impact of volatile compounds released by paper on cellulose degradation in ambient hygrothermal conditions, *Polymer Degradation and Stability* **98**(9): 1827–1837.

Vinther Hansen, B. and Vest, M. (2008). The lifetime of acid paper in the collection of the Royal Library, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing*

2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008, Faculty of Chemistry and Chemical Technology, Ljubjana, pp. 38–39.

Whitmore, P. M. and Bogaard, J. (1994). Determination of the cellulose scission route in the hydrolytic and oxidative degradation of paper, *Restaurator* **15**(1): 26–45.

Zervos, S. and Alexopoulou, I. (2015). Paper conservation methods. A literature review, *Cellulose* **22**: 2859–2897.

Zou, X., Uesaka, T. and Gurnagul, N. (1996). Prediction of paper permanence by accelerated aging. Part I: Kinetic analysis of the aging process, *Cellulose* **3**: 243–267.

Chapter 7

Cost

'The economic paradox of digital information is finding the correct financial strategy to collect sufficient revenues to pay for the benefits of digitization.'

Kingma 2000, p. 10

In this chapter, we address cost modelling, the third and last part of the Archival Preservation Management (APM) model. At first glance, cost might seem a straightforwards outcome to calculate. This is true if cost is simply understood as the sum of expenses. Partly as a result of that approach, '[...] many organisational decisions remain opportunistic, driven by funding availability' (Cassar; 1998, p. 5). Budgets generally run over operational planning windows of three to five years (Bell et al.; 2018), encouraging institutions to take decisions based on the sums of total costs over these periods.

Cost can also be approached as one of the sub-parts in a complex system. In a complex system, spending resources by taking a certain action will, directly or indirectly, affect other parts of the model, including their associated costs. For instance, at the beginning of this thesis (Chapter 2) we referred to the example given by Varlamoff (2004) who rightly argued that, in order to make an informed decision about providing collection access by microfilming, it is not sufficient to calculate the costs of microfilming alone. In a complex system, the costs of microfilming would be studied not in isolation, but in relation to other parts of the model, in this case the money saved by reducing remedial treatments as the risk of wear and tear will decrease after the creation of a surrogate. Moreover, by approaching costs as part of a complex system, costs are then calculated for the mid-term and long-term: what might seem a good strategy at the present time, might become expensive in the long-term.

In the APM model, the costs of three major preservation measures are calculated: collection care, energy consumption in repository facilities, and digitisation. Therefore, this chapter begins by examining several models that have been proposed in these three major areas which will help us to identify key elements for the APM model later on. Then, the boundaries and the structure of the model are presented. Similar to the preservation and access part, the APM-cost model has been developed for use both independently or in combination with the other two parts of the APM model. To demonstrate its independent use, we use the model to compare the costs related to the two services of providing collection access, namely digitisation and the reading room. Then, we demonstrate how the APM-cost model can be expanded by using the outputs of the APM-access model. The example shown in the last section of the chapter reports the costs of the preservation strategies discussed in Chapter 6 related to deacidification and environmental conditions in physical repositories.[1]

7.1 Cost model examples

This section provides a brief overview of models types proposed for collection care, digital collections management, and energy consumption.

7.1.1 Collection care

Cost-effectiveness analyses are not uncommon in collection care, as supported by the following three examples:

1. In 1998, Cassar (1998) presented cost–benefits appraisals for collection care. The appraisal consists of building a decision matrix where options are evaluated according to pre-stated criteria, with an assigned weight. Then, the results are evaluated by taking the costs of each option into account.

2. In 2011, Büllow and Brokerhof (2011) explored the applicability of the 'quality-adjusted life years' (QALYs) concept, a model used to measure cost effectiveness in health care, to support collection risk management. An interesting aspect of this approach is that 'collection quality' is measured by incorporating a range of aspects: values, accessibility, development, use, life expectancy and costs. In this cost-effectiveness model, the effectiveness

[1] Parts of Section 7.3.1, including Figure 7.8, have been published in Cristina Duran-Casablancas, Marc Holtman, Matija Strlič, & Josep Grau-Bové, The end of the reading room? Simulating the impact of digitisation on the physical access of archival collections (2022), Journal of Simulation, https://doi.org/10.1080/17477778.2022.2128911

of measures, understood as the product of the quality and life expectancy (read 'risk reduction'), are evaluated against the costs, resulting in annual cost per QALY.

3. Similarly, Michalski and Karsten (2018) presented, in 2018, a cost-effectiveness model based on numerous risk assessments conducted at the Canadian Conservation Institute. In this model, effectiveness is also defined as risk reduction. Cost-effectiveness is calculated in annual terms as the risk reduction (units of fractional loss per year) divided by cost (euros per year).

These three models have several aspects in common. None of them were developed for a specific type of objects, but for general collection care and risk management in heritage collections ranging from museums and historic houses to archives and libraries. They are mostly conceptual models to show how cost-effectiveness can be of use in collection management. As they are broadly oriented, no standard framework for a systematic calculation of the costs is provided. In addition, as cost-effectiveness models, the models do not simply calculate the costs of certain measures, but attend specifically to the relationship between cost and effectiveness. In these models, effectiveness is defined beforehand in order to evaluate scenarios according to pre-determined criteria.

7.1.2 Cost models for digital collections

In contrast to the models seen in the previous section, cost models for digital collections are characterised by a highly detailed and well-defined structure. The aim of these models is to include all costs related to the lifecycle of digital collections, in order to obtain the total process costs of providing digital access to the collections. Most cost models for digital collections follow the terminology of the standard Open Archival Information System (OAIS) (CCSDS; 2012), a reference model that summarises functional entities to ensure the long-term preservation and access to digital information. These functional entities are well defined, published, and well established in the preservation community, and therefore used in most of the cost models for digital collection.

In recent years, there has been considerable interest in the development of cost models for digital collections. A detailed review on these models was a topic of the 4Cproject (*4C Collaboration to clarify the costs of curation*; n.d.), a collaborative effort to clarify the costs of digital curation and preservation. The review of the ten main models available until 2014 can be found in Bogvad et al. (2014), including a critical description and references of each model.

In general, cost models for digital digital collections are comprehensive, with fixed and incidental cost being included for each activity, based on the OAIS functional activity framework. In addition, tools, such as Excel spreadsheets and guidelines are provided. However, as digital preservation is still in active development, new initiatives (Beagrie and Daphne; 2017; NCDD; 2017) have emerged since then.

Of the reviewed models, we found that two models had the potential to be implemented as part of the APM model: the LIFE^3 (the third of Lifecycle Information for E-Literature three project phases) (LIFE; 2010) and DEN Digital cost model (Gillesse et al.; 2010).

Lifecycle Stage	Creation or Purchase	Acquisition	Ingest	Metadata Creation[2]	Bit-stream Preservation	Content Preservation	Access
Lifecycle Elements	Conceive Activity*	Selection	Quality Assurance	Re-use Existing Metadata	Repository Administration	Preservation Watch	Access Provision
	Selection and Preparation*	Submission Agreement	Deposit	Metadata Creation	Storage Provision	Preservation Planning	Access Control
	Transport*	IPR & Licensing	Holdings Update	Metadata Extraction	Refreshment	Preservation Action	User Support
	Digitisation*	Ordering & Invoicing	Reference Linking		Backup	Re-ingest	
	Digitisation QA*	Obtaining			Inspection		
	IPR*	Check-in					

FIGURE 7.1: Model structure of LIFE^3 (Life Cycle Information for E-Literature) (Ayris et al.; 2010)

The LIFE project was a collaborative effort between University College London (UCL), The British Library, and the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow. The model provides costing estimates throughout the lifecycle of digital collections (Fig. 7.1), where lifecycle stages and activities are included. A similar approach had already been proposed by the British Library in 2003 for collection management, in order to 'demonstrate the long-term consequences of what the library takes into its collections, by making explicit the financial and other implications of decisions made at the beginning of the lifecycle for the next 100 plus years' (Shenton; 2003, p. 254). Figure 7.2 visualises the lifecycle stages of physical collections, showing strong similarities to the stages of the digital collections.

FIGURE 7.2: Lifecycle collection management proposed by Shenton (2003).

In the British Library lifecycle model (Shenton; 2003), the costs in each stage are calculated independently, and possible interactions between stages are not included. As the model focuses on management activities, access has been omitted. Therefore, the model presents a static approach to collection management, where potential interactions (e.g. level of information in the catalogue and the frequency of use; conservation needs due to high frequency of use) are not included.

The second model that we would draw our attention to is the DEN Digital cost model (Gillesse et al.; 2010). This cost model was designed to provide an approximation of the costs involved in digitisation projects, from scanning of the collections to making the scans available online. Similar to the LIFE model, an Excel spreadsheet is used to compute the costs. As the model focuses mainly on one process, the digitisation of analogue collections, excluding the preservation of digitally born collections, the model presents a comprehensive, but accessible, calculation of a digitisation project's implementation costs. The component costs are divided into personnel costs and others (mostly equipment-related). One-time costs (e.g. equipment) are spread over the their lifetime, or over the project's duration.

7.1.3 Energy cost models

There is an extensive body of literature focused on museum and historical building environments. A comprehensive literature overview can be found in Lucchini (2016), and Kompatscher et al. (2019), covering various aspects such as energy efficiency, current museum indoor climate, and evaluation of indoor climate on collection preservation. Remarkably, only few studies can be found that specifically address the energy consumption in repositories (Hong et al.; 2012; Kompatscher et al.; 2019).

Energy costs in buildings (expressed in $kW\,h\,m^{-2}$ or $kW\,h\,m^{-3}$) are calculated as the result of energy loads needed to maintain a certain set-point in temperature and relative humidity (RH). The energy consumption models simulate the heat and vapour flows to and from the environment and the surrounding physical elements, taking into account the temperature and RH differences between supply air and room air. The models calculate the energy consumption in different environmental scenarios, independently of the ventilation system used and the efficiency and energy demand of the ventilation system. Therefore, models deliberately exclude the heating, ventilation and air conditioning (HVAC) systems as model inputs. In essence, these models are based on the same principles of the calculator for energy use in museums proposed by Padfield (2010). Padfield's model is available as a online tool. In Figure 7.3 we present Padfield's model schematically. The model is a simple energy balance, which considers gains and losses of thermal energy through walls and ventilation:

1. Heat loss through the solid building materials, e.g. walls, (*conduction loss*), and air exchange (*ventilation loss*).
2. Energy needed to maintain the indoor relative humidity by evaporating or condensing water (*energy of condensation/evaporation*).
3. Heating energy is offset by waste heat from lighting and equipment.

Based on Figure 7.3, we developed a model that approximates Padfield's model. For a ful description of the developed model, including equations, see Appendix C[2].

In more elaborate models, software such as EnergyPlus (Hong et al.; 2012) or Hambase Heat Air and Moisture modelling and simulation (de Wit; 2006) are used to introduce a detailed description of the characteristics of the building and

[2]This part of the model was developed in collaboration with Josep Grau-Bové.

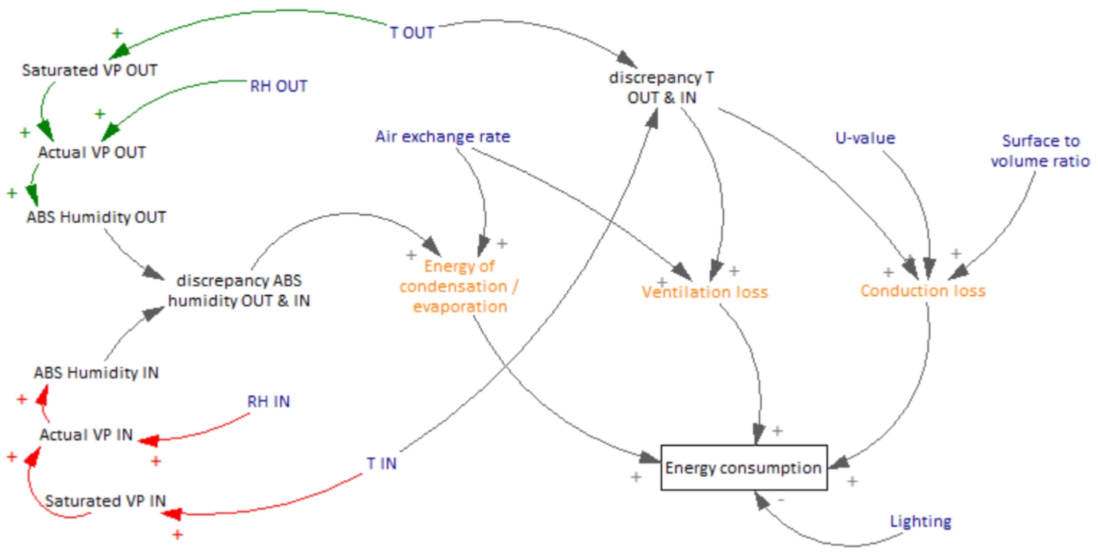


FIGURE 7.3: Scheme of the energy balance model to calculate energy costs in museums proposed by Padfield (2010), where three main mechanisms (in orange) are modelled to calculate the energy consumption (output of the model), depending on the input variables (in blue).

its materials. Models built specifically for environments containing paper collections (Hong et al.; 2012; Radon; 2018; Kompatscher et al.; 2019) may include the modelling of the moisture exchange with the paper mass, accounting for latency in the paper's response.

In order to gain a better understanding of the differences between the models, we analysed the research results of more sophisticated models presented by Kramer et al. (2015) and Hong et al. (2012) against the simplified model proposed by Padfield. The three models agree that, as (de)humidification has a limited contribution to the total energy load, the energy loads are attributable mainly to heating and cooling. More specifically, cooling was found to be responsible of almost 70% of the energy use in Hong's results. Moreover, Hong's long-term predictions are that the cooling load is expected to increase if the mean external temperature increases from 12.6 °C in 2009 to 14.5 °C in 2050, and to 17.7 °C in 2080, from a energy load of 33.4 $\mathrm{kW\,h\,m^{-2}}$ in 2009 to 41.5 $\mathrm{kW\,h\,m^{-2}}$ in 2080 (an increase of 24.3%). However, this steady increase in energy load as result of the rise in external temperature is not supported in Padfield's model under similar inputs. Padfield's model calculates the monthly energy demand. In those months when the external temperature is closer to the target level of the room temperature of 18 °C, the energy load will be low. This is particularly the case of the scenario in 2050 when the mean external temperature is 15.4 °C. Therefore, according to Padfield's model, and opposite to Hong's predictions, the energy demand will decrease if the external temperature increases from 12.5 °C to 15.4 °C, even if other values for air exchange rate (AER) are used (Fig. 7.4).

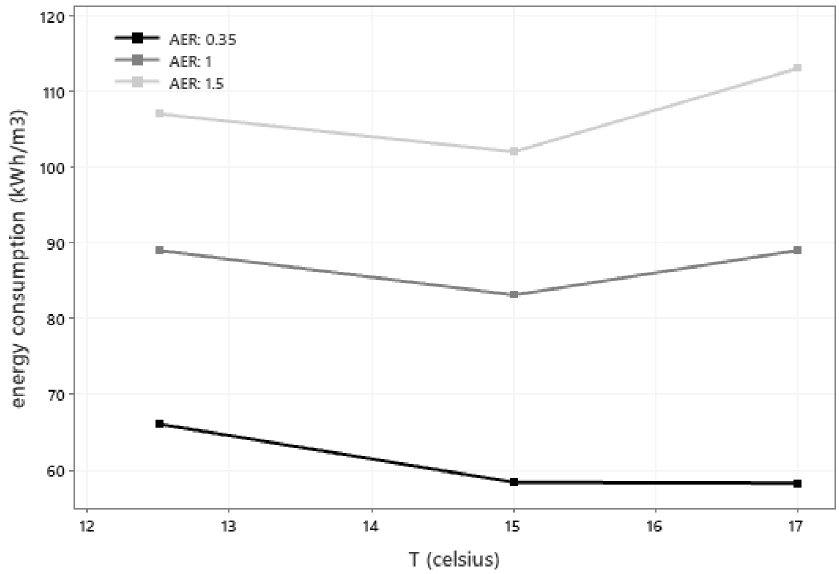


FIGURE 7.4: Calculation of energy loads according to Padfield's model (2010), using different AER inputs in three climate scenarios: mean external temperature of 12.6 °C in 2009 to 15.4 °C in 2050 and 17.7 °C in 2080.

As mentioned, these models do not include the HVAC system, and they calculate the energy loads based on the temperature and RH difference between supply air and room air. However, HVAC systems work very differently from these models. Typical air handling units (AHUs) used to regulate and circulate air as part of the HVAC system in museums have been described by Kramer et al. (2016). These are the elements of Kramer's AHU, each of them contributing to the energy demand:

1. mixing section: outdoor air mixed with re-circulating air
2. dust filter
3. counter-flow cooling coil: with a water supply temperature of 12 °C and return temperature of 16 °C
4. steam humidifier

5. dehumidification coil with bypass: with a water supply temperature of 6 °C and return temperature of 10 °C

6. belt-driven centrifugal fan

7. heating coil: with a water supply temperature of 45 °C and return temperature of 35 °C

8. filtration section: electrostatic, chemically active carbon and end-filter

The AHU uses deep cooling to extract the moisture of the air, which subsequently needs to be reheated in order to meet the indoor temperature in the (exhibition) rooms. According to Kramer's study (2016), heating and dehumidification are responsible for a large proportion of the total energy use. It is expected that significant energy savings can be achieved by relaxing the indoor climate specifications when less deep cooling is required and, consequently, reducing the energy needed for post-heating after dehumidification.

TABLE 7.1: Quality of Envelope (QoE) specifications described by Kramer et al. (2015).

	QoE1	**QoE2**	**QoE3**	**QoE4**
Exterior wall	400 mm brick	400 mm brick	400 mm brick, 100 mm insulation	100 mm brick, cavity, 150 mm insulation, 100 mm brick
Glazing	Single	Double	Double, low-E	Double, low-E
U-value	$5.7\,W\,m^{-2}\,K^{-1}$	$3.2\,W\,m^{-2}\,K^{-1}$	$1.4\,W\,m^{-2}\,K^{-1}$	$1.31\,W\,m^{-2}\,K^{-1}$
Convection factor	0.010	0.030	0.030	0.047
Solar gain factor	0.80	0.70	0.65	0.31
Infiltration rate	$1\,h^{-1}$	$0.4\,h^{-1}$	$0.2\,h^{-1}$	$0.11\,h^{-1}$

A model that includes the energy consumption of a HVAC has not been built yet for the heritage field. For now, Padfield's is the most accessible and well accepted model, as an indication of the least amount of energy needed, assuming perfect efficiency of the system used to achieve the climate. Though the model makes several simplifications, it can still be used to explore the magnitude of energy savings due to changing one of the following inputs: inside climate (temperature and RH), or building characteristics, such as surface to volume ratio, U-value (thermal transmittance of the external surface) and AER.

U-value and AER might not be known by institutions. A useful indication of U-values is given by Kramer et al. (2015) who distinguished four Quality of Envelope scenarios (Table 7.1). AER, according to Padfield (2010), ranges vary from 0.03 ACH for well isolated, windowless buildings, to around 2 for a typical house, and 5 and more for busy public rooms. AER values for repositories have been reported between 0.08 and 0.18 ACH (Twinn; 1997; Di Pietro et al.; 2016).

7.1.4 Key findings for the APM-cost model

In the previous sections we reviewed different approaches of cost models: from broadly oriented to highly comprehensive models, from cost-effectiveness models in collection care to activity-based cost models for digital collections. Cost-effectiveness models are useful when aims and criteria are clear (e.g. increase preservation level, reduce costs). In that case, as stressed by Henderson and Waller (2016), one's goal needs to be defined before decision making starts. However, before defining goals and criteria, it is important to be well informed of the options and their consequences for the whole system. The decision-making process starts with reflection, and, as Michalski correctly observed, '(w)hat-if games are, in fact, the arithmetic version of reflection' (Michalski; 2018, p. 202). In that sense, the aim of the APM model is to promote reflection, presenting (semi-)quantitative outputs, but avoiding any evaluation of these outputs. Specifically for the APM cost model, a comprehensive computation of costs is needed in order to allow the comparison of options. Including the same amount of detail in the calculations is a precondition for comparing actions or activities. Therefore, the activity-based cost model seems a more adequate approach for the APM cost model, as they provide the right framework for a systematic computation of costs.

Another important aspect to be included in the model is the concept of the collection lifecycle. Collections undergo different stages during their lifecycle, that require different management decisions, at different costs. For example, it is good to differentiate between the management of the collections in the first year, when they are acquired or digitally created (scanning), as those initial are expected to be higher than in later, retention-focused, years. This difference is a crucial element when evaluating the costs over the mid- and long-term.

Regarding the energy cost models, most of the models use complex software to model the characteristics of the buildings and their materials, and to calculate the energy loads. The theoretical concepts behind these sophisticated models do not greatly differ from the model proposed by Padfield. The models are energy balance models, where thermal energy is gained and lost through walls and ventilation. As Padfield's model does not require sophisticated software, it can be easily reproduced and included in the APM model. However, as discussed previously, as long as the HVAC system loads are not included in the calculation, the energy balance models provide merely a broad indication of the minimal energy loads needed to maintain a certain set-point in repositories, and, therefore, the value of this information is limited when used for comparing potential preservation actions.

7.2 APM-Cost model description

7.2.1 Purpose and boundaries of the model

The developed cost model can be used independently or in combination with the preservation and access part of the APM model. Using both together allows computation of the total costs of preserving and providing access to the collections, as well as allowing comparison of specific preservation measures (e.g. deacidification versus lowering the temperature and RH in the storage facilities). Whereas, for the first purpose, the model needs to be as exhaustive as possible, for the second propose, a selection of variables is sufficient. In this chapter, we discuss the APM cost model in its completeness (see Appendix D for description). In the next chapter, a overview of the APM model is given, including only the variables of the APM cost model which are required for the APM model (Chapter 8).

The APM model includes only paper collections, specifically bound and unbound records, and the digitisation of these collections. The uniformity of this collection type helps to narrow the model and its complexity. For example, part of the complexity of cost models for digital collections arises when the models include digitally born documents along with the diversity of formats to be digitised. Our model includes the digitisation of archival records in one format, considering only their average size (e.g. one gigabite), and repair, metadata and digital preservation are expected to play only small roles compared to their greater prominence when managing digitally born archives.

It is possible to include economic adjustments in the model, such as inflation/

deflation, depreciation, or cost of return. However, this level of detail is difficult to predict, and probably not relevant when the aim is to compare the effects of present actions now and in the future, as such financial adjustments would apply similarly to all of the measures under consideration. Similarly, as the focus of the model is on the preservation measures rather than the general costs of a institution, the model only accounts for costs related to the activities of interest, and not necessarily costs that apply to all activities, such as infrastructure cost adjustments for the building. Following the methodology presented by Michalski and Karsten (2018, p. 189), we chose to applying no complex adjustments, and one-time costs such as equipment purchases were spread over their expected life times.

7.2.2 Main structure of the model

The main structure of the model follows the approach of the DEN Digital cost model (Gillesse et al.; 2010), but is extended to physical collections. The model distinguishes two main stages in the lifecycle of archival collections: acquisition and retention. Each of these stages comprises several activities. Whereas acquisition activities will take place just once, the retention activities occur repeatedly throughout the lifecycle.

TABLE 7.2: Activities included in the APM-cost model.

Archival collection		**Digital collection**	
Acquisition	Inventory Rehousing Placing	Creation	Selection Creation Metadata Control Ingest
Retention	Storage Conservation Access	Retention	Storage Preservation Access

Table 7.2 shows the activities included in each stage. Formally, digitisation activities belong to the retention stage, for paper collections. However, as a new (digital) collection is formed when collections are digitised, and the activities in both types of collections are notably similar, we will treat them separately in order to compare them.

The costs of each activity are recorded in detail, including personnel costs (salaries) and equipment costs, which include all costs that are not expressed as an hourly rate. The model computes the annual costs for the whole collection as well as the costs of each activity per shelf-metre of the collection.

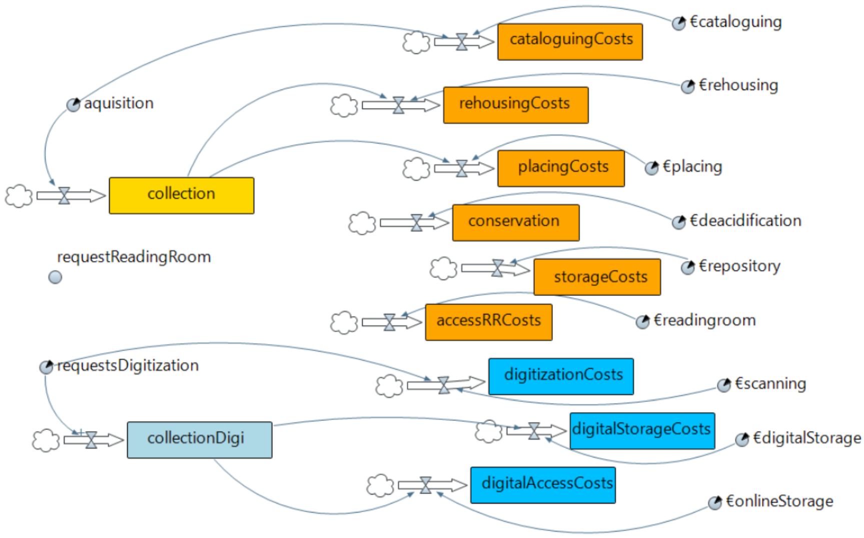


FIGURE 7.5: Simplified scheme of the APM cost model

Figure 7.5 shows how the costs are computed in the model, particularly if changes in collections size or the number of requests are taken into account in the calculations. We observe that storage and reading room pose fixed costs, independent of the amount of archive being preserved. The outputs of these activities will be constant during the retention years of a collection. In contrast, the costs of other activities, such as rehousing, placing and digital storage, does depend on the size of the archive that undergoes the activity, and will change accordingly when computing the costs.

The model was developed and tested in collaboration with the Amsterdam City Archives (SAA) and Utrechtsarchief (UA). A detailed description of the APM-cost model, as Excel spreadsheet, is given in Appendix D.

7.3 Experiments

7.3.1 Digital access vs reading room access

As discussed in Chapter 5, digitisation has been embraced as strategy to enhance access to collections. As the demand for accessing collections digitally is expected to continue growing over the coming decades, knowing at what cost it comes is essential in order to make informed decisions. The digitisation of collections is also affecting the use of the collections in the reading room. Hence, if costs are approached as part of the complex system, then it is not sufficient to calculate the costs of providing digital access in the collections by itself; analysing the cost of digitisation in relation to the cost of preserving and accessing physical collections will provide new insights. This question has been addressed by several studies (Lesk; 1996; Ashley; 1999; Chapman; 2006), most of which are case studies conducted in library collections, that compare the costs of digital production and storage of a book to the traditional costs of keeping collections in repositories for access in the reading room. One of the main conclusions is that the cost of providing digital access not only depends on the quantity (e.g. the number of scans produced and stored), but that it also responds to a number of choices made for the whole process of digital production and storage, regarding quality, format, integrity and risk. In addition, how the cost is calculated and presented (e.g. total cost for the whole collection, cost per volume or cost per use) will provide different information (Kingma; 2000; Shenton; 2003): calculating the total cost will inform us of the required investment, whereas cost per volume or per use can be used for analysing cost effectiveness.

Compared to these studies, the strength of the proposed APM cost model lies in its dynamical cost computation by (1) introducing the difference between acquisition and retention years, and (2) linking the costs to the outputs of the APM access model. Hereafter, several examples are discussed to illustrate these two points.

In the first example, we show how the model can be used to compute the total cost of preserving and providing access of (partly digitised) paper collections for a period of ten years. Continuing the study started in Chapter 5, the SAA collections are used as case study.

We start by computing the cost of the different activities, enumerated in Table 7.2 and Figure 7.5. The aim of this exercise is to visualise the contribution of each activity to the total cost of preserving and providing access to the collections, to put the cost of each activity in context of the whole budget. The following inputs were used as starting point:

- Collection size: 40,000 metres
- Acquisition: 1,000 metres/year

- Scans in storage: 35,000,000
- Digitisation: 20,000 scans/week

Figure 7.6, *left*, shows that providing digital access accounts for 36% of the annual budget for the preservation and access of the collections. Regarding the activities related to the paper collections, the most expensive activities are cataloguing new acquired records and storage in environmentally controlled repositories. In the case of the digitised collections, creation (scanning) and access (namely online storage) constitutes most of the cost. However, three remarks need to be made:

- The cataloguing cost is an estimate of what it will cost to catalogue the records acquired in a year according to a certain level of description, but, in practice, institutions might differently prioritise the collections to be catalogued and the level of description.
- As seen in previous sections, to calculate the energy loads in repositories, we use the simplified model proposed by Padfield, where the storage costs are calculated as the minimum energy load needed to maintain a certain temperature and RH in the repositories. However, if we chose a more realistic model including HVAC simulation, then higher storage costs estimates are expected (see discussion in section 7.3.2.1).
- As has been mentioned, digital costs are not only the result of the amount of available scans, but the result of the sum of decisions. In the case of the SAA, scanning standard archival record at large scale results in a relative low price per scan (€0.20/scan). SAA opts for the storage of lower quality scans (.jpeg) and relative small formats (0.002 GB/scan), resulting in relative less data to be stored. However, the higher costs are those related to the online storage of the scans (€2.85/GB). SAA chooses high-performance online access, which means that scans, besides being storage for preservation, are also online storage for fast access. It is clear that other choices will result in different distributions of the costs.

Therefore, the costs reported in these examples should be interpreted with caution.

A cost model is also useful to visualise the consequences of strategies over the middle-term, as well to identify leverage points. For example, in Figure 7.6, *right*, we observe the expected increase of online storage costs in the next ten years, as a consequence of the annual production of scans and, consequently, the increasing number of scans to be stored. We also note that online storage contributes

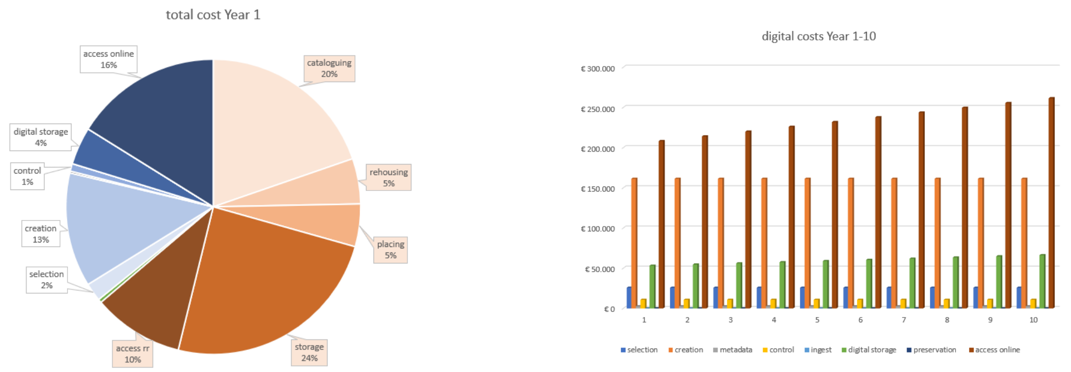


FIGURE 7.6: (*Left*) Annual cost of preserving and providing access of the collections at the SAA. (*Right*) Costs per activity related to digital collections during a period of ten years, with a production rate of 20,000 scans per week and 35,000,000 scans in storage in the first year. Online storage price is €2.85/GB per year.

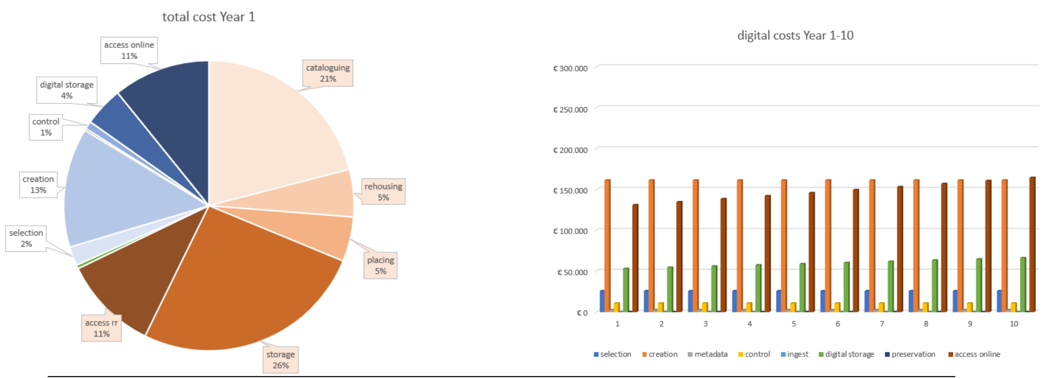


FIGURE 7.7: (*Left*) Annual cost of preserving and providing access of the collections at the SAA; (*Right*) Costs per activity related to digital collections during a period of ten years, with a production rate of 20,000 scans per week and 35,000,000 scans in storage in the first year. Price of online storage is €1.78/GB per year.

the most to the total cost, and, therefore, it is a potential leverage point to reduce costs. For instance, an strategy might be to choose a high-performance online system only for those scans that are frequently accessed; low-demand scans (comparable to collections with low demand in the reading room) could be consigned to a 'cold' storage which is notably less expensive than online storage. The result of choosing lower digital storage costs (e.g. €1.78/GB per year) is presented in Figure 7.7. By changing the strategy, the online storage cost could potentially be reduced to a level similar to the creation cost.

The model can also be used to explore strategies from a cost-efficiency point of view. For example, at what point is the Scan-on-Demand (SoD) request rate expected to surpass the request rate in the reading room, as more and more records become digital available? Knowing when this will happen is important to managers because it indicates that the reading room can probably be replaced by the SoD service.

In order to answer this question, we performed several runs to predict the number of requests in the reading room for a period of 10 years, under four different scenarios. The output of the agent-based modelling (ABM) approach using the SAA dataset seen in Chapter 5 was taken as the starting point of the run (Fig. 5.25). Table 7.3 summarises the four runs, with different inputs for the number of records digitised per year (R_s and R_b) and the SoD and block matches (M_s and M_b).

TABLE 7.3: Four runs to predict the effect of digitisation on the number of requests in the reading room at the SAA.

	SoD requests R_s	Block requests R_b	SoD match M_s	Block match M_b
Run 1	5000	0	0.42	-
Run 2	10000	0	0.42	-
Run 3	5000	5000	0.42	0.33
Run 4	5000	5000	0.42	0.25

As seen in the description of the APM-access model in Chapter 5 (section 5.5), the output of the ABM model is the annual number of requests (R_r), which is not necessarily equivalent to the number of records requested, as archival records can be accessed more than once in a year, resulting in multiple access requests per record. On the contrary, the unit of digitisation requests always refers to records. Therefore, Figure 7.8 shows the predicted decrease in the number of records, which are the access requests (R_r) converted to records, by applying a correction factor of 1.5 (based on the annual average of requests per record in the SAA dataset).

The predictions show that, only if the present digitisation production rate of 5,000 records per year is increased to 10,000, then, within 10 years the turning point will be reached when the annual number of requests in the reading room will be similar to the annual scan-on-demand production, making it an option to replace the reading room service by the scan-on-demand service. The results also show that different strategies (e.g. exclusively SoD and/or block digitisation program) produce similar outputs, as long as an annual production rate of 10,000 records is maintained.

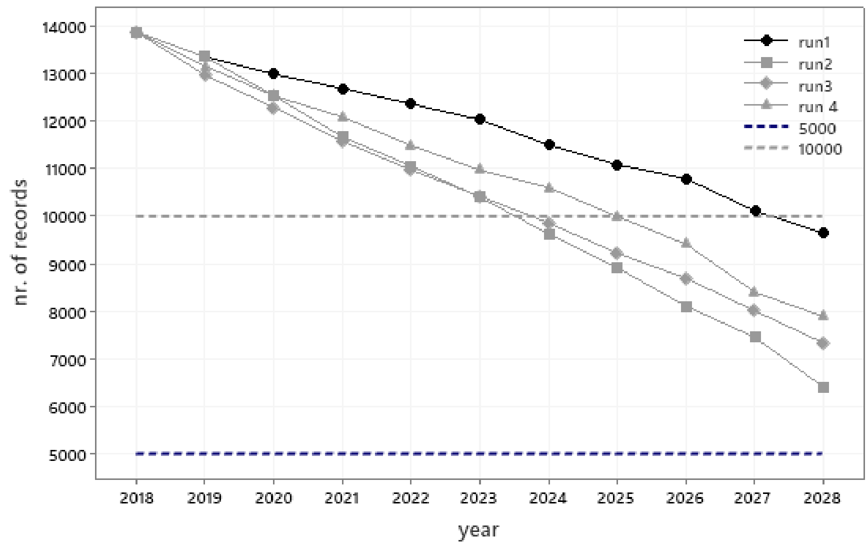


FIGURE 7.8: Prediction of the number of requested records in the reading according to four digitisation strategies as described in Table 7.3. The dashed line represents a digitisation production rate of 5,000 and 10,0000 records per year.

Doubling the scan production would come, of course, at a cost. Looking at the costs involved in the two processes of providing access of the collections, physically and digitally, the costs of maintaining the open reading room will remain similar over the years, whereas the costs for digital storage will rise annually due to the increasing amount of data generated by continuing scanning. Often, the creation of digital content (the action of scanning the archival records) is seen as the most expensive step of digitisation (Poole; 2010; Beagrie et al.; 2010). However, in the mid–long-term, rapid-access, online digital storage might become more expensive than the scanning itself, due to the increasing volume of data being stored. This is already the case at the SAA. Therefore, digitisation is a relatively expensive investment that must be based on a clear vision with regard to service provision and further developments on how collections will be used.

The costs of digitisation vary between locations but, once the costs are known for a specific archival institution, the capabilities of this model could be easily extended to economic forecasts. In the case of the SAA, around half of the investment needed to maintain a digitisation rate of 10,000 records per year could be covered by the cost savings of suspending the reading room service. However, when an institution contemplates the option of shifting to a fully digital access services, two aspects need to be considered. First, not only the quality of the scan, but even more importantly, the delivery time of the scan will affect the readers' satisfaction (see Chapter 5, section 5.3). Secondly, the need to provide physical access to a certain part of the collection will remain, either because standard digitisation of those records is not possible, or because readers will still require physical access in cases where other values, such as the materiality of the objects, matters.

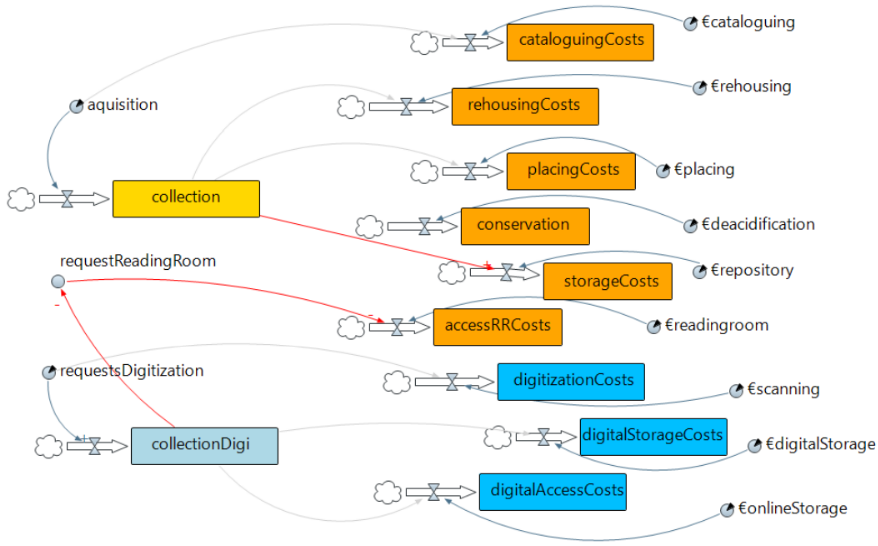


FIGURE 7.9: Scheme of the APM cost model per metre of archive. In red, the new dependencies that arise between variables when costs are calculated for one metre archive rather than for the whole collection (Fig. 7.5).

So far, we have focused on the total costs and the investment needed to provide digital access to the collections. Let us now look at the relative costs by comparing costs of activities for one linear metre of archive. If we now adapt Figure 7.5 (computation of total cost for the whole collection) to computing the cost per archive-metre (Fig. 7.9), we observe that the relationship between parameters changes slightly. For example, as we assume fixed repository storage costs, the annual increase in the collection size results in a decreased price of storage per metre. Similarly, the reading-room access costs per shelf-metre increase as the number the number of reading-room requests decreases. Simultaneously, the number of requests in the reading room decreases when the digitally available portion of the collection increases.

To illustrate the consequences of these relationships, in the following example, we compare the costs of the activities related to physical collections versus those related to digitised collections. We calculate the cost for one metre of archive during the lifecycle of ten years under the following three scenarios:

1. *One metre of archive not accessed at all*: costs include cataloguing and rehousing in the acquisition year, and storage cost during the retention years.

2. *One metre of archive physically accessed once per year*: costs include the above, plus annual costs for providing access in the reading room.

3. *One metre of archive digitally available*: costs include creation in the first year and digital storage and access during the retention years. No costs of activities related to the physical archive are included.

Figure 7.10 reports the cost for each of this type of metre archive during the ten first years of the lifecycle. The first year is the most expensive for physical collections as well as for the digitisation of one metre archive, due to cataloguing and scanning costs. During the retention years, changes in the costs are almost negligible, except for the case of the costs of reading-room access. As the reading room costs are fixed, as the number of records accessed there per year reduces due to increased digital access, the the access cost per metre of archive becomes more expensive.

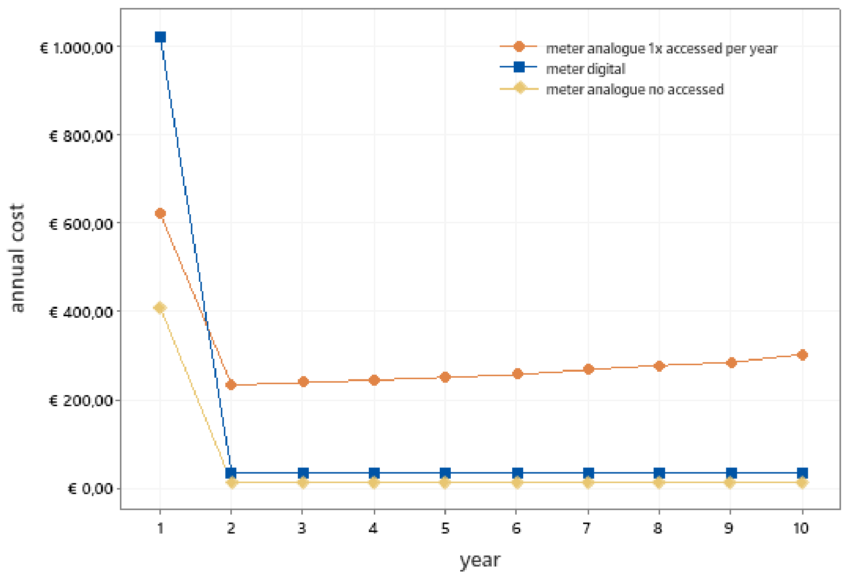


FIGURE 7.10: Annual costs per metre of archive during a period of ten years.

According to Figure 7.10, the costs during the retention years for the physical and digitised records are similar, which means that the costs in repositories are comparable to the costs of digital storage for one linear metre of archive. However, the costs of digitisation in the first year are more than double those of cataloguing and rehousing physical records. In this digital era, 'the disruptive impact of digitised and born-digital records has shaken ways of working as well as ways of interpreting the international standard for archival description ISAD(G)' (Garmendia; 2017, p. 2). For example, new automated cataloguing processes are being developed, in the case of born-digital records, though initiatives are also emerging to explore the feasibility of automated description for digitised records. But, taking into account the investment needed to digitised records for automated description, for now the traditional way of cataloguing by staff seems a more cost-effective strategy than digitising entire collections to facilitate automated description.

From these results, a new question arises, namely at what point is it more cost-efficient to digitise a record than to provide access in the reading room? To answer this question, we analyse the same data but, this time, we calculate the cumulative costs of of proving access, either digital or in the reading room. By doing so, we can calculate how many times a record would have to be requested in the reading room in order to break even with the costs of making it available digitally. Although that calculation will differ between institutions, depending on the chosen scan quality and the costs of consultation in the reading room, in the case of the SAA, we observe that, after five requests in the reading room, the point has been reached at which making and keeping a scan available is more cost-effective (Fig. 7.11).

Looking at these results, can one really conclude that providing digital collection access is a cost-effective strategy compared to providing access in the reading room? To answer this question, we need to take into account the statistics that show how often records are repeatedly requested in the reading room (Chapter 5, Fig. 5.2). In the case of SAA, we saw that, over a period of 20 years, over 50% of the requested records have been requested just once, and 8% had been requested more than five times. Taking these statistics into account, from a cost-efficiency point of view, only a small portion of the requested records in the reading room will be more cost-effectively accessed through digitisation than in the reading room within a period of 20 years.

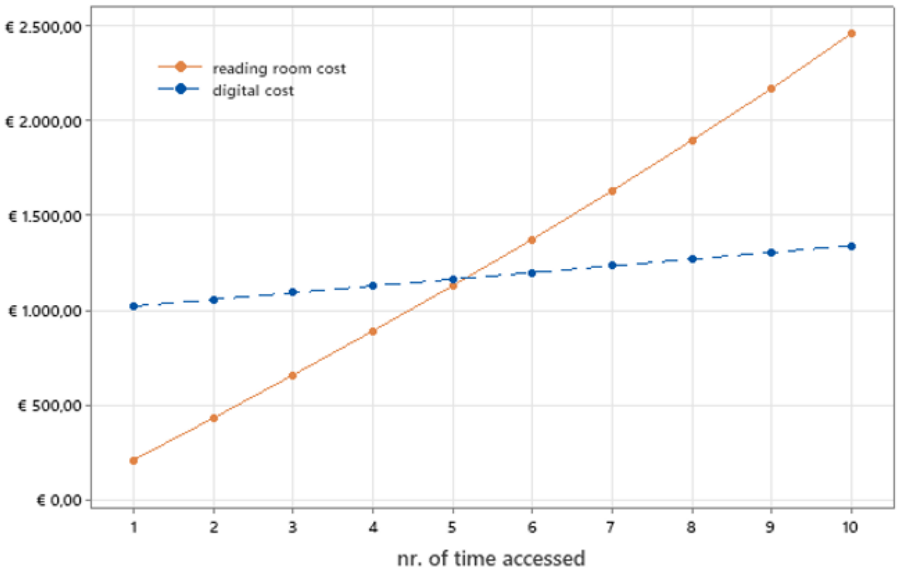


FIGURE 7.11: Cumulative cost of providing digital access versus access in the reading room per metre of archive, according to run 1 in Table 7.3.

However, given that no access data has been available to analyse the digital use of the collections in this research, the results of these analyses are only related to the level of demand in the reading room. Other potential benefits of providing digital access (e.g. increasing the overall access of the collections) are not included here. However, it is clear that digitisation is a relatively expensive investment which needs to be justified by a clear strategy on how the institution envisions the use of the collections.

7.3.2 Deacidification and environmental storage strategies

In Chapter 6 (section 6.7), we compared the effect of deacidification versus various environmental storage conditions on the preservation of paper collections. Note that deacidification is not a separate measure, but a added measure to the environmental control in storage facilities. More precisely, we were in fact comparing between lowering temperature and RH with respect to the traditional values of 18 °C and 50% RH, and maintaining the environmental conditions of 18 °C and 50% RH but supplemented by the deacidification of part of the collection.

In the next sections, we explore in more detail the costs of the two options in order to contextualise the investment required for these two preservation measures.

7.3.2.1 Energy costs

In this section, we explore the use of the model adapted from Padfield (see section 7.1.3) in order to calculate the annual energy use in a repository. The input values are summarised in Table 7.4.

TABLE 7.4: Input values to calculate the energy use in a repository.

	Input	Value
	U-value	$1\,W\,m^{-2}\,K$
	surface-to-volume ratio	1
	AER	$0.2\,h^{-1}$
	T and RH ext	Table 7.5
Experiment 1	T and RH int	18 °C / 50% RH
Experiment 2	T and RH int	16 °C / 45% RH
Experiment 3	T and RH int	5 °C / 40% RH

The energy use for the three strategies regarding environmental conditions in a repository are given in Table 7.5. As expected, when the set-point of the indoor temperature is similar to the outdoor temperature, then the energy loads are minimal, which implies that, according to this model, (de)humidification has little impact on the energy use. The results also indicate that different set-points of indoor temperature result in small change in energy loads, which seems not realistic, in particular the small difference between experiment 2 (16 °C / 45%RH) and experiment 3 (5 °C / 40%RH). As discussed in section 7.3.2.1, in this model a clearer effect on the energy loads can be observed if the U-value or the AER are changed, as seen in Figure 7.4. In that case, the magnitude of the increase in energy use is noticeable. Therefore, Padfield's model is a useful tool to explore (qualitatively) the magnitude of energy savings of these parameters (U-value and AER), but, to explore the difference between different environmental conditions, the model is less appropriate. Nevertheless, as an indication, Table 7.6 reports the energy load and cost for the three preservation strategies. The energy cost is calculated assuming that 1 km archive occupies about $40\,m^3$ (for an archive containing 50 km archive in $16.000\,m^3$) and the energy price is €0.20 $kW^{-1}\,h$.

On paper, and if we take an average outdoors temperature 10 °C as a reference, choosing a slightly lower temperature than the traditional 18 °C seems to be a

TABLE 7.5: Calculation of the energy use in a repository according to the inputs given in Table 7.4 and the exterior conditions (mean temperature and RH exterior between 1970–2018, de Bilt, The Netherlands) (KNMI; n.d.).

Month	T_{out} (°C)	RH_{out} (%)	**Exp. 1 18/50** (kW h m^{-3})	**Exp. 2 16/45** (kW h m^{-3})	**Exp. 3 5/40** (kW h m^{-3})
January	3.0	87	12.0	10.0	1.8
February	3.2	87	12.0	10.0	1.7
March	5.9	87	9.7	7.9	1.1
April	9.0	80	7.2	5.6	3.6
May	13.0	80	4.1	2.6	7.0
June	15.7	80	2.1	0.7	9.3
July	17.8	78	0.6	2.0	11
August	17.4	78	0.9	1.7	11
September	14.4	78	3.0	1.6	8.1
October	10.7	85	5.8	4.4	5.1
November	6.7	85	9.0	7.3	1.7
December	4.1	85	11.0	9.4	1.0
Total			77.5	63.8	62.0

more sustainable measure as the difference outside and inside temperature is reduced. Things get more complicated, however, when RH also needs to be controlled. As discussed in section 7.1.3, the model proposed by Padfield does not include how RH is regulated. Dehumidification can be energy demanding, as HVAC uses deep cooling to extract moisture from the air. From conversations with technicians specialised in HVAC systems, it became clear that lowering the RH will require a lower temperature of the water supply in the dehumidification coil (e.g. from 6 °C to 2 °C). In addition, the need for dehumidification will further increase as, the lower the temperature, the lower the air's moisture-holding capacity. Therefore, it is expected that lowering the RH in repositories will result in a higher energy demand which Padfield's model does not calculate. Furthermore, HVAC systems which are already in use in archival repositories have been generally built to meet the requirements of 18 °C / 50%RH. Therefore, it is questionable whether the HVAC systems currently in use have the capacity to meet

TABLE 7.6: Calculation of the energy use and cost in a repository as given in Table 7.5

Experiment	Energy use (kWh)	Energy cost (€)
Exp. 1 18/50	1,240,000	272,800
Exp. 2 16/45	1,020,800	224,576
Exp. 3 5/50	992,000	218,240

reduce the RH this far. If they cannot, extra investment will be needed to increase their dehumidification capacity.

Unfortunately, for now we are unable to investigate this relationship between temperature and RH further. We can not explore, for instance, whether a lower annual average temperature and RH could be achieved by simply relaxing the seasonal set-points (drier and cooler in the winter, wetter and warmer in the summer), which would result in energy savings (Hong et al.; 2012; Nadal; 2016). The question of whether lowering both temperature *and* RH will result in energy savings compared to the traditional 18 °C / 50%RH when HVAC systems are used cannot be answered for now.

7.3.2.2 Deacidification costs

The deacidification costs are usually calculated per kilogram archival material. Hence, the average weight per box is required to calculate deacidification costs. During the project on wear and tear in archival collections (Duran-Casablancas et al.; 2019), we collected data on the number of sheets per inventory item. Figure 7.12 shows the correlation between the number of sheets and the thickness of the stack, in millimetres. The scatter plot divides the data into two groups: paper prior 1850, which is mostly rag paper, and paper after 1850, which is bleached pulp paper and woodpulp paper. The width of standard archival boxes is 12 cm. Let us take the example of an inventory item with 5 cm thickness. According to the graph, a 5 cm stack of pre-1850 rag paper contains 250 sheets, whereas the same thickness of paper dating after 1850 comprises 400 sheets. Hence, depending on the type of paper, a box may contain 500 to 800 sheets, which makes a total of 4000 to 6400 sheets in one metre of archive with 8 boxes. These results are in agreement with the reported 5991 sheets per linear metre (historical archives) and 7000 sheets in government archives) in other studies (Poole; 2010).

A typical A4 sheet office paper weighs ca. 5 grams. Depending on the number

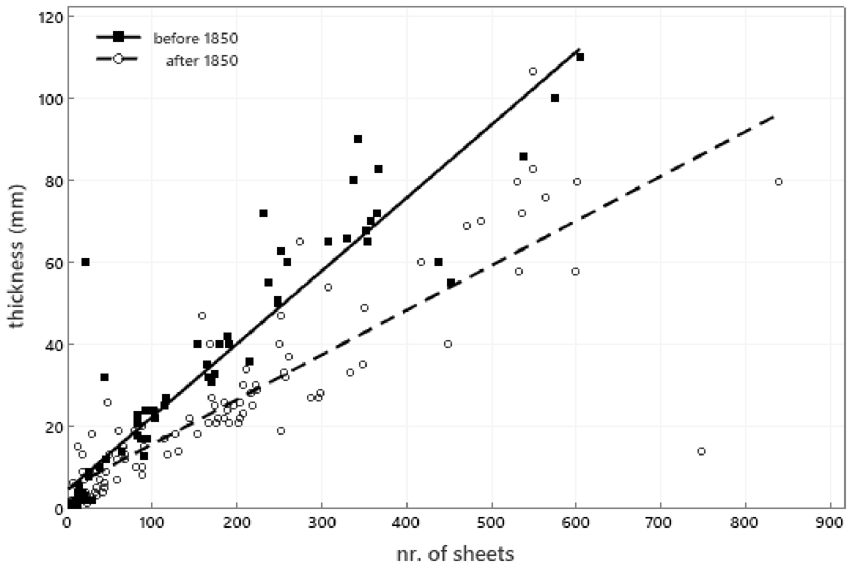


FIGURE 7.12: Correlation between thickness and the number of sheets of a inventory number, grouped by date of the records.

of pages in one box, the weight of one box is between 2.5 and 4 kg. As deacidification occurs mostly on paper dating after 1850, then 4 kg per box seems a good approximation. This is in line with 4.5 kg mentioned in the PaperTreat project (2012) and deacidification projects conducted at NIOD (see Chapter 8), where the reported average weight of one box is 4 kg.

The cost of deacidification will differ depending on the method used. For the following calculations, we took, as an example, the Bookkeeper process, with an estimated cost of €45/kg. If material needs to be preselected, for example due to the presence of photographic materials, then the price can increase up to 20%. Hence, the deacidification of one linear metre of archive using the Bookkeeper methods, without preselection, costs €1440 (= 4 kg/box × 8 boxes × €45/kg).

In Chapter 6, when we presented the dashboard of the APM preservation model (see Fig. 6.7), we briefly introduced how the costs of deacidification have been included in the model. The model provides the possibility to introduce deadicification costs according two inputs: the weight of one metre of archive (kg/m), and the price of deacidification per metre of archive (€/m). Then, the annual and total costs are calculated as output.

Let us now put the cost of deacidification into the context of the measures reviewed in the previous sections. For instance, compared to the digitisation costs, the deacidification of one metre of archive is slightly higher in the first year (€1440), compared to digitisation (€1100), but, during the following retention years, no additional costs are expected in the case of deacidification.

Compared to other preservation related activities, in 2009 the Dutch National Archives (NA) calculated the costs of keeping a linear metre of archive in storage for the PaperTreat project (2012). The following costs were included and calculated for one linear metre of archive:

- Storage in commercial air conditioned repository: €28/year
- Deacidification: €1080
- Rehousing three times in 100 years: €195
- Integrated pest control: €0.064/year (for a collection containing 1,000,000 boxes)

In order to calculate the annual cost per metre, the deacidification costs were divided by 300 years (the assumed prolonged life expectancy), making a annual cost of €3.5/year, and the rehousing by a period of 100 years (€1.95/year). According to the Dutch National Archives (PaperTreat; 2012), the preservation of one linear metre comes to €33,51 per year. In such approach, deacidification accounts for approximately 10% of the total costs of preserving one metre of archive, as the costs are spread in 300 years, instead of concentrated into the single investment. Therefore, the PaperTreat report concluded that '[d]epending on the condition of the material to be deacidified, the lifespan may be prolonged by several hundred years. From an economic point of view this is an extremely profitable investment. The drawback is that one has to invest currently to receive the benefits in the long term' (PaperTreat; 2012, p. 1). However, the conclusions of this study should be interpreted with caution, as they are based on the relative costs per metre of archive. As the Dutch National Archives have high relative repository costs, the deacidification costs are a relative low component of the total. In the case of the SAA and according to the APM cost model, the annual storage costs per metre are €14,49, due to the relatively low costs in repositories (€7,69/metre, energy cost). In that case, deacidification costs (€4.8/metre) count for one third of the total preservation costs. This example illustrates the complexity of the cost analysis, and that the need to develop energy load models for repositories closer to real-world practice, than the highly simplified models that are available so far.

Another important aspect to take into account when evaluating deacidification costs is the characteristics of the collections, such as type and size, as discussed in Chapter 6 (section 6.7). The effect of deacidification versus several environmental storage condition strategies was analysed in preservation terms. Now, we can compute the costs of deacidification for those cases. As an example, we show in Table 7.7 the costs of deacidification for different types and sizes of collections, to obtain similar preservation levels as environmental storage conditions of 16 °C and 40% RH.

TABLE 7.7: Cost of deacidification for different types of collections as described in Table 6.9, to obtain similar preservation levels as environmental storage conditions of 16 °C and 40% RH.

Collection type	**Size** (m)	% **deacidified**	**Annual costs** (€)	**Total costs** (€)
Library col. (L.1)	2,000	45	12,960	1,296,000
Library col. (L.1)	50,000	45	324,000	32,400,000
Acidic archival col. (C.1)	2,000	70	20,160	2,016,000
Acidic archival col. (C.1)	50,000	70	504,000	50,400,000
Modern archival col. (C.2)	2,000	60	17,280	1,728,000
Modern archival col. (C.2)	45,000	60	432,000	43,200,000
Mixed archival col. (C.3)	2,000	45	12,960	1,296,000
Mixed archival col. (C.3)	50,000	45	324,000	32,400,000

Choosing deacidification as a preservation measure means that institutions are taking their responsibility to preserve the collections in the long term *now*, instead of leaving to the uncertainty of whether, in the next hundred years, the temperature and RH in their repositories will be maintained to the advisable lower values, for example when subjected to expected challenges related to climate change. However, as discussed in Chapter 6 (section 6.7) and as shown in the calculations related to the deacidification cost, when the concern is the preservation of the collections as a whole, deacidification might be a highly expensive option. Such investment seems mostly justifiable when dealing with collections with a high percentage of content that, in the next 100 years, will reach the critical DP value of 300.

7.4 Conclusions

Since institutions deal with limited budgets, cost is a crucial aspect that needs to be considered in order to make well-informed decisions. Moreover, sustainability, understood as the responsibility to lower greenhouse gas emissions, has become a clear priority within institutions, which is reflected, for example, in recently drawn-up sustainability plans (see, for example, the NARA Climate Action Plan (2021)). In the case of archives, sustainability is a particularly relevant issue for the management of energy consumption in repositories as well as in digital data storage.

In this chapter, we proposed a model to calculate the costs over the collection lifecycle, making a distinction between the first year after acquisition and the following years of retention. Although the focus of the model is on the measures of deacidification, digitisation and storage facilities, other activities related to preservation and providing access to the collection are also included in the more comprehensive version of the APM-cost model in order to put the different measures into context.

Needless to say, an important aspect when comparing the cost of activities is that the same level of detail in the costs must be included. However, which level is required might differ depending on the aim of the analysis. For instance, in the presented examples that compare digital versus physical access to the collections, we included all the costs that are directly related to these services, including personnel costs, but excluding the costs that are common or not directly related to these services (e.g. cataloguing or rehousing). In the case of deacidification, we only included the costs related to deacidification itself, excluding personnel costs related to pre-selection or preparing collections for transport, in order to make it comparable to preservation measures related to lowering the temperature and RH in repositories, where costs accounts only for the energy demand in the repositories, but not the maintenance of the air conditioning systems or cleaning the repository facilities.

Most of the costs in the model can be easily calculated as the sum of inputs. To calculate the deacidification or digitisation cost of a metre of archive, the inputs are widely available and easy for end users to understand. In contrast, the calculation of the energy demand in repositories is not straightforward. In the APM-cost model, we used a well-accepted model that calculates the minimal energy needed to maintain the environmental conditions at a certain value, depending on the external environmental conditions. This is a simplified model that does not take into account how the temperature and RH are regulated. In order to compare the effects of different environmental strategies in repositories on the energy demand, a more sophisticated model is required that can predict energy loads closer to the actual energy consumption, which depends on the characteristics of the building and its HVAC system. Such a model is yet not available, which means that the APM-cost model requires further development before it can report valuable information on the energy costs in repositories.

Cost results from a multiplicity of decisions which are specific to each institution. Whereas in the APM-preservation model we were able to make some general recommendations based on the results of the experiments, in this chapter, the examples presented only illustrate how the model can be used, as the results may vary greatly between individual cases. We showed that a more interesting approach is to use the model in combination with the other parts of the APM model. Regardless of whether the model can accurately calculate the costs of particular activities, the same calculations can be used to examine single questions that go further than what for the model was initially built for. For example, we forecast the critical moment when the weekly number of requests in the reading room becomes similar to the weekly scan-on-demand production rate, which means that, by then, replacing the reading room service by the scan-on-demand will become viable, but also at other costs.

We also investigated how many times a record would have to be requested in the reading room in order to break even with the costs of making it available digitally, and how this information is related to the analysis of how often records have actually been subject to repeated requests. The analysis made clear that, although (mass) digitisation enables new uses of the collections (such as automated transcriptions), it is a relatively expensive investment and can only be justified if it serves the institutions' clear vision of how the collections are to be used.

References

4C Collaboration to clarify the costs of curation (n.d.).
URL: *https://www.4cproject.eu/*

Ashley, K. (1999). Digital archive costs. Facts and fallacies., *Proceedings of the DLM-Forum on Electronic Records (DLM '99)*.

Ayris, P., Wheatley, P., Aitken, B., Hole, B., McCann, P., Peach, C. and Lin, L. (2010). The LIFE3 Project: Bringing digital preservation to LIFE, *Technical report*.
URL: *http://www.life.ac.uk/3/docs/life3_report.pdf*

Beagrie, N. and Daphne, C. (2017). D4.9 Cost-benefit advocacy toolkit deliverable report, *Technical report*, CESSDA.

Beagrie, N., Lavoie, B. and Woollard, M. (2010). Keeping research data safe 2, final report, *Technical report*, JISC.
URL: *http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf*

Bell, N., Cassar, M. and Strlič, M. (2018). Evidence for informed preservation planning and advocacy. A synoptic view, *Studies in Conservation* **63**(sup1): 8–14.

Bogvad, K., Hougaard, E., Thirifays, A. and Bo, N. (2014). D3.1 Evaluation of cost models and needs & gaps analysis., *Technical report*.
URL: *cproject.eu/summary-of-cost-models/*

Bülow, A. E. and Brokerhof, A. W. (2011). The QALY in collection care: A cost effectiveness approach to collection management, *in* J. Bridgland (ed.), *ICOM-CC 16th Triennial Conference Lisbon 19-23 September 2011: Preprints*, Critério, Almada.

Cassar, M. (1998). *Cost / benefits appraisals for collection care*, Bigwood and Staple Ltd., Bridgwater, Somerset.

CCSDS (2012). Reference model for an open archival information system (OAIS).
URL: *http://public.ccsds.org/publications/archive/650x0m2.pdf*

Chapman, S. (2006). Counting the costs of digital preservation. Is repository storage affordable?, *Journal of Digital Information* **4**(2): 1–15.

de Wit, M. (2006). *Hambase. Heat, air and moisture model for building and systems evaluation*, Techinische Universiteit Eindhoven, Eindhoven.

Di Pietro, G., Ligterink, F., Porck, H. and de Bruin, G. (2016). Chemical air filtration in archives and libraries reconsidered, *Studies in Conservation* **61**(5): 245–254.

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part II: an epidemiological study, *Heritage Science* **7**(11).

Garmendia, J. (2017). Digital cataloguing practices at The National Archives, *Technical report*, The National Archives.
URL: *https://www.nationalarchives.gov.uk/documents/digital-cataloguing-practices-march-2017.pdf*

Gillesse, R., Jochems, R. and Maris, M. (2010). Handleiding rekenmodel digitaliseringskosten, *Technical report*, Digitaal Erfgoed Nederland.

Henderson, J. and Waller, R. (2016). Effective preservation decision strategies, *Studies in Conservation* **61**(6): 308–323.

Hong, S. H., Strlič, M., Ridley, I., Ntanos, K., Bell, N. and Cassar, M. (2012). Climate change mitigation strategies for mechanically controlled repositories. The case of The National Archives, Kew, *Atmospheric Environment* **49**: 163–170.

Kingma, B. (2000). The costs of print, fiche, and digital access. The Early Canadiana Online Project., *D-Lib Magazine* **6**(2).

KNMI (n.d.). Maand- en jaarwaarden van de temperatuur, neerslag en luchtdruk.
URL: *https://www.knmi.nl/nederland-nu/klimatologie/maandgegevens*

Kompatscher, K., Kramer, R. P., Ankersmit, B. and Schellen, H. L. (2019). Intermittent conditioning of library archives. Microclimate analysis and energy impact, *Building and Environment* **147**(October 2018): 50–66.

Kramer, R. P., Schellen, H. L. and van Schijndel, A. W. (2016). Impact of ASHRAE's museum climate classes on energy consumption and indoor climate fluctuations. Full-scale measurements in museum Hermitage Amsterdam, *Energy and Buildings* **130**.

Kramer, R., Schellen, H. and Van Schijndel, J. (2015). Energy impact of ASHRAE's museum climate classes. A simulation study on four museums with different quality of envelopes, *Energy Procedia* **78**: 1317–1322.

Lesk, M. (1996). Substituting images for books. The economics for libraries., *Symposium on document analysis and information retrieval, April 15-17, 1996, Las Vegas, Nevad*.

LIFE (2010). Life Cycle Information for E-Literature.
URL: *http://www.life.ac.uk/*

Lucchi, E. (2016). Multidisciplinary risk-based analysis for supporting the decision making process on conservation, energy efficiency, and human comfort in museum buildings, *Journal of Cultural Heritage* **22**: 1079–1089.

Michalski, S. (2018). Sharing conservation decisions. Tools, tactics, and ideas, *in* A. Heritage and J. Copithorne (eds), *Sharing conservation decisions. Current issues and future strategies*, ICCROM, Rome, pp. 183–204.

Michalski, S. and Karsten, I. (2018). The cost-effectiveness of preventive conservation actions, *Studies in Conservation* **63**(sup1): 187–194.

Nadal, J. (2016). Comprehensive preservation environments: Site-wide resource management and conservation outcomes for recap, *Up, Up and Away: High Density Storage for Library Materials 10-11 August, 2016 Library of Congress, Washington, D.C.*, ReCAP: The Research Collections and Preservation Consortium, Princeton, NJ.

NARA (2021). Climate action plan, *Technical report*, National Archives and Records Administration.
URL: *https://www.archives.gov/files/about/plans-reports /sustainability/climate-action-plan.pdf*

NCDD (2017). Onderzoek naar de kosten digitale duurzaamheid, *Technical report*, Nationale Coalitie Digitale Duurzaamheid (NCDD).

Padfield, T. (2010). Conservation physics. Calculator for energy use in museums.
URL: *http://www.conservationphysics.org/atmcalc/energyusecalc.php*

PaperTreat (2012). Evaluation of mass de-acidification processes. Deliverable 11. Specification of costs. PaperTreat project, *Technical report*.

Poole, N. (2010). The cost of digitising Europe's cultural heritage, *Technical report*, Collection Trust.
URL: *https://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf*

Radon, J. (2018). Energy use optimization in the building of National Library, *IOP Conference Series: Materials Science and Engineering* **415**(1).

Shenton, H. (2003). Life cycle collection management, *LIBER Quarterly* **13**(3-4): 254–272.

Twinn, C. (1997). Passive control of relative humidity to ±5%, *UK, Chartered Institution of Building Services Engineers, 1997, proceedings of CIBSE National Conference held Alexandra Palace, London, UK, 5-7 October 1997*, Vol. 2, pp. 77–89.

Varlamoff, M. (2004). The future, *in* J. Feather (ed.), *Managing preservation for libraries and archives*, Ashgate, Aldershot, chapter 8, pp. 159–174.

Chapter 8

Model application

In the previous chapters we developed the three sub-models of the Archival Preservation Management (APM) model separately. Before testing the model on actual collections, we present first an overview of the model, including inputs and outputs.

In the second part of this chapter, we investigate the application of the Archives Preservation Management (APM) model to actual collections. The underlying research question in this chapter is, to what extent does the model provide new information in relation to the information already collected for the input values?

8.1 Description of the APM model

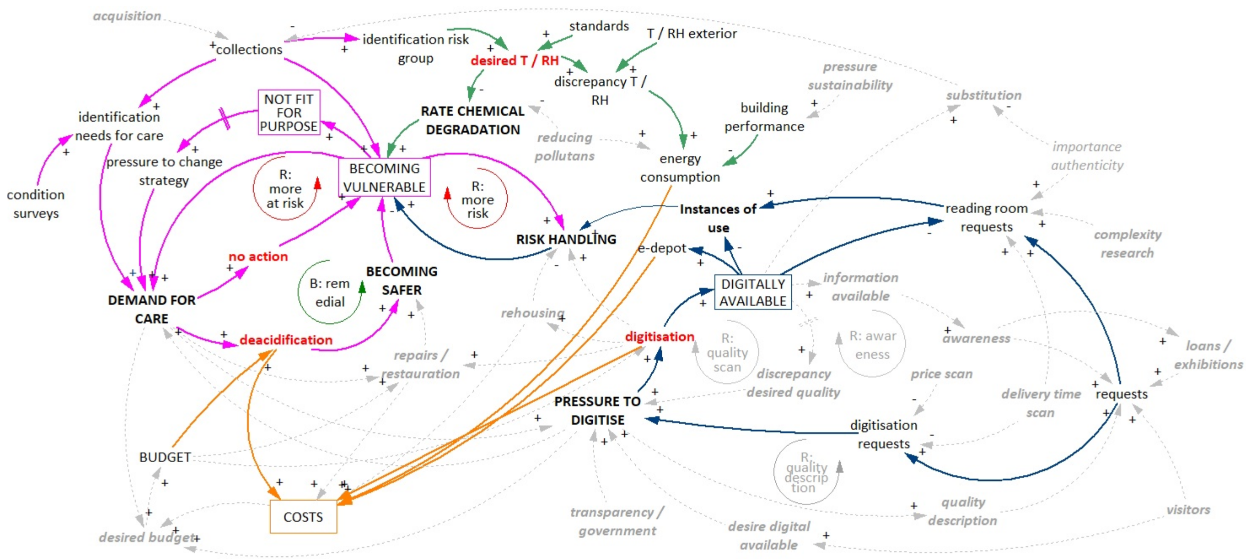


FIGURE 8.1: Causal loop diagram showing the variables included in the APM model, and the variables omitted (in light grey).

Fig. 8.1 shows the variables that have been included in the APM model in relation to the variables identified in the causal loop diagram. At this point, a note must be made: although the APM model is meant to be used by professionals from archival and library institutions, the three sub-models have not yet been integrated into one user-friendly tool. Figure 8.2 shows an impression of how the dashboard of such a tool would look at initialisation of the run.

The model was built using the software AnyLogic 8.6.

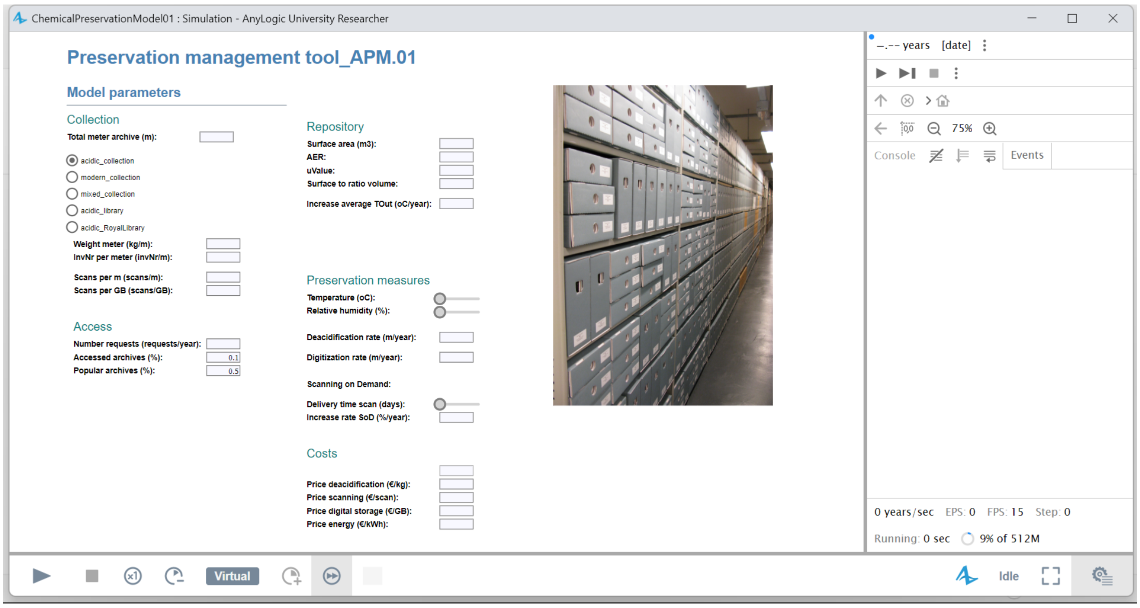


FIGURE 8.2: Mock-up of the dashboard including several inputs of the APM model as example.

8.1.1 Inputs

We distinguish between two types of inputs:

1. Inputs chosen before the initialisation of the simulation run.

   - pH and DP values of the agents (section 6.5).
   - outdoor temperature and relative humidity (Table 7.5)
   - pH distribution after deacidification (section 6.6.2.2)
   - records requested before beginning run (default value= 4%)
   - increase of newly requested records per year during the run (accessed once (p=0.003); accessed more than once (archive size*0.002))

2. Inputs chosen at initialisation of a simulation run by the end user. Table 8.1 summarises the inputs grouped according to the sub-model.

It is worth noting that the main unit used in the model is linear metres of archive. Those parameters whose unit is records (read inventory number) are converted into metres. An agent in this model is equivalent to both a certain number of metres and a certain number of records. Table 8.2 shows the variables that define a single agent.

As seen in Chapter 7, in the APM model, only those parameters related to cost calculate energy cost in repositories, deacidification price and digitisation price (creation and digital storage). For a description of the comprehensive APM cost model, see Appendix D.

8.1.2 Outputs

At every step of the run, the outputs are updated. The outputs are reported as annual values as well as the cumulative value at the end of the simulation run. See Table 8.3 for the list of outputs.

TABLE 8.1: List of inputs of the APM model.

Parameter	Default	Unit	Reference	Remarks
Preservation				
archive size	–	metre		
no. records/m	25	records/m		average number records per metre archive
storage *T*	18	°C	Table 6.7.2	environmental condition in repository
storage RH	50	%	Table 6.7.2	environmental condition in repository
deacidif. rate	0	m/year	Section 6.7.2	metres archive treated per year
max. pH deacidif.	6		Section 6.7.1	pH value of the paper
min. DP deacidif.	300	DP	Section 6.7.1	DP value of the paper
Access				
requests reading rooms	–	request/year		at initialisation
accessed archives	15	%		records requested more than once in the reading room at initialisation
popular archives	30	%		records requested more than once of accessed archives at initialisation
SoD	5000	m/year	Table 7.3	
block digitisation	5000	m/year	Table 7.3	
match SoD	42	%	Table 7.3	records requested *and* digitised
match block	33	%	Table 7.3	records requested *and* digitised
match once	30	%		records requested once *and* digitised
Cost				
weight box	4	kg	Section 7.3.2.2	
price deacidif.	45	€/kg	Section 7.3.2.2	
scans/m	5000	scans/m		
price scanning	0.20	€/scan	Section 7.3.1	
scan size	0.002	scans/GB	Section 7.3.1	
price digital storage	2.85	€/GB	Section 7.3.1	
repository size	–	m^3	Sect. 7.3.2.1	1 km archive = ca.40 m^3
AER	0.2	h^{-1}	Table 7.4	
U-value	1	$W\,m^{-2}\,K$	Table 7.4	
surface:ratio volume	1		Table 7.4	
price energy	0.20	€$kW^{-1}\,h$	Section 7.3.2.1	

TABLE 8.2: List of inputs/outputs of the agents in the APM model.

Variable	Remarks
Input	
pH	
DP	at initialisation of the run
deacidified	yes/no
requested	yes/no
popular	yes/no
digitised	yes/no
Output	
DP actual	DP value updated during the run
number of requests	
number of tears	

TABLE 8.3: List of outputs of the APM model.

Output	Remarks
Preservation	
chemical condition – critical	% collection (DP < 300)
chemical condition – fair	% collection (301 > DP < 800)
chemical condition – good	% collection (DP > 801)
deacidified	% collection has been deacidified
not deacidified	% collection has not been deacidified
years deacidification	no. years when treatment took place
Access	
wear&tear condition – critical	% collection (more than 75 tears)
wear&tear condition – poor	% collection (between 50 and 74 tears)
wear&tear condition – fair	% collection (between 25 and 49)
wear&tear condition – good	% collection (less than 24 tears)
digitally available	% collection has been digitised
not digital available	% collection has not been digitised
Cost	
deacidification	€
scanning	€
digital storage	€
energy demand repositories	kW^{-1} h
energy cost repositories	€

8.2 Case studies

In the first case study, we test the APM model on a small sized collection of just 2000 metres, the NIOD Institute for War, Holocaust and Genocide Studies. The second case study is the archival collection of the Utrechtsarchief (UA), the archive of the province and city of Utrecht in the Netherlands. The UA case study has been selected to explore whether the model can be applied to just one part of a larger collection, in this case 13 km of archive containing records dating between 1850 and 1960, out of the more than 30 km kept in the repositories.

The focus of the case studies lies on collections mostly containing paper dating between 1850 and 1960. Paper from this period is expected to have a lower durability during ageing due to the introduction of acidic rosin sizing and wood-based fibres around 1850 (Gurnagul et al.; 1993; Strlič and Kolar; 2005). Survey results (Sobucki; 2003; Konsa; 2007) have consistently reported low pH values in those paper collections until the 1980s, when neutral sizing was introduced into large-scale paper manufacturing (Baird et al.; 1997). Selecting this type of collection allows us to model collections where noticeable changes are expected within a time horizon of the 500 years.

8.2.1 NIOD Institute for War, Holocaust and Genocide Studies

The State Institute for War Documentation (Rijksinstituut voor Oorlogsdocumentatie, RIOD) was founded on 8 May 1945, three days after the Netherlands was liberated. Created as a war documentation centre, its main task was to collect and inventory documents about the Netherlands and the Dutch East Indies during the Second World War as well as conducting research in this subject. This has resulted in a collection mainly containing documents dating around the Second World War. However, modern material can also be found in the collection as from the 1990s when the field of research was broadened in terms of time and geography, and when the institute's name changed into Netherlands Institute for War Documentation (Nederlands Instituut voor Oorlogsdocumentatie, NIOD). In 2010, NIOD merged with the Centre for Holocaust and Genocide Studies (CHGS) to become the NIOD Institute for War, Holocaust and Genocide Studies.

FIGURE 8.3: NIOD building located in the city centre of Amsterdam, The Netherlands (credit: NIOD).

Since 1996, NIOD has been located in a monumental 19th century building on the Herengracht in Amsterdam (Fig. 8.3). Physical collections can be requested and accessed in the reading room facilities, and, since 2006, parts of the collection have been digitised through several programs. The NIOD collection, with 2,500 linear metres of archive, is stored in two repositories located in the basement of its building (Fig. 8.4).

FIGURE 8.4: Repository facilities at the NIOD (credit: NIOD).

8.2.1.1 Data collection

Chemical properties of the collection
The NIOD collection was sampled for information on the chemical properties of the collection (pH and DP). To obtain a representative sample of the whole collection, a systematic interval sampling method was designed to spread the sample throughout the collection, on the understanding that the collection—as with many heritage collections—follows a certain order (for example, sequencing by acquisition date or record date). For the sampling population, photos and other objects than archival records and pamphlets were excluded. The population was homogeneous and no stratified sampling methods were applied.

When conducting a sample survey in archives, the most common sampling units are metres, boxes or inventory numbers. In the case of the NIOD, a sampling frame (a list containing all inventory numbers) was not available. Regarding the use of number of metres as sampling unit, this is a good option when the shelves are completely filled with boxes, in which case the number of metres containing archival records can be counted within a relatively small time span. However, this was not the case of the NIOD repositories, where shelves are not always completely full. Since counting and marking the metres containing one metre of archive would be highly time-consuming, we chose another sampling unit, namely boxes. This is a good alternative to metres when the collection is mostly housed in the same type and size of box. In the case of the NIOD, eight standard, vertical archival boxes are stored in one linear metre. It has to be said that the standard 12 cm-width boxes were not the only type present. Boxes slightly thinner have also been used, which means that, on occasion, more than 8 boxes are found on a shelf. However, the proportion of discrepant boxes was small, so this difference was not corrected for in the sampling design.

A size of ca. 200 samples was chosen. Assuming a large population size, and a confidence level of 95%, then the expected maximum margin of error for categorical variables is 6.89%, or 5.79% margin of error with a confidence level of 90%. These are well accepted levels of precision in surveys conducted in heritage collections (ICN; 2000).

The following steps were taken for sampling:

1. Counting of boxes: the number of boxes were noted per each bay (element of a rack). A total of 17,442 boxes was counted.

2. Selection of sample boxes: every k^{th} box was selected after a random start, to obtain a sample containing 200 boxes, to ensure spread across the entire collection.

3. Selection of sheets: a record within the selected box, and a sheet within the record, were randomly selected.

DP model for groundwood paper

Near-infrared (NIR) spectrometry was used to obtain, non-invasively, information on the chemical and mechanical properties of single sheets of paper. The application of this technique for paper analysis has been described by Trafela et al. (2007), Cséfalvayová et al (2011), Mahgoub et al. (2016), and Brouwn et al. (2017).

196 sheets were measured using the SurveNIR instrument (Lichtblau, Germany). Using chemometric analysis, the near-infrared reflectance spectra are correlated with analytical data obtained on a reference collection using traditional analytical methods (Strlič et al.; 2007; Lichtblau et al.; 2008). The SurveNIR is a portable instrument that can be used on site. The measurements are taken within seconds, and as many as 100 measurements can be taken in one day if the selection has been prepared in advance. (In the NIOD survey, selection and measurement each took two days of work.) SurveNIR is, therefore, a useful tool for collecting information on large samples (Duran-Casablancas et al.; 2017, 2019; Brown et al.; 2020). The SurveNIR instrumental errors have been evaluated by Brown et al. (2020, p. 4): pH measurements have an error of 0.71 for rag paper, and 0.63 for bleached pulp and groundwood paper; the error for the DP is 400 and for TS around 8 N for rag and bleached pulp paper.

SurveNIR provided data on the following properties: paper type, pH, DP, tensile strength (N), tensile strength folded (N), protein (%), and rosin and lignin content (mg/g). SurveNIR distinguishes the following three paper types (Lichtblau and Dobrusskin; 2009, p. 22):

1. Rag paper, made from rag or cotton fibres.
2. Bleached pulp paper, made from (bleached) cellulose from which most of the lignin has been removed; paper containing less than 75 mg/g lignin is classified within this category.
3. Groundwood paper, made from mechanically sanded wood pulp. Papers with a lignin content exceeding 75 mg/g are classified as groundwood papers.

DP in the SurveNIR project was determined viscometrically. However, as lignin is not soluble in cupriethylenediamine, the solvent used for viscometry, the DP value of groundwood paper is missing from the SurveNIR samples database, and consequently the SurveNIR instrument does not provide data on the DP of this type of paper.

As discussed in Chapter 6, DP is an essential input in the dose—response function to predict the chemical degradation of the papers through their lifetime. As groundwood paper is regarded as not chemically stable, if this type of paper is excluded from the model due to missing data, then an important part of the collection with a higher risk for chemical and mechanical degradation is also excluded. Hence, being able to predict the missing DP values of groundwood paper would be a valuable contribution in order to obtain more completed and relevant results from the simulations.

Recently, the sample dataset of the SurveNIR project has been analysed resulting in a multivariate regression model that investigates the relationship between DP (dependent variable) and other variables collected on the paper samples of the SurveNIR project (independent variables) (Dong; 2020). In this model, the SurveNIR sample dataset is pre-processed by normalisation by min–max scaling that maps all values onto the range [0,1] to contribute equally in the regression model. As the model only includes complete entries, groundwood paper samples were excluded. Following this line of research, Orr has adapted this model for use with groundwood paper (Orr; 2021). Orr's model has been trained using only the output variables given by the SurveNIR instrument (pH, tensile strength (TS), protein, and rosin and lignin content). The approach taken to predict the DP of groundwood paper is composed of two models:

- **Model #1**
  1. Cut the SurveNIR training data, SurveNIR sample dataset, to pH < 7 and MW < 700000, in order to include the range of values common for groundwood paper.
  2. Predict DP for groundwood samples in the training set.
- **Model #2**
  1. Train a model for DP from pH, lignin, rosin, protein, TS, including the predicted DP values of Model #1.
  2. Apply Model #2 to the input dataset collected during the survey to predict DP.

The Model #1 is based on the correlation between DP and molecular weight (MW), as found in the SurveNIR dataset by Strlič et al. (2020), revealing that the slope of the linear correlation is equal to 0.77 times the molecular weight of a derivatised glucose monomer (519 g mol^{-1}), which is in agreement with other research results (Łojewski et al.; 2010; Kes and Christensen; 2013). As MW is known for groundwood paper, MW and other parameters included in SurveNIR dataset are used to predict the DP of the groundwood paper samples in the proposed models. However, it is known that, whereas small amounts of lignin do not interfere with MW determination using the procedure used in the SurveNIR project, the effect of large amounts of lignin is unknown (Strlič et al.; 2020). Taking into account this unknown, and that the model to predict DP has not yet been validated with other data sets, the predicted DP of groundwood paper needs to be interpreted with caution. In addition, the cumulative error of the model is 402. Taking into account that we are mostly interested from a preservation point of view on papers with a DP between 300 and 800 (paper at risk of becoming unfit for use in the next 100 years under the storage at 18 °C and 50%RH), such error (50%—130% error in DP) contributes substantially to the predictions, and similarly for uncertainties related to the pH, discussed in Chapter 6 (section 6.6.2.1). Despite the limitations and the results being for now indicative, the proposed methodology could be a starting point for solving the prediction of DP values for groundwood paper.

Environmental conditions

The climate conditions in the NIOD repositories are logged twice per hour. HVAC systems keep the temperature and RH stable within the values prescribed in the Dutch Archive Regulation (Dutch Admin.; 2014). In 1995 the Archives Act was enacted in the Netherlands (Dutch Admin.; 1995). The Act covers all public archival documents, and also those private archives stored in public repositories (Ketelaar et al.; 2016). The Archives Regulation from 2014 provides further detail under the Archives Act 1995 . The Archive Regulation includes standards for materials, media, quality of information and archival store rooms. Regarding the environmental conditions in the repositories, the Archive Regulation prescribes a temperature 18(2) °C and RH 50(5) %, and ventilation rate of at least 0.1 ACH and circulation rate of at least 1.5 ACH.

The dimensions of the repositories are 12 m × 13 m × 2.46 m (length × width × height), equivalent to ca. 380 m^3 per repository.

Collection access in the reading room
The access information of records that have been requested in the reading room is automatically recorded by the collection management application MAIS-Flexis. Access history data since 2006 are available.

Digitisation projects
NIOD started digitising the collections in 2016. So far, a scan-on-demand (SoD) service has not been introduced, and all digitisation projects are initiated by NIOD. Most of those projects entailed the digitisation of whole collections, and only in a few occasions has part of a collection has been digitised. It is important to note that, since archival records in the NIOD collection might contain sensitive information, digitisation does not necessarily mean that the scans are also made available on the website. For privacy reasons, it may be decided to provide online access only to certain (parts of) collections.

8.2.1.2 Data analysis

Preservation
The survey was conducted to obtain the input data required for the preservation part of the Archival Preservation Management (APM) model. Although not essential for the APM model, we also conducted several data analyses to deepen our understanding of general characteristics of the collection we are dealing with. The means and prevalences of the collected variables were calculated with 95% confidence intervals (CI). As the sampled population was relatively large ($n = 196$), the 95% CIs of the prevalence are within the range of ±6%.
Here follows a summary of the main statistics:

- The collection consists mostly of records dating between 1940 and 1949 (71(6) %). However, part of the collection dates after 1970 (12(4) %) due to recent acquisitions.
- According to SurverNIR measurements, the two types of paper present in the collection are bleached pulp paper (76(6) %), and groundwood paper (24(6) %).
- Regarding the chemical characteristics of the collections, the DP values show a bi-modal frequency distribution, with a first peak related to groundwood paper with a mode around 1100, and a second peak for good-quality modern paper with a higher mode around 2100 (Fig. 8.5). The mean DP value is 1873, and just 2% of the collection has a DP value below 800, which might be explained by the low percentage of the collection (26(6) %) with a pH below 6. The good condition of the most of the collection is further supported by just 15(5) % of the collection having a TS value below 40.

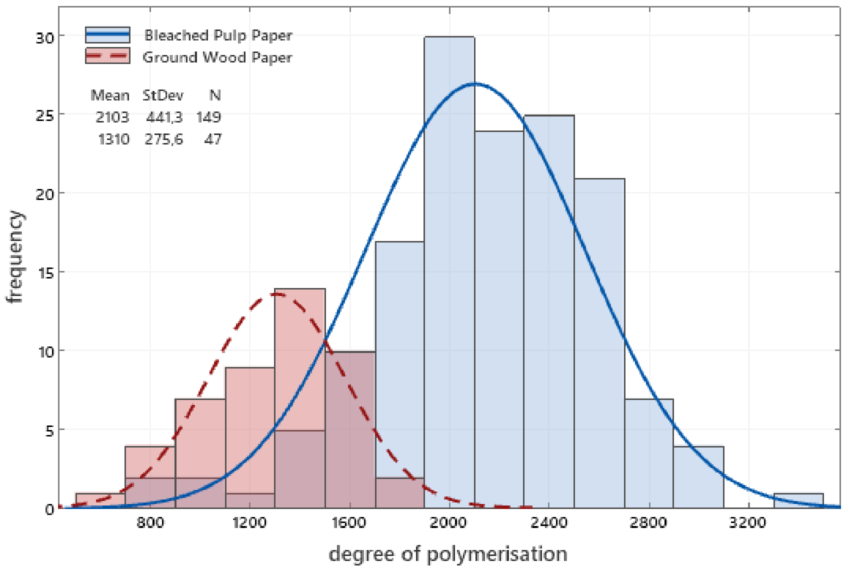


FIGURE 8.5: Histogram of the DP values for paper surveyed at the NIOD.

Since most of the collection dates from around the Second World War, a larger part of the collection was expected to be found in fair or poor condition. However, during the sample survey, it was noticed that most of the selected papers were official records, made of good-quality raw materials. Figure 8.6 (top and bottom) supports that most of the collection is in good condition, and only a small group of papers, mostly groundwood papers, have a pH value below 6, and a TS below 40 N. In previous studies, a TS value below 40 N was found to significantly increase the risk of failure (Duran-Casablancas et al.; 2019).

The principal compounds analysis (PCA) was conducted to explore the relationship between the variables. The PCA seems to indicate that low pH values are correlated to the rosin content in the sizing, and less to the lignin content (Fig. 8.7). However, low pH is not directly associated to the rosin itself, but to the aluminium content from the alum used to precipitate rosin acids onto fibres during sizing (Małachowska et al.; 2020). This association between pH and aluminium was found also by Strlič et al.'s mining of a database of historic European paper properties (Strlič et al.; 2020). But, in the same study, a strong correlation was found between pH and lignin as well as rosin, explained by the degradation of rosin and lignin over time perhaps contributing to the acidity. The NIOD dataset shows a negative correlation between pH and rosin, and not to the lignin content that is negatively correlated to TS. At the same time, as expected, pH is the variable most closely correlated to DP. As the measured pH values are not lower than 5, almost no samples with DP values below 800 were found.

Based on this analysis, we can conclude that groundwood paper collections dating around the Second War World can also contain (good) quality bleached pulp paper. It is worth noting that, in the NIOD archives, groundwood paper seems to be found less frequently after 1960 (Fig. 8.8). These results are in agreement with surveys conducted in other collections (Sobucki; 2003).

Although the results of this survey need to be interpreted cautiously due to the instrument uncertainty and the model uncertainty used to predict the missing values of DP, the results offer a good indication of the general characteristics of the collection and the groups at risk due their chemical properties. The survey results show that most of the collection is in good condition. Confirming these results, during the survey it was also observed that the collections showed almost no signs of mechanical degradation. Taking into account the date of the collection, these results were slightly unexpected. However, if the causes of failure, understood as mechanical degradation, are examined for the NIOD collection, then factors prior to use (design and manufacture) and factors during use (maintenance and usage) can be found which may explain the stable condition of the collection under study (Duran et al.; 2019):

1. Design: the design factor is less relevant in this case compared to collections with bindings, since the NIOD collections mostly comprises unbound

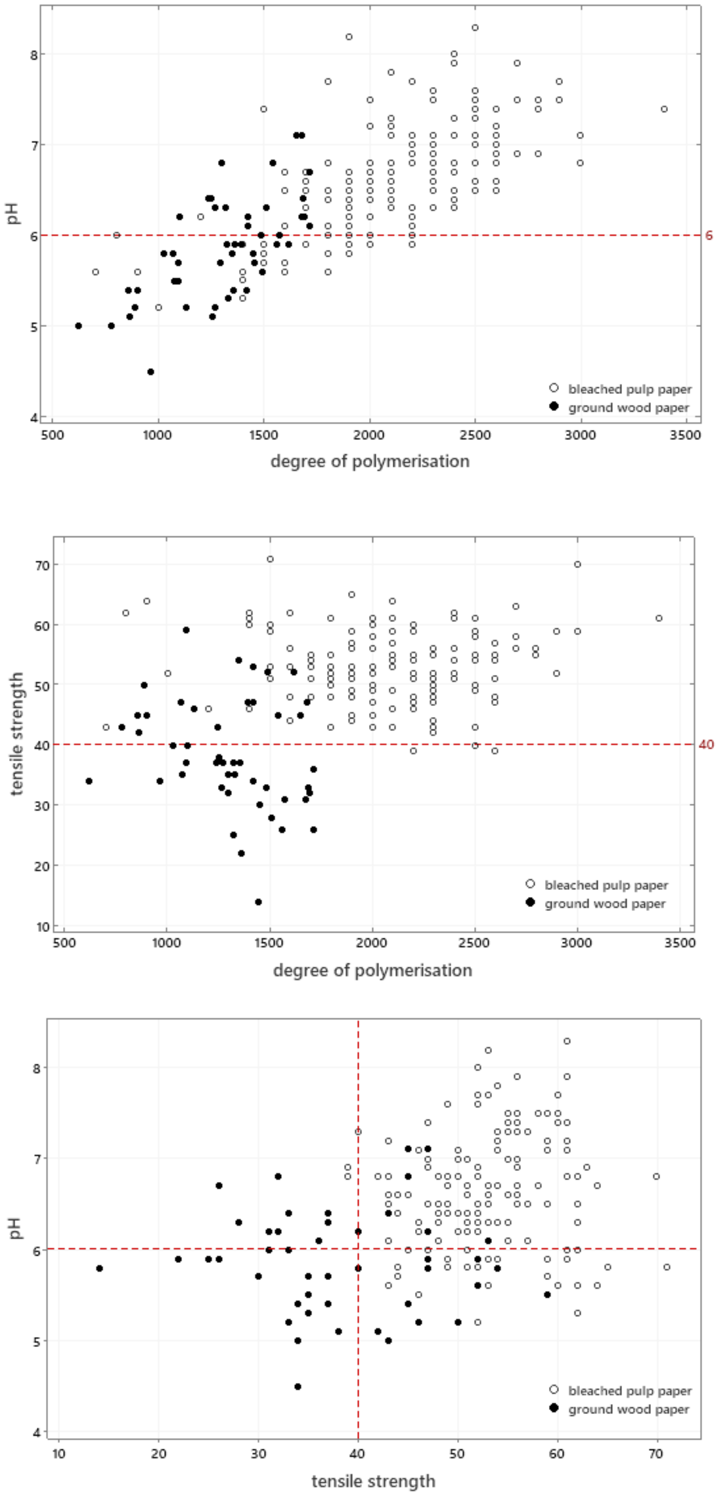


FIGURE 8.6: Scatterplot showing the DP and pH (*top*) and tensile strength (*middle*), and pH and tensile strength (*bottom*) values for the surveyed paper at the NIOD. The dashed horizontal line separates objects with pH < 6 and TS < 40.

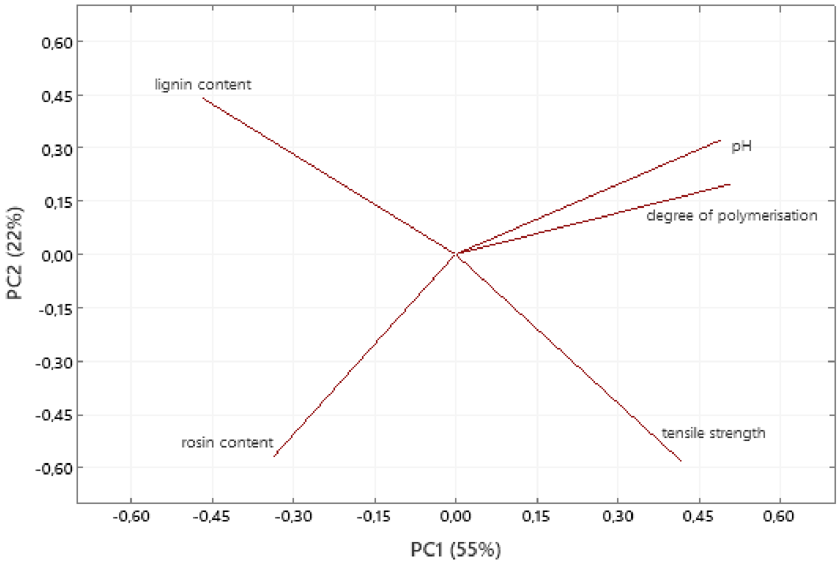


FIGURE 8.7: Loadings of the PCA, showing correlations between pH, DP, tensile strength, rosin and lignin content. The % variance explained by the principal components is in parentheses; n=196

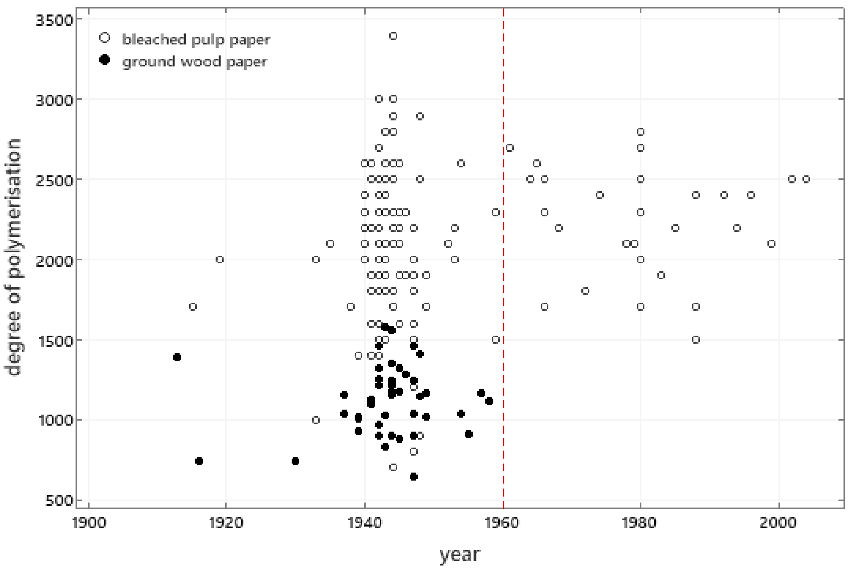


FIGURE 8.8: Scatterplot showing the DP values of the surveyed papers at the NIOD, dating from 1910 to 2010, grouped by paper types.

sheets.

2. Manufacturing: most of the collection is formed by official documents which are mostly made of bleached pulp. Groundwood papers were frequently found in pamphlets or books, and these constituted only a small part of the collection.

3. Maintenance: the maintenance of the collection is excellent. The documents have been rehoused in archival quality boxes and are kept in repositories with annual average temperature of 18.27°C and RH 50.3%, in a very stable climate (Fig. 8.9).

4. Usage: in the next section we will see that the frequency distribution of instances of use in the reading room follows the same Pareto distribution seen in other collections, which means that just a small part of the collection has been heavily used. In addition, one important aspect is that inventory items comprise thin stacks of paper; thick stacks of papers are an exception in this collection. In previous studies, it has been observed that the thickness of the stack plays an important role in the accumulation of tears (Duran-Casablancas et al.; 2019).

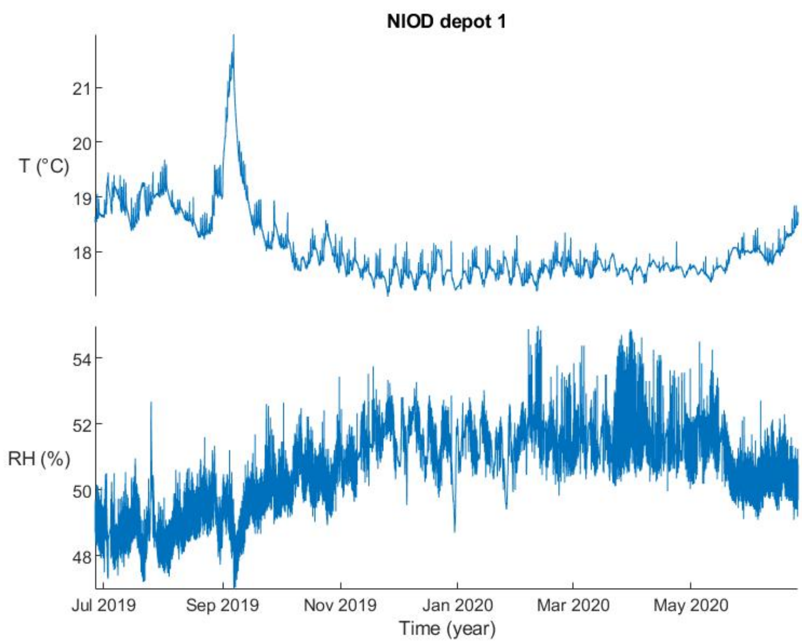


FIGURE 8.9: Storage conditions (temperature and RH) in the NIOD repository, from July 2019 to June 2020.

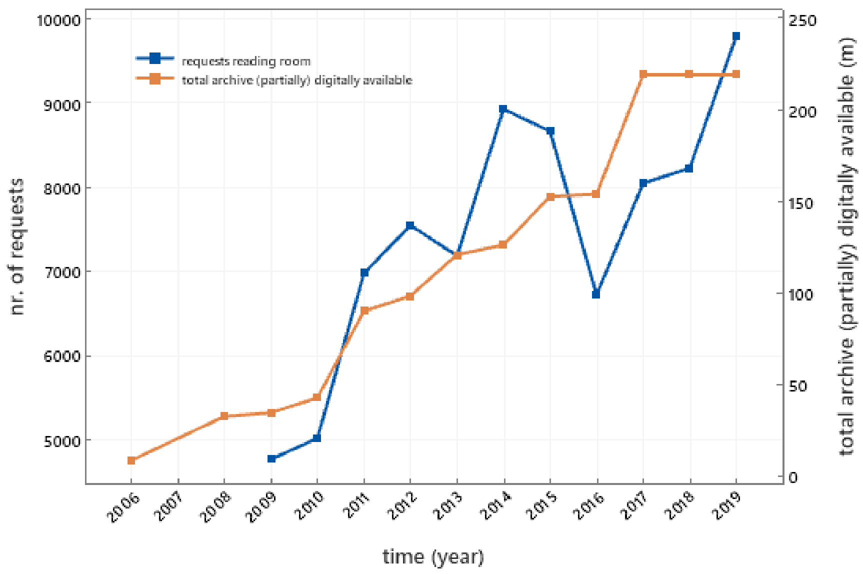


FIGURE 8.10: Number of requests in the reading room and cumulative number of metres of archive (partially) digital available at the NIOD.

Access

NIOD has digitised a total of 258 linear metres in 12 years, which is equivalent to ca. 10% of the collection. Compared to other archives, this percentage shows that NIOD has actively been digitising the collections. However, differently from other institutions, the usage data of the NIOD shows a upward trend on the number of requests in the reading room during the last ten years (Fig. 8.10).

The use of the collections in the reading room is comparable to that seen in other archives: 60% of the collections have been accessed only once during the studied period, and only 1.5% of the records have been requested, on average, more than once per year (Fig. 8.11). Of the total 89,216 requests for access, 48,016 requests belong to just 15 collections. Notably, not one of the top 10 collections have been made digital available. Therefore, it is not surprising that digitisation has not affected the use of the physical collections so far.

In addition, of the 258 m digitised, just 128 m are available online, whereas 91 m are partially available, and 39 m have not been made digitally available. The restrictions on digital access are related to copyrights, and, even more importantly, to privacy. Depending on the type of record, the NIOD staff will decide whether permission can be granted, according to the Archive Law and the Law for Protection of Personal Data.

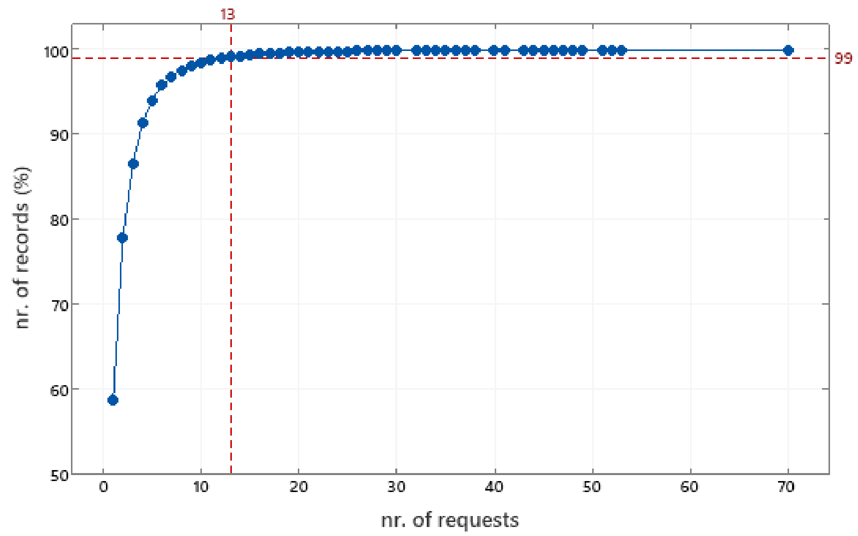


FIGURE 8.11: Cumulative distribution (percentage) of the numbers of records that have been requested in the reading room between 1 and 70 times during a period of 10 years at the NIOD.

Another important aspect to take into account is that, due to the nature of the NIOD collections, visitors attach a higher value to the access to the original than in many other archives. Therefore, the access policy of NIOD, differently from other archives, is that records can still be accessed in the reading room even after having been made digital available (see Fig. 8.12).

All these aspects may explain why the digitisation of the collections has had no effect on reducing the use of the collections in the reading room. On the contrary, a clear increase in the use of the collections has been observed in the last ten years. Although a thorough analysis of the possible causes has not been conducted, it is likely that the digitisation of the collections has been one of the drivers behind the increased collections use (see Fig. 8.12 *right*), a trend that has also been observed in other studies (VanSnick and Ntanos; 2018).

8.2.1.3 Simulation

In this case study, only the preservation part of the Archives Preservation Management (APM) model was simulated. The access part of the APM model is not applicable, since this case study does not follow the main hypothesis of the model, namely that there will be a reduction in reading-room requests due to digitisation.

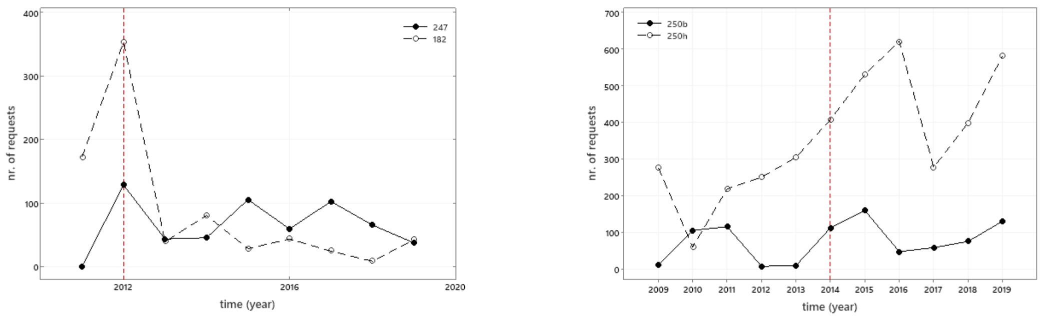


FIGURE 8.12: Number of requests of four NIOD collections; two collections fully digitally available (*left*), and two collections partially digitally available (*right*). The dashed vertical line indicates when the collections were digitised.

Table 8.4 summarises the computational experiments and their results. Different strategies regarding deacidification as well as a slight decrease in the temperature and/or RH were tested in comparison to the baseline (the actual conditions in the repository were 18 °C and 50% RH). As input data, the DP and pH values collected during the survey were used. According to the survey, in Year 1 of the simulation, 2% of the collection is in fair condition (300 $<$ DP $<$ 800) and 98% in good condition (DP $>$ 800).

TABLE 8.4: Comparison of strategies for the NIOD collections after a time horizon of 500 years.

Scenario	*T*	**RH**	**Deacidification**			**Output**
			rate	**duration**	**annual cost**	**poor–fair**
	(°C)	(%)	(m/year)	(year)	(€)	**condition** (%)
1 Baseline	18	50	0	0	0	**2–23**
2	16	50	0	0	0	**1–15**
3	16	45	0	0	0	**0–12**
4	18	50	25	10	36,000	**1–18**
5	16	50	25	10	36,000	**0–11**

As discussed in the previous sections, the NIOD case study represents a collection containing mostly bleached pulp paper with relative high DP and pH values.

The simulations show that, if the actual strategy is maintained, then, after a time horizon of 500 years, 2% and 23% of the collection will be in poor and fair condition respectively. Since the percentage in poor condition is very small, changing the strategy will mostly affect the part of the collection in fair condition. Just a slight decrease in the temperature will induce a noticeably beneficial reduction of the percentage of the collection reaching a DP value below 800. Similar results can be also obtained by deacidification. In this case, to obtain similar results as by lowering the temperature, 25 m would need to be deacidified annually over 10 years. However, for a relatively small institution as NIOD, deacidification can be an expensive investment. To illustrate this point: in recent years, NIOD has been conducting small-scale deacidification projects. These projects deacidified an annual average of 10 m.

In view of these results, NIOD's actual strategy of investing in good storage conditions and conducting small deacidification projects seems to be well justified. A leverage point that could be suggested is a slightly lowering of the temperature, provided that the HVAC systems have the capacity to maintain the RH below the 50% value.

8.2.2 Utrechtsarchief

UA constitutes the memory of the city and the province Utrecht in the Netherlands, with 30 km of archival records, books, photographs and films. The collection, dating from the 12th century to the present day, is stored in environment-controlled repository facilities. The collections can be accessed in the reading room. The repository and reading room buildings date from the 1980's (Fig. 8.13).

In 2006, UA started to digitise their collections with the scan-on-demand (SoD) service, among other digitisation programs. An analysis of the access and digitisation data has been presented in Chapter 5 (section 5.6.2); here, we analyse the data used for the preservation part of the model.

8.2.2.1 Data collection

Chemical characterisation of the collection
A sample survey was conducted to collect information on the chemical characteristics of the collection. The selected sample papers were measured using near-infrared (NIR) spectrometry. Here follows a description of the sampling method,

FIGURE 8.13: UA buildings (credit: Het Utrechtsarchief).

and the processing of the data obtained with SurveNIR, the tool used for the measurements.

The cataloguing system was used to create a list of archive collections with archival records dating from 1850 to 1960. The list included 534 archive item numbers, corresponding to a total of 12,864 linear metres. Besides the location in the repository of each archival collection, the list also included the number of metres per collection. With this information, we constructed a list with the cumulative sum of shelf-metres. Then, according to systematic interval sampling, every k^{th} metre was selected to obtain approximately 150 samples from the population.

In the repository, one inventory number within the selected shelf-metre range, and then one sheet within the inventory number item, were randomly selected. In inventory numbers with more than 100 sheets, a sheet was selected from the first 50 sheets of the stack. Measurements were taken on the upper right corner of each selected sheet using the SurveNIR tool. Measurements were taken for 152 paper samples.

The mean and prevalence of the collected variables of the survey were calculated with 95% confidence intervals (CI). With a sampling population of $n = 152$, the 95% CIs of the prevalence fell within $\pm 8\%$.

As seen in the previous case study, the measurement were taken using the SurveNIR tool, and, therefore, the missing data regarding DP values for groundwood paper samples were estimated using the regression model based on the SurverNIR dataset.

Environmental conditions
The UA uses an HVAC installation to regulate the temperature and RH and to filter the air in the repositories to meet the Archives Regulation prescription. The environmental conditions in each repository are logged hourly.

Costs
For this case study, the intention was to make full use of the APM model, including costs in the different computational experiments, since cost is seen as an intrinsic part of the model. Regrettably, it was not possible to obtain the price of the digital storage, therefore limiting the extent of the analysis. For examples of how the cost part of the model can be applied, we refer to the cases presented in Chapter 7.

8.2.2.2 Data analysis

Preservation
The survey data were analysed to gain a better understanding of the characteristics of the collection and the (chemical) condition of the archival records. Here follows a summary of the main statistics:

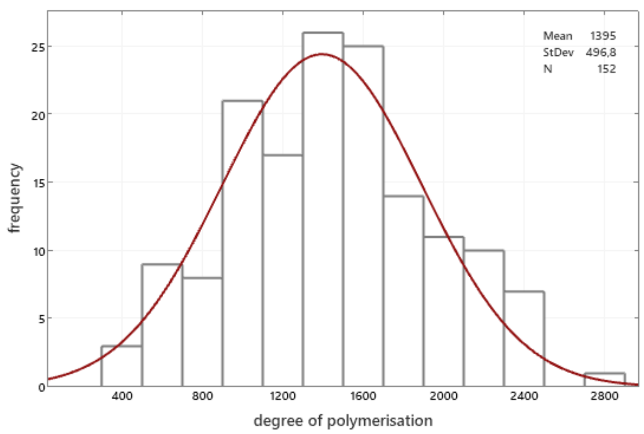


FIGURE 8.14: Histogram of DP values for the surveyed paper at the UA.

- Of the sampling population, $(76 \pm 7)\%$ of records date from 1850 to 1960, corresponding to 8,876–10,419 m of the 12,864 shelf metres of storage space. $(16 \pm 6)\%$ and $(4 \pm 4)\%$ of the sampling population pre-date 1850 and post-date 1960, respectively. As expected, how archival collections are formed does not necessary follow the division in paper quality from 1850–1960; inevitably, records were selected which did not satisfy the dating criteria, even when only archives of this period were selected according to the cataloguing data.

- According to SurveNIR measurements, the type of paper most present in the collection is bleached pulp paper $(56 \pm 8)\%$), followed by rag paper $(38 \pm 7)\%$) and groundwood paper $(24 \pm 6)\%$).

- Regarding the chemical characteristics of the collections, $(28 \pm 7)\%$ of the collection have a pH below $(6.5 \pm 3)\%$

Figure 8.14 shows that the frequency distribution of the DP is normal with a mean of 1395. As seen in Chapter 5 (section 6.5), in modern collections we expect to find a slightly right-skewed distribution. But in this case, including paper dating before 1850 and after 1960 results in the normal distribution seen. However, as expected, papers with a low pH and DP are mostly found between 1850 and 1950 (Fig. 8.15, *left*). These results are similar to those from a comparable survey conducted at the SAA with more than 500 measurements (Duran-Casablancas et al.; 2017, 2019) (Fig. 8.15, *right*). Although the UA sample is smaller than the survey conducted at the SAA, similar patterns can be observed. For example, the results in both archives show that bleached pulp paper took over from rag paper after 1900.

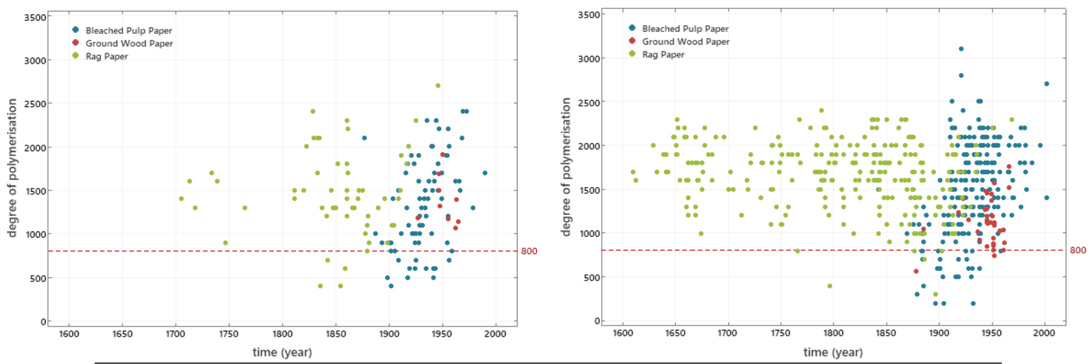


FIGURE 8.15: Scatterplot showing the DP values of the surveyed papers dating from 1600 to 2000, grouped by paper type. *Left*, values of the surveyed paper at the UA; *right*, at the SAA. The dashed horizontal line separates objects with DP < 800.

However, a difference between SAA and UA dataset emerges when the pH values are plotted against DP (Fig. 8.16). The pH in the UA dataset is, in general, higher, and whereas in the SAA all points with a DP below 800 also have a pH below 6, this is not the case in the UA. As seen in NIOD case study above, all NIOD DP measurements below 800 also have a pH lower than 6, similar to the

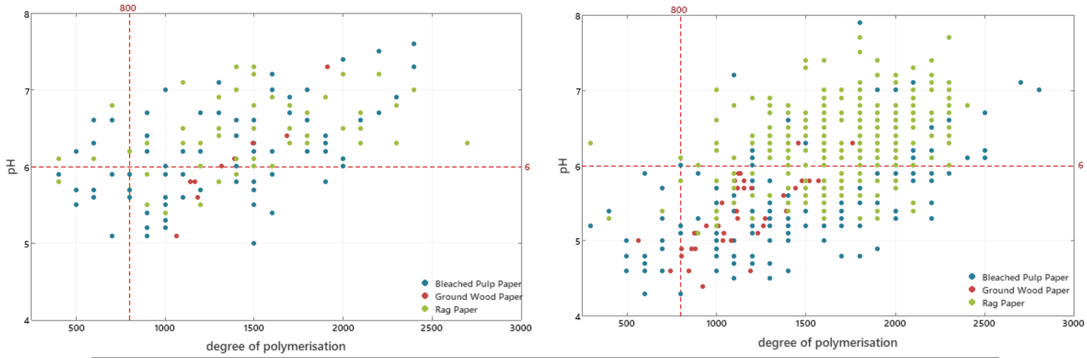


FIGURE 8.16: Scatterplot showing DP and pH for paper from 1600–2000, grouped by paper type. *Left*, values of the surveyed paper at the UA; *right*, at the SAA. The dashed horizontal line separates objects with DP < 800 and pH < 6.

SAA observations (Fig. 8.6). The correlation between pH and DP seen in the SAA and NIOD dataset is more in line with the theoretical expectation that a lower pH is expected to result in a lower DP (Zou et al.; 1996; Strlič and Kolar; 2005). This correlation is missing in the lower part of DP values in the UA. The question arises whether the observed difference could be the result of systematic error (e.g. instrumental error of the SurveNIR tool). We further explored this question by analysing differences between the two datasets by comparing all variables measured by the SurveNIR instrument.

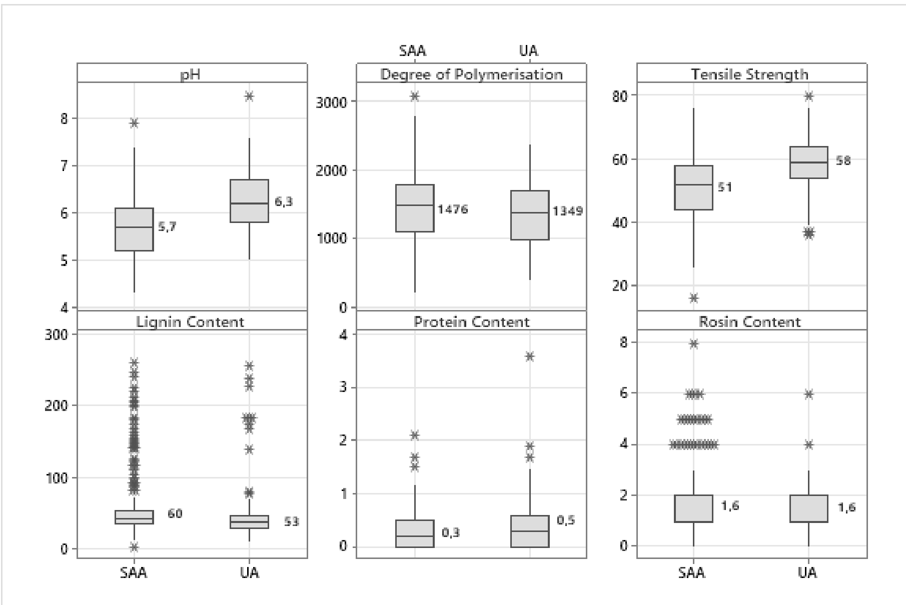


FIGURE 8.17: Comparison of papers at the SAA and UA, including bleached pulp and groundwood paper after 1850.

Figure 8.17 reveals that the average lignin content is slighter in the SAA than in the UA collection. The average rosin content is the same in the two collections, however the protein content, an indicator of gelatin sizing, is a slightly more in the UA papers. As discussed in the NIOD case study (section 8.7), it has been suggested that the degradation of lignin and rosin might contribute to paper acidity. If we take into account the small difference in the rosin and lignin content in the two collections, this difference might be one of the factors behind the higher pH found in the UA papers. Similarly, we have also seen that lignin content relates to tensile strength in modern paper collections, which can also been observed in the UA collections, resulting in a higher average TS compared to the SAA values.

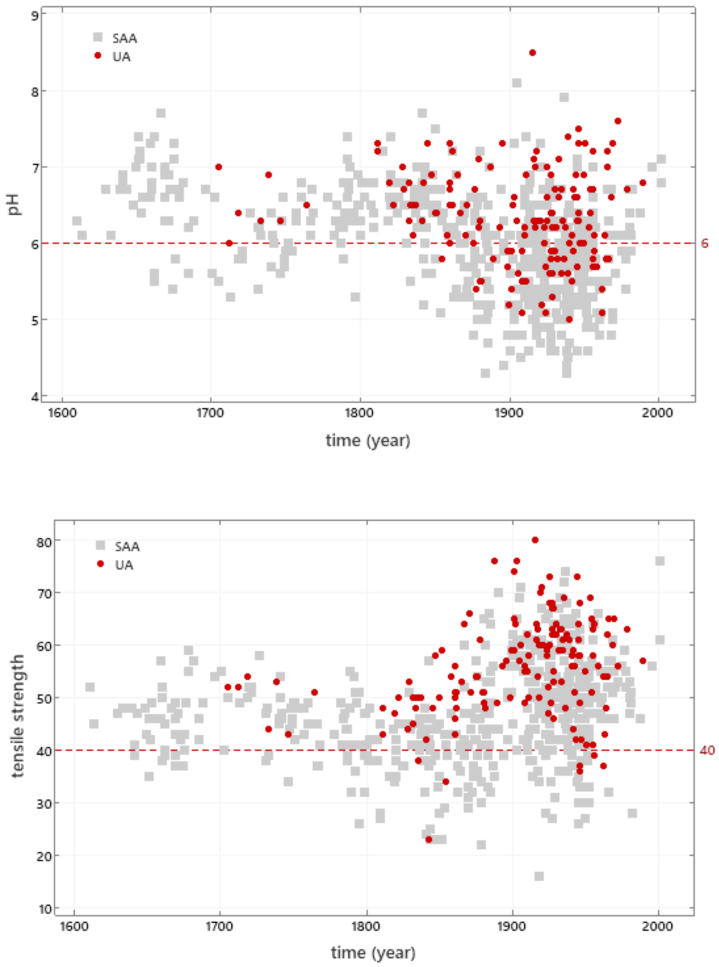


FIGURE 8.18: Scatterplot showing the pH (*top*) and tensile strength values (*bottom*) for paper at the SAA and UA, dating from 1600–2000. The dashed horizontal line separates objects with pH < 6 and TS $<$ 40.

When the pH and TS values of the UA are plotted on a timeline, the UA values are seen to be within the range of values seen in the SAA, but mostly in the higher regions (Fig. 8.18). These results seem to indicate that the differences cannot be explained by an instrumental error. The differences between the two datasets might also respond to the maintenance of the collections and storage conditions in the past. In this sample, it remains unclear why the average DP is lower at the UA than the SAA.

8.2.2.3 Simulation

Environmental data shows that climate in the UA repositories is stable and follows the requirements prescribed in the Archive Regulation 2014 (Fig. 8.19). Therefore, as baseline the set point of 18 °C and 50% RH was chosen for an exploration of slight changes to the storage environment.

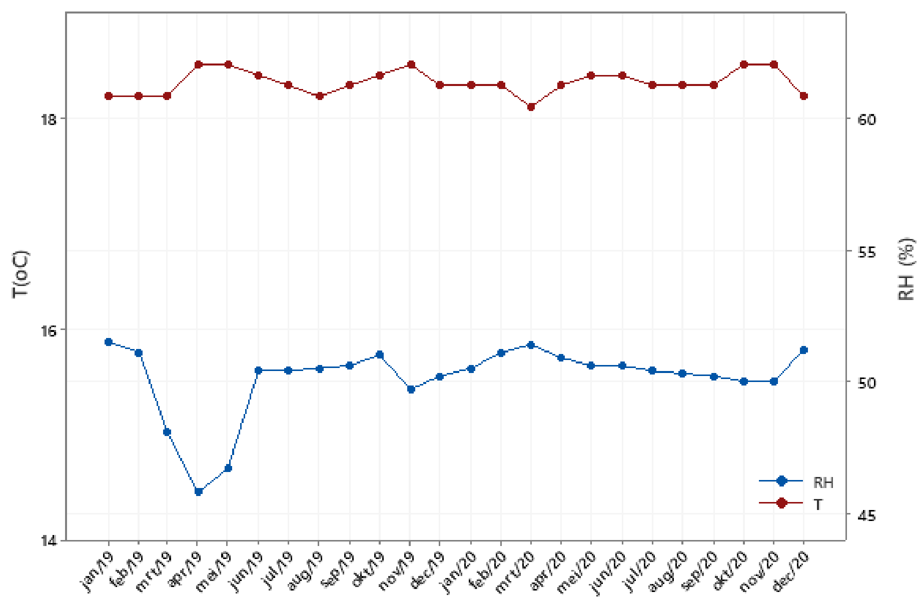


FIGURE 8.19: Monthly average temperature and relative humidity in one of the repositories of the UA over a period of two years.

Table 8.5 summarises the conditions and results of six scenarios. The experiments are in line with previous results seen in the simulations in Chapter 6 (section 6.7). By slightly decreasing temperature and/or RH, the percentage of the collection in poor condition is halved at a time horizon of 500 years, compared to the baseline. However, as in the NIOD case study, the percentage of the collection in poor condition is already relatively small, so the benefit of changing the storage conditions can be questioned. It is interesting to note that changing temperature has a positive effect on the parts of the collection most at risk, independently whether the RH is maintained at 50% or even increased to 55%. Decreasing the temperature from 18 to 16 °C seems a small difference, however, a temperature gradient across a container (read repository) will affect the RH values (Padfield; 2006). This is because, as discussed in Chapter 7 (section 7.3.2.1), the lower the temperature, the lower the air's moisture-holding capacity. The buffer capacity of the paper, as well as the capacity of the HVAC, will determine whether RH can be maintained close to 50%. However, the experiments shows that even the RH increases to 55%, the most chemical unstable papers in the collection will benefit from lowering the temperature by 2°C. These findings are in line with the recommendations given by Michalski and Pedersoli (2016).

TABLE 8.5: Comparison of strategies for the UA collections after a time horizon of 500 years.

Scenario	***T***	**RH**	**Deacidification**			**Output**
			rate	**duration**	**annual cost**	**poor–fair**
	(°C)	(%)	(m/year)	(year)	(€)	**condition** (%)
1 Baseline	18	50	0	0	0	**7–38**
2	16	50	0	0	0	**2–30**
3	16	45	0	0	0	**2–30**
4	16	55	0	0	0	**3–37**
5	16	40	0	0	0	**0–27**
6	18	50	100	10	144,000	**5–36**

In the same experiments (Table 8.5), we also explored whether deacidifying part of the collections could have a significant long-term impact on the condition of the collection. The chosen scenario was the deacidification of 100 m per year over a period of 10 years. Currently the UA is digitising an average of 120 m per year (ca. 4,000 records per year). The calculated digitisation price of one metre at the UA is €1399 (slightly higher than the €1100 calculated for the SAA). Therefore, in the case of the UA, the required annual budget for deacidification ($100\,\mathrm{m} \times €1440\,\mathrm{m}^{-1} = €144,000$) is comparable to the costs of digitisation ($120\,\mathrm{m} \times €1399\,\mathrm{m}^{-1} = €167,888$). However, the results show that the impact of such a measure of deacidification is almost unnoticeable, and a major investment would be required in order to obtain similar results to slightly lowering the temperature and RH.

TABLE 8.6: Comparison of strategies for the UA collections on wear and tear after a time horizon of 100 years.

Scenario	Input			Output		
	T (°C)	RH (%)	Rate digi. (req./year)	Critical m (σ)	Poor m (σ)	Fair m (σ)
1 Baseline	18	50	0	92 (38)	110 (14)	348 (34)
2	16	45	0	82 (20)	78 (12)	363 (38)
3	18	50	4,000	84 (34)	46 (5)	272 (31)

In the next series of experiments (Table 8.6), we compare prospective preservation measures on the risk of accumulating tears due to handling in the reading room. The input of the first 10 years is the reproduction of the dataset of the last 10 years, as discussed in Chapter 5 (section 5.6.2). Around 4,000 m of the collection (166,268 records) were requested during that period. For the baseline, we explored the accumulation of wear and tear if the digitisation had stopped from 2020, but with collection use remaining the same for a period of 100 years. In that case, at the end of the run, ca. 550 m are at risk of accumulating tears. This is the result of both use and the chemical condition at the time of use. Then, we compared these results with the impact of two conservation measures: slightly lowered temperature and RH, and digitisation. As expected, reducing reading-room use via digital access would decrease the number of records that accumulate tears across the three condition categories (critical, poor and fair). The results of changing the storage conditions but continued baseline-level reading-room use although also expected, are interesting. If we slow down the chemical degradation, the records accumulate tears at a slower rate. Therefore, the results show that this preservation measure would reduce the number of records reaching the critical and poor condition stage, but it is interesting to notice that the slower (albeit continued) accumulation of tears results in a higher number of records remaining in fair condition.

8.2.3 Conclusions

The two presented case studies provided the opportunity to test, firstly, the process of collecting input data for the Archives Preservation Management (APM) model, and, secondly, how the outputs can serve to support preservation management decisions. Regarding the input data, three different sources of data are required for each part of the model (preservation, access and costs). The data collection for each part of the model presented different challenges.

For the data for the preservation part of the model, a survey, a common tool for assessing archival collections, can be conducted to collect data on the chemical characteristics of the collection (DP and pH). If the SurveNIR, a tool that collects data on the DP and pH of the paper non-destructively and on a large scale (ca. 100 measurements per day) is accessible, then the selection and measurement of ca. 200 observations can be carried out within four days.

However, although the advantages of the SurveNIR tool are clear, unfortunately there is also an important limitation, namely SurveNIR does not provide DP values for groundwood paper. To fill the missing data, we presented a model that predicts the DP value from other SurveNIR measurements. The error of the presented model for DP is ca. 400. This is also the error reported in the literature related to the SurveNIR's instrumental error. It is good to realise that, as we are interested in finding the moment when the paper collections will reach DP < 800 (fair condition) or DP < 300 (poor condition), but are hampered by the instrumental error and the model in predicting DP values of the group of interest (groundwood paper), there is a clear uncertainty for those groups of papers that we are most interested in, namely those that might reach the critical DP of 300 in the next hundred years. Keeping this uncertainty in mind, the output of the model should not be used to predict a collection's life expectancy, but only for comparing the impacts of preservation measures, as already pointed out in previous chapters.

While designing and conducting a survey is more likely to be closer to the work and skills of the end-users of the APM model, who are responsible for preservation management, obtaining and analysing the input for the access of the model requires substantial data analysis skills. The raw data on collection use, provided by the collection management systems, needs to be pre-processed and analysed in order to obtain the input data for the model (e.g. number of SoD requests, number of records requested and digitised per year, etc.; see Table 8.1 for the list of inputs). Although such analysis will provide a better fit of simulations to the actual realities, some simplifications can be done when data are not available, since the archival collections studied in this project indicate certain commonalities such as the same Pareto frequency distribution of the accumulated number of requests per record, and how different digitisation projects correspond to the most heavily requested records in the reading room.

Regarding the input data for the costs, most of the inputs are relatively straightforward (e.g. price of digital storage per scan; time needed for tasks within a process). However, these case studies showed that getting the right data and values can be more difficult than one may have anticipated. To our knowledge, based on our experience during this project, there are two main reasons for the difficulties encountered: first, the input data is provided by different departments within the archive, and therefore, the collaboration of large number of colleagues is required. Secondly, some of the costs need to be estimated as averages. The combination of these two factors can make it difficult, or even impossible, to collect the data within a short time frame. We expect that the archives in this study are not an exception and that similar problems might be encountered in other institutions when collecting input data for the APM cost model. These difficulties, together with the limitations of the cost model, especially when the energy costs in repositories are not available, had as consequence that full use of this part of the model was not possible, so only the easily obtained costs could be introduced as input, entailing a simplification of the results.

As seen in Chapter 6, the degradation of good-quality paper (i.e. high DP values, often in combination with pH-neutrality or slight acidity) are expected to reach a critical DP value far beyond the 500-year horizon. Hence, we might argue that, as we are mostly interested in the deterioration of a certain group of papers (with low DP values, often in combination with low pH), then we should focus our attention on this group instead of on the whole collection in order to obtain a good characterisation of the group of interest. In addition, when analysing the whole collection, if the group of interest turns out to be relatively small, the majority will dominate, making it difficult to put the results into true perspective.

In light of these arguments, in the Utrechtsarchief case study, the target population (the complete collection of individual items from which samples are selected) was defined as those collections dating between 1850 and 1950, whereas, in the

NIOD case study, the whole collection was defined as the target population. The survey data of the two case studies showed that, although the selected papers mostly date from the period when less chemically stable paper is expected, still, the simulation results showed that just a small part of the collection are prone to reach a DP lower than 800 after a time horizon of 500 years. Nevertheless, we could still test the value of the information provided by the APM model in comparison to information already obtained during the data collection process. Here follow some reflections on this topic:

1. From the perspective of the end-user:

    - Regarding the data analysis required for the access part of the model, the data analysis outcomes are so rich that they can be considered to provide enough evidence and insights to justify and to develop digitisation strategies, without having to use models. The data analysis is already capable of providing information on how digitisation is affecting the use of the physical collections, as well as the effects of different digitisation strategies.
    - Similarly, those end-users with (advance) knowledge on preservation will be able to form accurate opinions on the urgency of undertaking preservation actions, by analysing the data collected on chemical characteristics of the collections, and on demand levels in the reading room.
    - That would mostly be the case when dealing with durable collection with a relatively low level of demand. One may argue that, in those cases, the model just confirms what it was expected and that data analysis would have been probably sufficient.
    - Even when this seems true, the added value of using a simulation model is that it allows us to test the contribution of each parameter by quantifying his impact on the end result, even for those parameters that might seem ‘obvious’ to the experienced end-user. By using mathematical models, mental models become communicable and testable, and hence promote evidence-based decision-making within institutions.

2. From the perspective of the researcher:

    - Simulations are a valuable virtual laboratory for conducting sensitive analysis: from testing whether (parametric) uncertainty matters, to revealing those parameters that contribute the most to the end results.

- However, data analysis has shown to be crucial to formulating hypotheses and, subsequently, to interpret and put the results into context.

References

Baird, B. J., Krentz, J. and Schaffner, B. (1997). Findings from the condition surveys conducted by the University of Kansas Libraries, *College & Research Libraries* **58**(2): 115–126.

Brown, N., Coppola, F., Modelli, A., Amicucci, F., Lichtblau, D. and Strlič, M. (2020). Non-destructive collection survey of the historical Classense Library. Part I: Paper characterisation, *Heritage Science* **8**(1): 1–11.

Brown, N., Lichtblau, D., Fearn, T. and Strlič, M. (2017). Characterisation of 19th and 20th century Chinese paper, *Heritage Science* **5**(47).

Cséfalvayová, L., Strlič, M. and Karjalainen, H. (2011). Quantitative NIR chemical imaging in heritage science, *Analytical Chemistry* **83**(13): 5101–5106.

Dong, Y. (2020). Multivariable linear regression analysis of degree of polymerization and paper composition, *Master thesis*, University College London.

Duran, C., Grau-Bové, J. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part I: exploring the concepts of reliability and epidemiology, *Heritage Science* **7**(1).

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2017). Data mining in collections: From epidemiology to demography, *ICOM-CC 18th Triennial Conference, 2017 Copenhagen*.

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part II: an epidemiological study, *Heritage Science* **7**(11).

Dutch Admin. (1995). Act of the Minister of Education, Culture and Science, of 28 April 1995, regarding the sustainability and proper and accessible condition of archive records and the construction and setup of archive rooms and depositories (Archiefwet).
URL: *https://wetten.overheid.nl/BWBR0007376/2020-01-01*

Dutch Admin. (2014). Regulation of the Minister of Education, Culture and Science, of 17 December 2009, no. WJZ/178205 (8189), regarding the sustainability and proper and accessible condition of archive records and the construction and setup of archive rooms and depositories.
URL: *https://wetten.overheid.nl/BWBR0027041/2014-01-01*

Gurnagul, N., Howard, R., Zou, X., Uesaka, T. and Page, D. (1993). The mechanical permanence of paper: a literature review, *J. Pulp Pap. Sci.* **19**: 160–166.

ICN (2000). De steekproef alshulpmiddel bij collectiebeheer, *ICN Informatie* **4**.
URL: *https://kennis.cultureelerfgoed.nl/index.php/Steekproef_-_hulpmiddel_bij_collectiebeheer*

Kes, M. and Christensen, B. E. (2013). A re-investigation of the Mark-Houwink-Sakurada parameters for cellulose in Cuen: A study based on size-exclusion chromatography combined with multi-angle light scattering and viscometry, *Journal of Chromatography A* **1281**: 32–37.

Ketelaar, E., Huysmans, F. and van Mensch, P. (2016). Netherlands: archives, libraries and museums, *in* M. J. Bates and M. N. Maack (eds), *Encyclopedia of Library and Information Sciences*, third edition edn, CRC Press, pp. 3874–3900.

Konsa, K. (2007). Condition survey for the Estonian national preservation policy, *Restaurator* **28**(4): 239–255.

Lichtblau, D. and Dobrusskin, S. (2009). Das SurveNIR – System zur Bewertung des Zustands und des Alterungsverhaltens von Papier, *Technical report*.

Lichtblau, D., Strlič, M., Trafela, T., Kolar, J. and Anders, M. (2008). Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics – a new instrument, *Applied Physics A* **92**(1): 191–195.

Łojewski, T., Zieba, K. and Łojewska, J. (2010). Size exclusion chromatography and viscometry in paper degradation studies. New Mark-Houwink coefficients for cellulose in cupri-ethylenediamine, *Journal of Chromatography A* **1217**(42): 6462–6468.

Mahgoub, H., Bardon, T., Lichtblau, D., Fearn, T. and Strlič, M. (2016). Material properties of Islamic paper, *Heritage Science* **4**(34).

Małachowska, E., Dubowik, M., Boruszewski, P., Łojewska, J. and Przybysz, P. (2020). Influence of lignin content in cellulose pulp on paper durability, *Scientific Reports* **10**(1): 1–12.

Michalski, S. and Pedersoli, J. L. J. (2016). *The ABC Method. A risk management approach to the preservation of cultural heritage*, Canadian Conservation Institute, Ottawa.
URL: *https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf*

Orr, S. (2021). Multivariable linear regression analysis of degree of polymerisation for groundwood paper, *Unpublished - personal communication* .

Padfield, T. (2006). The interaction of water vapour with paper, *Conservation Physics* .
URL: *http://www.conservationphysics.org/vapap/vapap.pdf*

Sobucki, W. (2003). Survey of the preservation status of the 19th and 20th Century collections at the National Library in Warsaw, *Restaurator* **24**(3): 189–201.

Strlič, M. and Kolar, J. (eds) (2005). *Ageing and stabilisation of paper*, National and University Library, Ljubljana.

Strlič, M., Kolar, J. and Lichtblau, D. (2007). The SurveNIR project—a dedicated near infrared instrument for paper characterization, *in* T. Padfield, K. B. Scientific, C. Mads, C. Christensen, B. Eshøj, M. Ryhl-Svendsen and D. Thickett (eds), *Museum microclimates*, National Museum of Denmark, Copenhagen, pp. 81–84.
URL: *http://www.conservationphysics.org/mm/musmic/musmic150.pdf*

Strlič, M., Liu, Y., Lichtblau, D. A., De Bruin, G., Knight, B., Winther, T., Kralj Cigić, I. and Brereton, R. G. (2020). Development and mining of a database of historic European paper properties, *Cellulose* **27**(14): 8287–8299.

Trafela, T., Strlič, M., Kolar, J., Lichtblau, D. A., Anders, M., Mencigar, D. P. and Pihlar, B. (2007). Non-destructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation, *Analytical Chemistry* **79**: 6319–6323.

VanSnick, S. and Ntanos, K. (2018). On digitisation as a preservation measure, *Studies in Conservation* **63**(sup1): 282–287.

Zou, X., Uesaka, T. and Gurnagul, N. (1996). Prediction of paper permanence by accelerated aging. Part I: Kinetic analysis of the aging process, *Cellulose* **3**: 243–267.

Chapter 9

Conclusions

> *'The unbuilt dome of Santa Maria del Fiore had therefore become the greatest architectural puzzle of the age. [...]. Even the original planners of the dome had been unable to advise how their project might be completed: they merely expressed a touching faith that at some point in the future God might provide a solution, and architects with a more advanced knowledge would be found.'*
>
> King; 2000,p. 14

We have produced a mathematical model to compare the impact of several preservation measures on the preservation and use of collections, considering the budgets available in institutions. The model, referred as the Archival Preservation Management (APM) model, is composed of three interconnected sub-models, where the output of one sub-model is used as input in another sub-model. As we built further upon the model in each chapter, we showed that the sub-models can be used independently or in combination with each other. In the preservation sub-model, one may explore the impact of deacidification and changed environmental conditions on the chemical preservation of the collections, including—or not—the costs of those particular measures. The access sub-model inform us about how the number of requests in the reading room evolves depending on the amount of collection material becoming digitally available. This part can be then linked to the costs of digitisation, as well as to the usage patterns in the preservation sub-model and how that affects the accumulation of wear and tear, in conjunction with the chemical condition of the collections. The third sub-model calculates the costs involved in the processes of preserving and providing access to the collections, depending on the choices made.

The three research questions formulated at the beginning of this project have ridden along with us throughout the development of the model. It seems, therefore, a logical choice to group the conclusions with their associated research questions. In addition, from the beginning we also formulated a particular aim, namely that the developed model was meant to be used by those engaged in the preservation of collections. In this last chapter, we reflect on whether we succeed in this quest, together with a discussion on the strengths and limitations of the proposed model. Finally, we suggest areas of future work based on the unsolved challenges encountered during this thesis project.

9.1 Research question 1: Can the preservation management of archival and library collections be modelled as a complex system?

Beyond doubt, the answer to this question is a definitive yes: the preservation management of collections is a complex system. We are explicitly differentiating our system from merely complicated systems, in which, once the system has been formulated, it will lead to the same result. Conversely, complex systems are characterised by their uniqueness, and therefore, the outcome will be more difficult to predict beforehand.

At a macro-level, the complexity emerges from the different parts that the whole system is composed of, and whose purposes are not necessarily the same as the purpose of the whole system. In the case of the preservation management of archival collections, we have identified three parts or, more specifically, three sub-systems: preservation, access and cost. Whether these sub-systems could be equally found in other heritage collections remains to be answered; the system under study in this thesis has been narrowed to paper collections, and, more specifically, paper archival records. We have seen that potential competition may arise between the provision of digital access to the collections versus the preservation of the physical collections. Despite the undeniable benefits of digitisation, from enhancing collection access to reducing the risk of mechanical degradation due to handling, the discussion with practitioners made clear that it is feared that digitisation will continue growing, claiming more resources to the detriment of preservation, following the pattern of *success to the successful*, or maybe of *shifting the burden*; only in the long term we will realise that the measures that we were taking were not targeting the records more at risk.

The APM model has been developed with the aim of comparing the impact of preservation measures on the three sub-systems—preservation, access and cost. Acknowledging that we are dealing with a complex system means accepting that conflicts between the sub-parts might not be soluble. Nevertheless, with a model, such as the APM model, we can still identify leverage points where intervention may be fruitful.

At a micro-level, complex systems are characterised by the many components, with individual characteristics, interacting among each other and with their environment. We have hence observed that the same preservation measure might have a different impact when applied in different types of collections, due to their different characteristics and use, but also depending on when the measures are applied. The computational experiments used datasets collected from actual collections, and we created datasets for different types of collections, representing acidic, modern and mixed collections.

The literature review showed that choosing either a macro- or micro-view of system behaviour determines which simulation modelling approach will be the most adequate: either system dynamics (SD), with its macro-level approach, or agent-based modelling (ABM) with a micro-level approach. We have explored both approaches, and decided to use mainly ABM to build the APM model. The ABM approach is closer to the system under study, and allows to apply micro-scale modelling where archival records, represented by agents, can undergo a specific development depending on chemical characteristics, instances of access, or whether they have deacidified or digitised. The power of micro-modelling is, at the same time, its main drawback: from the extensive number of inputs needed, to the opacity of grasping how the outputs are generated. Finding the right level of detail in the model has been a continuous thread throughout this thesis.

Through computational experiments, and using generated datasets representing archival collections and datasets obtained by sampling actual archives, we have shown the power of simulations. We have tested the present (baseline) preservation strategy against other strategies, considering the effects of changed storage conditions and/or deacidification, and examined how digitisation programs affect physical collection use in the reading room. The APM model is a virtual laboratory where the effects of duration and timing of various preservation measures can be explored through the possibility of virtual intervention at each step of the simulation run.

Due to the chosen ABM approach, the APM model is composed of deterministic elements (e.g. degradation of cellulose) as well as stochastic elements (e.g. selection of records for digitisation or reading-room access). We have found that, for the deterministic part of the model (chemical preservation), a collection of 2,000 agents is sufficient, whereas for the stochastic part (access) 4,000 agents are needed to obtain a better fit.

9.2 Research question 2: How can data collected in actual archives and libraries be used for the mathematical formulation of the model and/or to validate the model?

When possible, we have incorporated well accepted models that are already available. We found that some aspects of preservation have been influenced by important research, such as the development of the dose–response equation for the acid-catalysed degradation of cellulose, while other areas are still in a more nascent phase of development (e.g. models to calculate the energy costs in repositories), and others, such as the modelling of collection use in the reading room, have received very little attention.

We have started filling this research gap by addressing how digitisation affects collection use in the reading room by analysing the usage data that are automatically collected by collection management systems. These records form the input data needed in the access sub-model, and have also helped for validation. The two main findings are the Pareto frequency distribution found for the instances of use in the reading room, and how the digitisation of heavily accessed records in the past is resulting in the reduction of requests for reading-room access.

We have built the model for reading-room access requests in SD and in ABS. From the SD model, we have learnt that the decreased number of access requests in the reading room is the consequence of digitising records that were previously being accessed there. But, from the ABM approach, it became clear that the decrease happens in relation to how often the records have been requested in the past. The ABM approach takes the records' access histories into account, resulting in a better fit with the actual data.

The archive-based case studies showed a noteworthy similarity in the frequency distribution of the accumulated access requests over a period of 10 to 20 years. The common Pareto distribution describes how 50–60% of the accessed collections have been requested just once over the observation period, and only 5% were accessed more than once per annum. This uneven level of interest has strong implications for the modelling of accumulated wear and tear. The wear-and-tear function calculates the number of tears per 100 sheets, depending on the DP and number of requests. As the group of objects requested is small, it is important to model the objects with low DP and high request frequencies. The model takes the DP and the number of requests into account when calculating the number of tears, but the selection of records for access, digitisation or deacidification occurs randomly. Therefore, for now, the model provides only a broad indication of the *risk of tear accumulation*, in contrast to predicting the actual accumulation of tears.

The wear-and-tear function takes into account both DP and instances of use. But, for items rarely used, other factors might play more important roles (e.g. thickness of the stack, type of housing) in mechanical degradation. We have conducted epidemiology studies, focusing on acidity and handling, as an attempt to validate the wear-and-tear function. For now, no conclusive results have been obtained, due to the difficulty of isolating factors without further, and longer-term, detail about the collections' maintenance and use histories. However, the insights obtained so far from the observational studies encourage us to continue this line of research in the future.

The model includes only digitisation of the collections as a main driver of the decrease in requests observed in archival collections, and most of the variables (e.g. number of digital requests per year) are modelled exogenously. Nevertheless, we have attempted to identify the factors to model the willingness of readers to choose scan-on-demand (SoD) access in favour of the reading room, by conducting a survey among archival readers. According to the survey, expedience matters: if the delivery time of a scan is reduced to less than one week, more visitors will choose the SoD service. The survey also revealed that a small group of readers resists SoD, and another small group chooses exclusively SoD. Remarkably, most of the readers interchange the use of SoD with visiting the reading room. These outcomes could be the used as basis for a more elaborated version of the APM model.

For the preservation sub-model, we have conducted surveys to characterise the collections (pH and DP of the paper), dating after 1850. Confirming earlier results, the condition of the surviving archival collections was found to be more stable than is reported for (acidic) library holdings. Crucial for this type of survey is the availability of a technique, such as near-infrared measurement, that allows to collect data at high speed in order to meet the required sample size within a practicable time. The drawback of this technique is that, since the DP of groundwood paper is complicated to measure, the instruments do not currently provide direct DP measurements for these groundwood papers, which constitute an important group of interest due to their expected low DP and pH values. However, we made a first attempt to develop a regression model to predict the missing values that went some distance on this front.

9.3 Research question 3: To what extent can the simulation modelling provide quantitative data to enable the comparison of different preservation actions, such as climate control, deacidification or digitisation?

Regarding the preservation part of the model, we have seen that there are several sources of uncertainties, due to instrumental errors when measuring the DP and pH values, the model proposed to predict the DP values of groundwood paper or the dose–response function to model cellulose degradation. This uncertainty will matter the most for the group of records that we have been most interested in throughout this thesis, namely those that will reach the DP of 300 within 500 years. Identifying this group is a matter of immediate practical importance, as even a slight lowering of storage temperature reduces the upper limit on the pH and DP ranges of interest by 0.5 and 100, respectively (Fig. 9.1). The narrowness of these values explains the model's lower predictive power which, nevertheless, still provides a tool for exploring explore what-if scenarios, at least semi-quantitatively, for the collections as a whole.

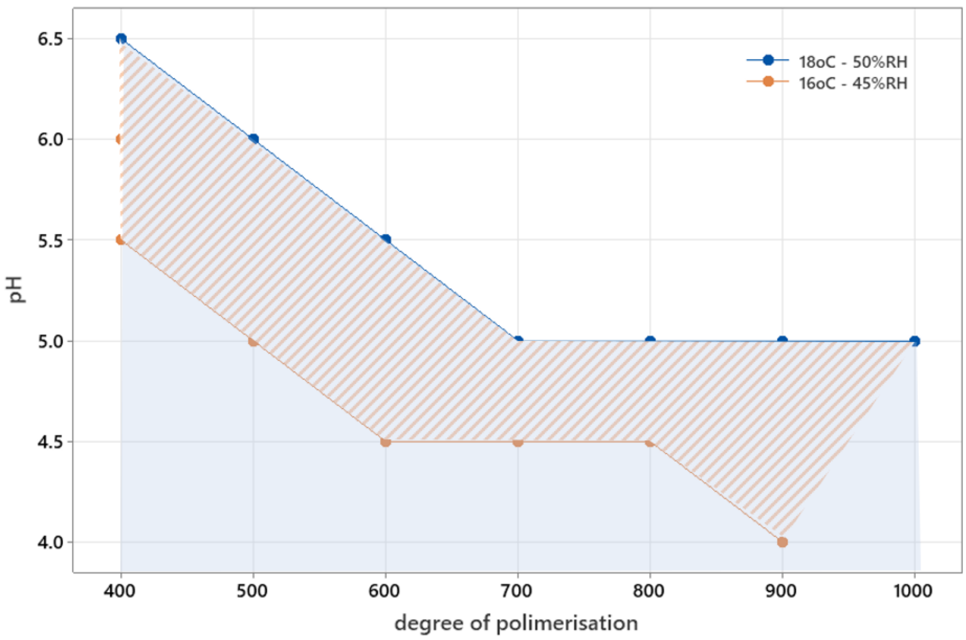


FIGURE 9.1: pH values of records that are expected to reach the critical DP of 300 within the time horizon of 500 years, depending on the DP_0 value and according to two storage climate scenarios. If the temperature and RH are slightly lowered, the group of records at risk is reduced from those with a pH value between 5.5 and 6.5 to those with a pH value between 4.0 and 5.5.

This general characterisation of the collections—from the frequency distributions of DP and pH—will determine how many of the group of interest will decline to poor or fair condition within a given time horizon and, therefore, how effective the measures are. As all of the collections examined include records in this risk group, slightly lowering the temperature will have a noticeable effect on reducing the percentage of records at risk, but similar results can also be obtained by deacidifying part of the (acidic) collections.

The above illustrates how decision makers are confronted with three options: maintaining the present preservation strategy, taking instant action by the deacidification of acidic part of the collection, or taking the long-term measure of changing the storage environment. During the decision-making process, the model shows its value in permitting micro-simulations that allow all combinations to be tested, from different types of collections to the durations and timing of measures. Being able to fully explore possible scenarios provides decision makers with more valuable information to decide which is the 'best' option for their own collection, than if they were merely briefed about expected results.

This thesis did not include a full modelling of the deacidification costs versus the energy costs of maintaining the repositories under different environmental set-points. The APM model would benefit from the development of HVAC-aware energy consumption models, specific for repositories. Although these models are currently being developed, until they become available as working tools, an important aspect of managing archival collections is missing from the APM model, and, more broadly, the full challenge of dealing with finite financial resources.

The analysis of usage and digitisation data has shown that scan-on-demand is succeeding in digitising those records frequently accessed in the reading room, since the readers are the ones making the selection. Interestingly, digitisation projects started by the institution itself are slightly less successful on this front, either because whole archives are selected for digitisation (including records that have never been accessed), or because popularity of the archive is not the only driver behind digitisation projects. This information is used in the access part of the model to predict how the number of requests in the reading room is expected to decrease, depending on the digitisation strategy.

However, due to the Pareto distribution describing the frequency of instances of use, from a preservation point of view, the predictions of the access part of the model are of limited relevance, as handling in the reading room will not be the decisive factor for the accumulation of wear and tear. Nevertheless, predicting how reading-room and digital services will evolve in the next years informs future strategies on the providing collection access.

Regarding the costs related to digitisation, the model makes it clear that such a service requires a considerable investment not only to make the scans, but also for storage, and for making the scans available online. Although institutions have been digitising for years, we have noticed that clear digitisation strategies are still not fully in place, and the importance of including the mid- and long-term costs of the whole digitisation process is not always recognised during decision making.

From the start of this thesis, we have argued that, when modelling a complex system, the model should include the main variables that explain how the system behaves. In our case, those variables related to preservation, access and costs. As the model's developers, we tested the relevance of several variables. Throughout the development of the model, we found out that, if we want to link the accumulation of tears with usage, a great level of detail is needed in the modelling, which makes the model unsuitable for our intended end-users. In addition, in view of the analysis of the records' usage histories, the access sub-model is, at the same time, less influential as long as the annual mean number of requests is low. Therefore, we can conclude that, whereas the preservation part (including future costs) provides relevant information and meets the aim of informing decisions in a clear way, the access part finds its strength in exploring questions related to collection management (e.g. how to provide access), but would require simplification for use in a larger model. That simplification could, in the future, be advanced through further sensitivity analysis.

9.4 Strengths and limitations of the proposed model

Let us in this concluding section reflect on the strengths and limitations of the proposed model in the context of the reviewed models and tools in Chapter 2.

Compared to other models and tools, one of the most innovative aspects of the proposed model is that the exploration of what-if scenarios is at the core of the model. In the proposed model, simulation modelling is applied to test preservation strategies, where input data can adopt any value, not only at initialisation but also at any point in the lifetime of collections. By doing so, the proposed model differentiates itself from the indoor climate evaluation models that use detailed data collected in repositories as input data, or risk assessment where what-if scenarios to evaluate the effectiveness by different options on reducing risks are not formalised, but estimated by means of mental simulations. Differently from the existing models, the APM model is designed to explore strategies by 'playing' with different inputs to learn more on the effect of actions. Although a limited range of aspects has been included for now, the model has demonstrated its applicability and value as virtual laboratory where decision-makers can explore the effect of preservation actions which cannot be tested in the real world.

One might therefore argue that the main contribution of this PhD has been the application of simulation approaches in the field of heritage collections. Particularly, agent based modelling (ABM) allowed us to build a model very close to the real world, for example by defining states of the agents (e.g. archival records) that change according to the actions taken. The output of the model is therefore dynamic, showing outputs over time. This is, in essence, different to the static character of other models, for example those reporting a preservation index.

The proposed model operates in this context where empirical data is used, but a certain abstraction and simplification are also required to operate on a strategic level. Similarly to risk assessment, the focus is on the collection, and not on the individual cases. Therefore the model does not aim to include all preservation aspects involved in preservation management. For example, due to the one year time step during a simulation run, the model does not allow mould risk to be modelled, or inform pest management strategies. Furthermore, the model does not capture processes that do not contribute significantly in view of the whole (for example, the risk of wear and tear during the one handling during the digitisation process). For this (operational) level of information, other models are more suitable and will be needed to complement the information given by the proposed model.

The modular design of the model has two clear benefits. First, the model can be further expended with other variables as more data become available. Secondly, and more importantly, the modular design captures the complex nature of the system under study, with potentially conflicting aims between the sub-models. As shown during the development of the causal loop diagram, practitioners are well aware of this potential conflict of interests; but activities started in a sub-model might have beneficial (side-)effects in other parts of the system. Although these relationships have been identified during the development of the causal loop diagram, they have not been fully developed in the mathematical model. Nevertheless, even the model does not aim to solve the dichotomy, the proposed model might help to explore whether this (general) dichotomy matters in the particular case of archival institutions. By testing different options, one might investigate for instance whether the strategies in one part (digitisation) are having a positive effect on preservation or whether the limited budget (e.g. for deacidification treatment projects) might have a negative long-term impact. The model, in his current state, cannot fully answer these questions, but offers a glimpse of the potential applications of the model.

Let us now balance the strengths of the model against the limitations, which are partly the result of the difficulties encountered during the development of the mathematical model: some of them are characteristic of working with models, whereas others are specific for the APM-model.

- Model boundaries are defined by the included variables, endogenous and exogenous, as well as by the omitted variables, those variables chosen not to be included to keep the model 'manageable'. The APM-model is not an exception and omitted variables have been part of the discussion during the development of the CLD as well as in the mathematical model. For example, regarding the access of collections, a long-term downtrend of visitors in the reading room was observed which might respond to social-cultural factors. Omitted variables are an intrinsic limitation when working with models, but also an essential part that needs to be taken into account when interpreting the outputs of the model.

- The APM-model includes well-accepted models. But even these models are subject to limitations and uncertainties, due to assumptions and simplifications. Cellulose degradation models incorporate uncertainties related to the use of artificial aging to study the kinetics of cellulose degradation. The energy demand model is a small, comprehensive model, that although useful for some purposes, makes assumptions on those aspects that we are particularly interested in. Likewise, wear and tear models for paper collections have only recently been proposed. For now the wear and tear model includes a limited number of variables (for example, thickness of the stack is not taken into account), and the model is indicative of the part of the collections at risk. Herewith an important observation needs to be made that applies for the whole model, namely that even the results are presented numerically, they should primarily be evaluated qualitatively.

- This is even truer for other parts of the model, such as access, that have been modelled for the first time in this research. The proposed access-model is at the initial phase of modelling, when models are meant to support the development of hypotheses.

- The collection of data in the actual setting of the collections has been a spearhead in this research. It is clear that, in order to increase confidence on the proposed model, a larger corpus of historical data is imperative to further develop hypotheses, to validate parts of the model and to be used as input data of the model. Although the benefits of using historical data are well known, this research has shown, once more, that working with historical data is challenging, since this type of data can not be 'controlled'. As long as a larger corpus of data is not available, results will need to be interpreted with caution.

- A tangible impact of this lack of reliable data has resulted on the omission of several variables in the final mathematical model. Variables that were included in the causal loop diagram, e.g. related to the preferences of the readers, or a more ambitious model of wear and tear, remain for now in the design phase.

- The challenges of working with historical data were further stressed in the case studies. As empirical data can not be controlled, the case study data did not provide us with the opportunity to fully test the model and its relevance to collection management.

- Sensitivity analysis has been conducted to illustrate how this type of analysis could be used to investigate whether parametric uncertainty matters. Although it was out of the scope of this project, a systematic approach to sensitivity analysis will contribute to clarifying the level of uncertainty present in the model. Similarly, sensitivity analysis could be extended to identify the variables that have greater impact in the dynamic behaviour of the system, and at the same time, supporting whether variables could be omitted in the model.

- Compared to other fields that also deal with complex systems and long-time horizons, such as climate models, the modelling of (paper) heritage collections is still nascent. Predictions and forecasting are far from possible due to the enormous corpus data and research that is needed to achieve this last stage in modelling. The model is for now a comparative tool that helps to develop and test hypotheses.

- The proposed model, referred to as Archival Preservation Management model, has mostly been tested for archival paper collections. Although it can be expected that most of the dynamics seen in the model are also applicable to library collections, this has not been specifically addressed or tested in this research.

This analysis has shown that the proposed model, as a whole, should be regarded as 'in progress' rather than 'user-ready'. This research has revealed the potential of the approach, which, for now, offers a departing point for future research to work on those challenges and limitations that could not be fully solved or addressed in this thesis.

9.5 Future work

This last reflection brings us to the last section of this thesis, where future work is suggested to further advance research on the modelling of preservation management of archival and library collections. The suggestions are mostly about finding the balance between increasing detail in some variables where needed to improve the insights generated, versus simplifying the inputs as much as possible to improve the practicability for application.

Without doubt, the most significant improvement to the model will be from including the energy costs of maintaining a certain set-point in the storage facilities. It is imperative that the energy consumption in repositories is modelled to reflect how the HVAC systems found in archival repositories work. The model could then be further extended to contribute to the current debate on whether preservation strategies need to be reconsidered in view of planetary climate change. The development of a model including HVAC energy consumption is currently being carried out by a research group at the Eindhoven University of Technology (Decision Support for Sustainable Archives project). The results of this research are expected to be directly applicable to the further development of the APM model.

Also needed is a better model to predict the DP values of groundwood paper, which are missing in the tools used to collect chemical characteristics of the paper. Groundwood paper is regarded is one of the less durable paper types, so it is essential to incorporate a good understanding of how it responds to preservation measures when modelling the life expectancy of collections.

We have seen that the model could better fulfil its aim of end-user-friendliness if some parts of the model are simplified, in the sense of reducing the level of detail needed for the inputs, as well as the level of detail in the outputs. For example, we have indicated that the access part of the model could be simplified if, instead of modelling how records accumulate instances of use, it could work with an average. It would be interesting to further explore whether the aggregates of the SD, representing an average, could be an option. At the same time, we have seen that some of the inputs are to some extent comparable between the studied archives. Therefore, another area of research could be to find 'default' data descriptors in order to simplify the data analysis, as well as alleviating the demands for input data that the current model requires.

A further simplification could be also tested for the wear-and-tear model, for example by including, as output, the percentage of the collection in a certain condition, but still not digitally available.

On the other hand, one of the most important features in ABM is that agents, in our case archival records, can be characterised in detail. One variable that would be worth exploring is the date of the records. By including the date, a certain probability of being digitised or accessed in the reading room could be included. This new level of information could be relevant because date can be used to determine the material characteristics of the records (e.g. rag, bleached pulp or groundwood paper), and potentially also the popularity of the collections among the readers.

In the end, we chose a bottom-to-top approach as a simulation model, by modelling the records to explore the dynamics of collections, since the focus of the model was mainly on simulating collection lifetimes. We briefly introduced the potential of using SD to understand the system a whole. This macro-approach could be further applied to model how decisions are made in each of the processes in archival and library institutions. Herewith, both the conflicts and the synergies between departments will become more visible.

There are also several other areas of research, perhaps not directly applicable to the model, which could provide valuable information to develop strategies regarding the provision of access to collections. Some areas that we would like to suggest are studies on readers' preferences for the use of SoD or the reading room; how can we measure whether the use of digital content is comparable to the instances of use in the reading room; and whether other trends can be identified that, besides digitisation, are affecting the use of the collections in the reading room.

I strong believe that the study of the collections themselves are the principal source of information to directly inform decision making or, for indirect information, using (simulation) models. This is certainly true for those areas which cannot be easily studied experimentally (e.g. wear and tear). Therefore, it is advocated that epidemiological studies are very much needed to collect further evidence to support decision-making in preservation management.

9.6 Final (personal) reflection

For five years, I have considered myself a hands-on conservator learning modelling for my PhD research. While working on this research, I have been conducting data analyses and learning how to build models, step by step, in the most logical manner. But, at the same time, being a practitioner myself, I found myself in the position of being one of the end-users of the model that I was working on, and caught myself on repeated occasions asking this troubling question: 'Do we really need models, that look into the next (hundreds of) years, to take decisions now, in the present?'

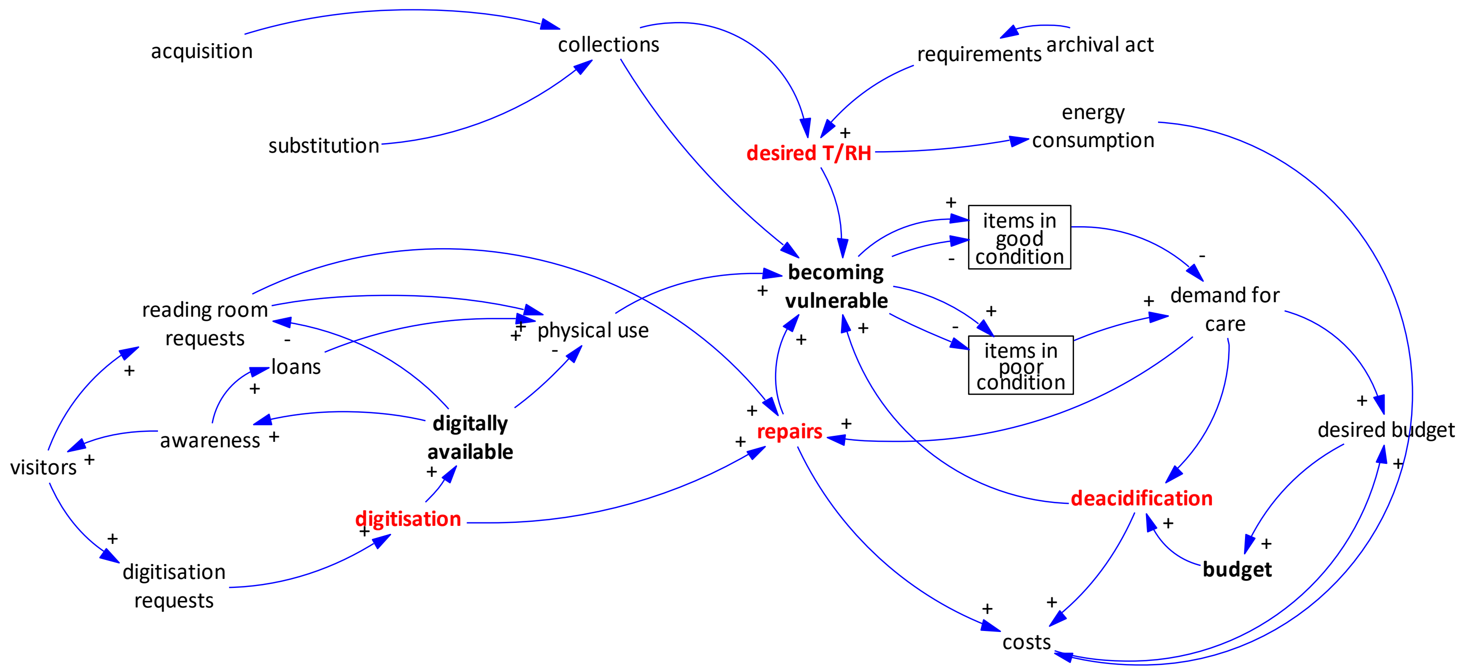
Now, at the end of this journey, it feels that the time has come to answer this question, and, maybe not surprisingly, for now my answer needs to remain two-fold: yes and no. I learned to recognise the power of models and, in particular, simulation, to investigate those measures whose impact goes beyond our own experience. At the same time, I also learned not to underestimate the usefulness of data analysis that provides insights without the need of using models. In short, this PhD project has reassured me that models, data analysis, or probably a combination of both, helps us to justify the preservation choices we are taking, despite all the unknowns and uncertainties, which will be resolved as we further move from the *now* to the *tomorrow*.

Appendix A

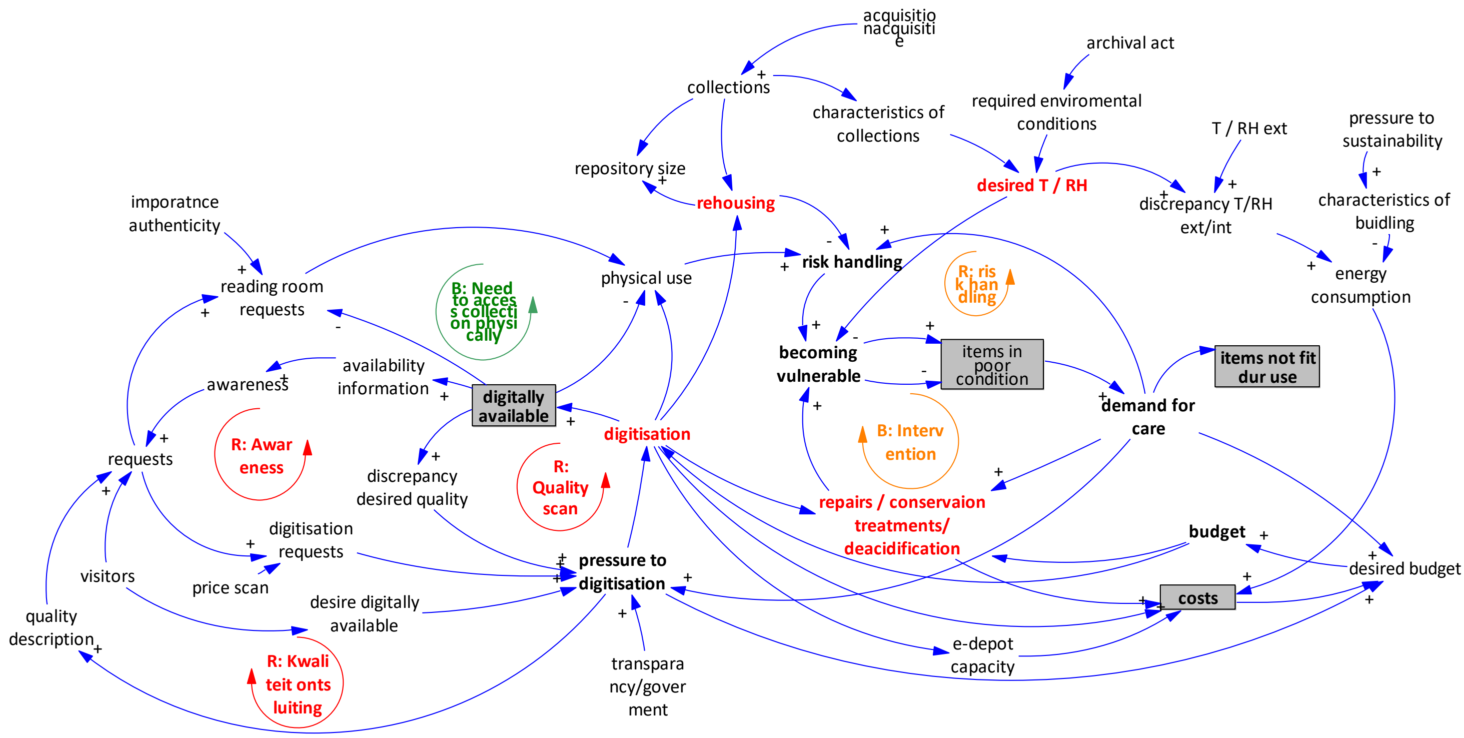
Causal loop diagram

This appendix includes the causal loops diagrams built during the group model sessions. The same exercise was repeated with three different groups of professionals.

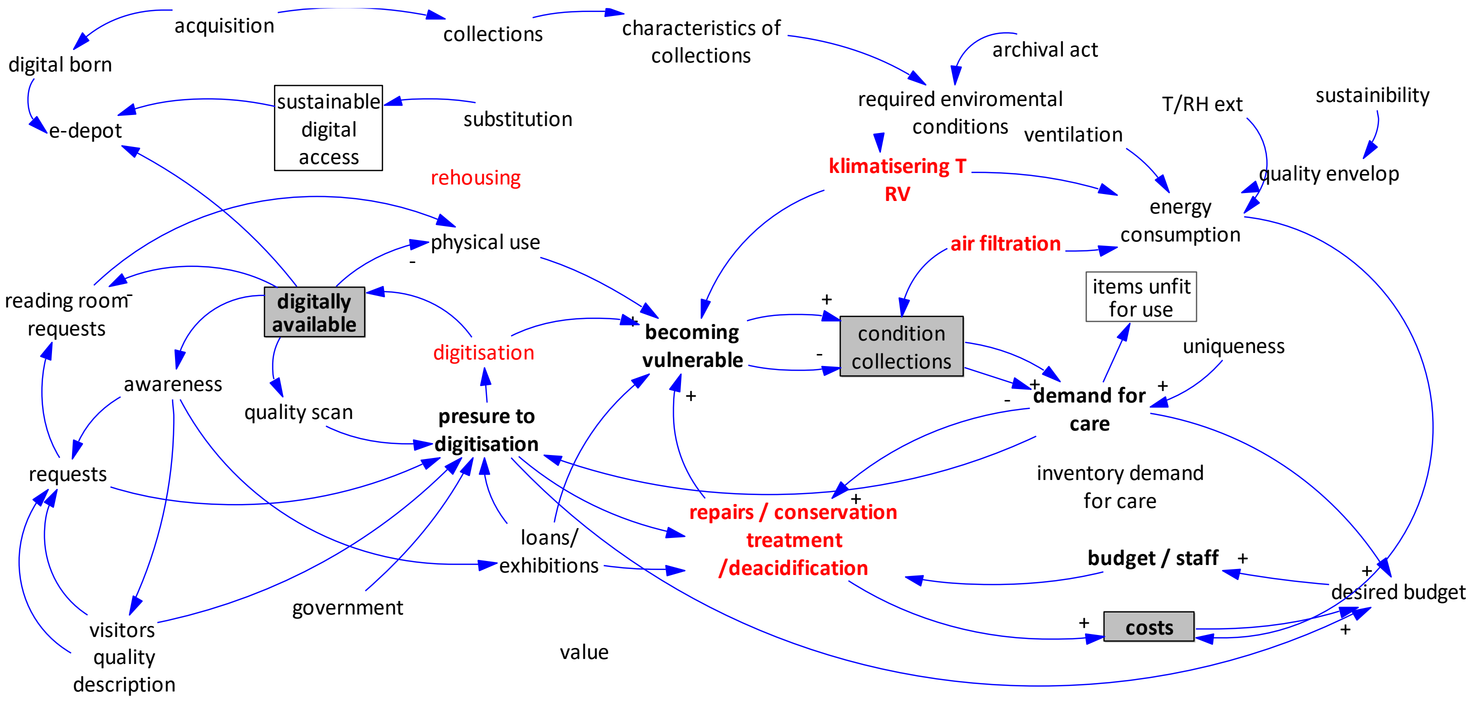
Participatory session group 1



Participatory session group 2



Participatory session group 3



Appendix B

Questionnaire visitors reading room

Questionnaire on the use of the reading room

This questionnaire is part of the research project on archival preservation and management carried out within the Conservation Research Program of Metamorfoze. This project is also part of a PhD project at the University College London (UCL). More information about the project can be found at the blog page of Metamorfoze: https://www.metamorfoze.nl/kennis-onderzoek/kennisblog/wat-zijn-de-effecten-van-papierconservering.

Dear Visitor of the reading room,

Welcome to our reading room. Today you have chosen to access the archival records in the reading room instead of using the online service Scanning on Demand. In order to understand the preferences and considerations of our visitors we would like to ask you a few questions regarding the reading room and the Scanning on Demand services. The questionnaire is anonymous and will take about 5 minutes of your time.

We thank you for answering the questions.

Question 1

Divide 100 points between the following sentences and fill in the missing information:

I would consider using the online service Scanning on Demand instead of accessing the original archival records in the reading room:

a. if the opening hours of the reading room would be more limited, namely .... day(s) per week ...... points

b. if I could place as many requests for digitization as I need to, namely .... requests per week ...... points

c. if the delivery time of the scan would be shorter than 7 days. I would like to get the scan within ...... day(s) ...... points

d. Never. If I can choose, I prefer to access the original archival record instead of a digital version ...... points

e. I have no choice, because one or more archive items that I need are not eligible for digitization. ...... points

f. Otherwise, namely... .... points

TOTAL 100 points

Question 2

Divide 100 points between the following two sentences so that the division of points will reflect your research method.

For my research,

a. I need to access a certain archival record first to know which record I will request next ...... points

b. I can request several archival records at the same time ...... points

TOTAL 100 points

Question 3

How many archival records would you like request today?

.... archival records

Question 4

Which type of visitor do you identify most with?

- ☐ Historian
- ☐ Teacher
- ☐ Student
- ☐ Journalist
- ☐ Research
- ☐ Image researcher
- ☐ Otherwise, namely.......................................

Question 5

How often do you use the reading room service?

- ☐ This is my first time
- ☐ Weekly
- ☐ Monthly
- ☐ A few times a year

Question 6

How often do you use the online service Scanning on Demand?

- ☐ Never
- ☐ Weekly
- ☐ Monthly
- ☐ A few times a year

Appendix C

Calculator for energy use

Energy balance model to calculate energy use described by Padfield (https://www.conservationphysics.org/atmcalc/energyusecalc.html).

Screenshots showing the Excel formulae and layout of the calculator.

AER	1	1/h
U Value	1	W/m2K
Surface to volume ratio	2	1/m
Lighting and appliances	0	kWh/m3
Setpoint T	18	K
Setpoint RH	50	K

Yearly Energy consumption	=SOM(O13:O24)	kWh/m3

A	B	C	D	E	F
	T out	RH out	T in	RH in	Abs H out (g/kg)
Jan	25	73	=B6	=B7	=1000*0,622*(H13/(101325-H13))
Feb	25	73	=D13	=E13	=1000*0,622*(H14/(101325-H14))
Mar	25	73	=D14	=E14	=1000*0,622*(H15/(101325-H15))
Apr	25	73	=D15	=E15	=1000*0,622*(H16/(101325-H16))
May	25	73	=D16	=E16	=1000*0,622*(H17/(101325-H17))
Jun	25	73	=D17	=E17	=1000*0,622*(H18/(101325-H18))
Jul	25	73	=D18	=E18	=1000*0,622*(H19/(101325-H19))
Aug	25	73	=D19	=E19	=1000*0,622*(H20/(101325-H20))
Sep	25	73	=D20	=E20	=1000*0,622*(H21/(101325-H21))
Oct	25	73	=D21	=E21	=1000*0,622*(H22/(101325-H22))
Nov	25	73	=D22	=E22	=1000*0,622*(H23/(101325-H23))
Dec	25	73	=D23	=E23	=1000*0,622*(H24/(101325-H24))

G
Saturated VP at this T (Pa)
=10 ^ (28,59051 - (8,2 * (LOG($B13 + 273,16)) + (0,0024804 * ($B13 + 273,16)) -(3142,31 / ($B13 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B14 + 273,16)) + (0,0024804 * ($B14 + 273,16)) -(3142,31 / ($B14 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B15 + 273,16)) + (0,0024804 * ($B15 + 273,16)) -(3142,31 / ($B15 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B16 + 273,16)) + (0,0024804 * ($B16 + 273,16)) -(3142,31 / ($B16 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B17 + 273,16)) + (0,0024804 * ($B17 + 273,16)) -(3142,31 / ($B17 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B18 + 273,16)) + (0,0024804 * ($B18 + 273,16)) -(3142,31 / ($B18 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B19 + 273,16)) + (0,0024804 * ($B19 + 273,16)) -(3142,31 / ($B19 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B20 + 273,16)) + (0,0024804 * ($B20 + 273,16)) -(3142,31 / ($B20 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B21 + 273,16)) + (0,0024804 * ($B21 + 273,16)) -(3142,31 / ($B21 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B22 + 273,16)) + (0,0024804 * ($B22 + 273,16)) -(3142,31 / ($B22 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B23 + 273,16)) + (0,0024804 * ($B23 + 273,16)) -(3142,31 / ($B23 + 273,16)))*100000
=10 ^ (28,59051 - (8,2 * (LOG($B24 + 273,16)) + (0,0024804 * ($B24 + 273,16)) -(3142,31 / ($B24 + 273,16)))*100000

H	I	J	K
Actual VP at this T&RH (Pa	Abs H in (g/kg)	Saturated VP at this T (Pa)	Actual VP at this T&RH (Pa)
=(C13/100)*G13	=1000*0,622*(K13/(101325-K13))	2336,95932404742	=(E13/100)*J13
=(C14/100)*G14	=1000*0,622*(K14/(101325-K14))	2336,95932404742	=(E14/100)*J14
=(C15/100)*G15	=1000*0,622*(K15/(101325-K15))	2336,95932404742	=(E15/100)*J15
=(C16/100)*G16	=1000*0,622*(K16/(101325-K16))	2336,95932404742	=(E16/100)*J16
=(C17/100)*G17	=1000*0,622*(K17/(101325-K17))	2336,95932404742	=(E17/100)*J17
=(C18/100)*G18	=1000*0,622*(K18/(101325-K18))	2336,95932404742	=(E18/100)*J18
=(C19/100)*G19	=1000*0,622*(K19/(101325-K19))	2336,95932404742	=(E19/100)*J19
=(C20/100)*G20	=1000*0,622*(K20/(101325-K20))	2336,95932404742	=(E20/100)*J20
=(C21/100)*G21	=1000*0,622*(K21/(101325-K21))	2336,95932404742	=(E21/100)*J21
=(C22/100)*G22	=1000*0,622*(K22/(101325-K22))	2336,95932404742	=(E22/100)*J22
=(C23/100)*G23	=1000*0,622*(K23/(101325-K23))	2336,95932404742	=(E23/100)*J23
=(C24/100)*G24	=1000*0,622*(K24/(101325-K24))	2336,95932404742	=(E24/100)*J24

L	M
Energy of condensation / vaporization (in kWh/m3)	Ventilation Loss (kWh/m3)
=ABS(2257*(I13-F13)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D13-B13)*1,005*1,2*24*30/3600
=ABS(2257*(I14-F14)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D14-B14)*1,005*1,2*24*30/3600
=ABS(2257*(I15-F15)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D15-B15)*1,005*1,2*24*30/3600
=ABS(2257*(I16-F16)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D16-B16)*1,005*1,2*24*30/3600
=ABS(2257*(I17-F17)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D17-B17)*1,005*1,2*24*30/3600
=ABS(2257*(I18-F18)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D18-B18)*1,005*1,2*24*30/3600
=ABS(2257*(I19-F19)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D19-B19)*1,005*1,2*24*30/3600
=ABS(2257*(I20-F20)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D20-B20)*1,005*1,2*24*30/3600
=ABS(2257*(I21-F21)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D21-B21)*1,005*1,2*24*30/3600
=ABS(2257*(I22-F22)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D22-B22)*1,005*1,2*24*30/3600
=ABS(2257*(I23-F23)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D23-B23)*1,005*1,2*24*30/3600
=ABS(2257*(I24-F24)*$B$2*30*24/1000/1,2/3600)	=$B$2*(D24-B24)*1,005*1,2*24*30/3600

N	O
Conduction Loss (kWh/m3)	Total (kWh/m3)
=(D13-B13)*$B$3*$B$4*24*30/1000	=L13+ABS(M13)+ABS(N13)+$B$5
=(D14-B14)*$B$3*$B$4*24*30/1000	=L14+ABS(M14)+ABS(N14)+$B$5
=(D15-B15)*$B$3*$B$4*24*30/1000	=L15+ABS(M15)+ABS(N15)+$B$5
=(D16-B16)*$B$3*$B$4*24*30/1000	=L16+ABS(M16)+ABS(N16)+$B$5
=(D17-B17)*$B$3*$B$4*24*30/1000	=L17+ABS(M17)+ABS(N17)+$B$5
=(D18-B18)*$B$3*$B$4*24*30/1000	=L18+ABS(M18)+ABS(N18)+$B$5
=(D19-B19)*$B$3*$B$4*24*30/1000	=L19+ABS(M19)+ABS(N19)+$B$5
=(D20-B20)*$B$3*$B$4*24*30/1000	=L20+ABS(M20)+ABS(N20)+$B$5
=(D21-B21)*$B$3*$B$4*24*30/1000	=L21+ABS(M21)+ABS(N21)+$B$5
=(D22-B22)*$B$3*$B$4*24*30/1000	=L22+ABS(M22)+ABS(N22)+$B$5
=(D23-B23)*$B$3*$B$4*24*30/1000	=L23+ABS(M23)+ABS(N23)+$B$5
=(D24-B24)*$B$3*$B$4*24*30/1000	=L24+ABS(M24)+ABS(N24)+$B$5

Appendix D

APM-Cost model

Available at:

https://github.com/cristinadur/APMcostmodel
/blob/main/Life%20cycle%20collection%20management%20costs_04.xlsx

Appendix E

APM-Preservation model

Documentation created by the software AnyLogic 8.6.

Model: ChemicalPreservationModel01 - kopie

Name	Value
General	
Model time units	years
System Dynamics solver	
Differention Equations Method	Euler
Algebraic Equations Method	Modified Newton
Mixed Equations Method	RK45+Newton
Absolute accuracy	1.0E-5
Time accuracy	1.0E-5
Relative accuracy	1.0E-5
Fixed time step	0.001
Advanced	
Java package name	chemicalpreservationmodel01
File Name	C:\Users\crist\Dropbox\Cristina\UCL\PhD\AnyLogic\ChemicalPreservation\ChemicalPreservationModel01\ChemicalPreservationModel01 - kopie.alp

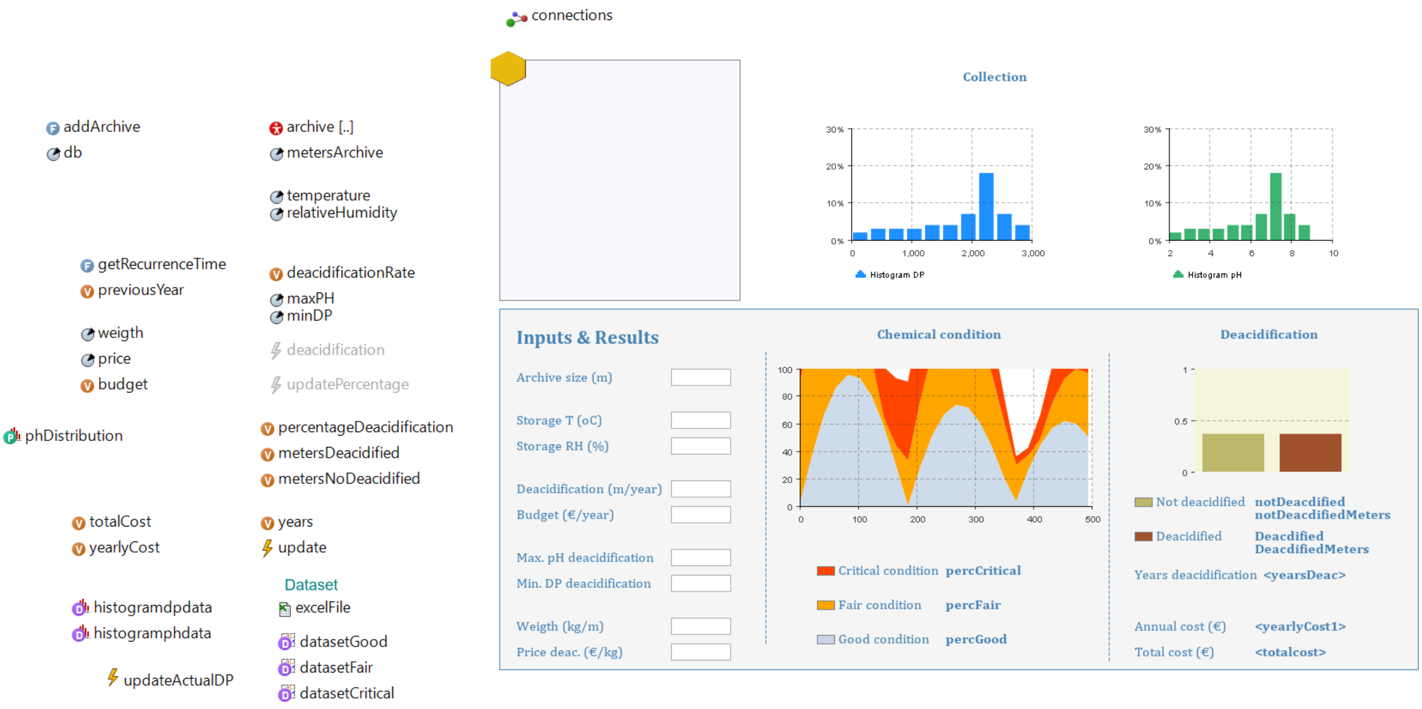
Agent Type: Main

Name	Value
Agent actions	
Startup code	addArchive(db); applyLayout();

Destroy code	excelFile.writeDataSet(datasetCritical,1,1,1);
	excelFile.writeDataSet(datasetFair,1,1,3);
	excelFile.writeDataSet(datasetGood,1,1,5);

Agent in flowcharts	
Use in flowcharts as	Agent
Dimensions and movement	
Speed	(10 : MPS)
Rotate animation towards movement	true
Rotate vertically as well (along Z-axis)	false
Space and network	
Space Type	Continuous
Dynamic: Width	260
Dynamic: Height	260
Dynamic: z Height	0
Layout Type	Arranged
Layout Type Apply On Startup	true
Network type	User-defined
Network Type Apply On Startup	true
Enable steps	false
Advanced Java	
Generic	false
Advanced	
Logging	true
Auto-create datasets	true
AOC_DATASETS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties

Name	Value
Limit the number of data samples	false



Parameter: temperature

Name	Value
General	
Array	false
Default value	18
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: relativeHumidity

Name	Value
General	
Array	false
Default value	50
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text

Name	Value
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: metersArchive

Name	Value
General	
Array	false
Default value	5000
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: maxPH

Name	Value
General	
Array	false
Default value	6
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: minDP

Name	Value
General	
Array	false
Default value	0
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: db

Name	Value
General	
Array	false
Default value	"acidic_collection_2000"
Type	String
Show at runtime	true
Show name	true
Value editor	
Label	Type of collection
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: weigth

Name	Value
General	
Array	false
Default value	32
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: price

Name	Value
General	
Array	false
Default value	45
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Function: getRecurrenceTime

Name	Value
General	

Name	Value
Return type	double
Return type:	Returns value
Show at runtime	true
Show name	true
Function body	
Body	Date nextDate = toDate((getYear()+1), getMonth(), getDayOfMonth(), getHourOfDay(), getMinute(), getSecond()); Date currentDate = date(); double difference = differenceInCalendarUnits(DAY, currentDate, nextDate); return difference;
Advanced	
Access type	default
System dynamics units	false

Function: addArchive

Name	Value
General	
Return type:	Just action (returns nothing)
Show at runtime	true
Show name	true
Function body	
Body	selectAndDoForEach(rs -> { add_archive(rs.getDouble("ph"), rs.getDouble("dp")); }, "SELECT * FROM " + db + ";");
Advanced	
Access type	default
System dynamics units	false

Arguments:

Name	Type
db	String

Custom Distribution: phDistribution

Name	Value
General	
Custom Distribution Definition Type	Ranges
Type	Continuous
Show at runtime	true
Show name	true
Data	
Load From Database	false

null	null	null
3.5	4.0	2.0
4.0	4.5	2.0
4.5	5.0	3.0
5.0	5.5	4.0
5.5	6.0	11.0
6.0	6.5	9.0
6.5	7.0	16.0
7.0	7.5	14.0
7.5	8.0	25.0
8.0	8.5	51.0
8.5	9.0	77.0
9.0	9.5	56.0
9.5	10.0	10.0
10.0	10.5	1.0
10.5	11.0	1.0

Event: update

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	int currentYear = getYear() +1;
	if(previousYear != (currentYear)){
	
	//update percentage deacidification//
	percentageDeacidification
	=(100*deacidificationRate)/metersArchive;
	
	//update deacidification action//
	List<InventoryNumber> acidicArchives = filter(archive, a -> a.ph <
	maxPH && a.dp > minDP);
	double n = (int)(archive.size()*percentageDeacidification)/100 >
	acidicArchives.size() ? acidicArchives.size() :
	(int)(archive.size()*percentageDeacidification)/100;
	for(int i = 0; i < n; i++){
	  InventoryNumber randonAcidicArchive =
	randomFrom(acidicArchives);
	   randonAcidicArchive.ph = uniform(7, 9.5);
	   // randonAcidicArchive.dp = randonAcidicArchive.actualDp;
	  //randonAcidicArchive.ph = phDistribution();
	  randonAcidicArchive.deacidified = true;
	  acidicArchives.remove(randonAcidicArchive);
	 }
	
	//update yearly cost deacidification//
	yearlyCost = weigth * price * deacidificationRate;
	
	//update years and cost deacidification//
	List<InventoryNumber> deacidifiedArchives = filter(archive, a ->
	a.ph < maxPH && a.dp > minDP && a.deacidified == false);

Name	Value

	if (deacidificationRate != 0 && deacidifiedArchives.size() > 0)
	{years = years + 1;
	totalCost = totalCost + yearlyCost;
	}
	
	//update meters deacidified//
	//metersNoDeacidified =
	((double)archive.NDeacidifiedNo()/archive.size())*metersArchive;
	//metersDeacidified =
	((double)archive.NDeacidifiedYes()/archive.size())*metersArchive;
	previousYear = currentYear;
	}

Event: updateActualDP

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	histogramdpdata.reset();
	for(InventoryNumber in: archive){
	    double H = pow(10, -in.ph);
	    double LnH = log(H);
	    double H20 = 100*pow((-log(1-relativeHumidity/100)/(285.655-1.67*temperature)),(1/(2.491-0.012*temperature)));
	    double LnK = 36.98+0.367*H20+0.244*LnH-14299.49/(temperature+273.15);
	    double K = exp(LnK);
	    double Years = (1/300.-1/in.dp)/K;
	    in.actualDp -= (in.dp-299)/Years;
	    in.onChange();
	    histogramdpdata.add(in.actualDp);
	}
	histogramphdata.reset();
	for(InventoryNumber in : archive){
	   histogramphdata.add(in.ph);
	}

Variable: percentageDeacidification

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: years

Name	Value
General	
Initial value	0
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: metersDeacidified

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: previousYear

Name	Value
General	
Initial value	-1
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: metersNoDeacidified

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true

Name	Value
System dynamics units	false

Variable: totalCost

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: yearlyCost

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: budget

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: deacidificationRate

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	

Name	Value
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Bar Chart: PopularDigital

Name	Value
General	
Chart Scale: To	1
Chart Scale: From	0
Scale type	Fixed
Lock	false
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Appearance	
Bars relative width	0.8
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	810.0
Width	220.0
y	340.0
Height	150.0
Legend	
Show legend	false
Chart area	
Chart Area: X Offset	30.0
Chart Area: Width	180.0
Chart Area: Y Offset	20.0
Chart Area: Height	120.0
Chart Area: Background Color	beige
Advanced	
Show name	false

Chart Items:

Title	Color	Value
not deacidified	darkKhaki	zidz(archive.NDeacidifiedNo(), archive.size())
deacidied	sienna	zidz(archive.NDeacidifiedYes(), archive.size())

Time Stack Chart: chartChemical

Name	Value
General	
Lock	false

Name	Value
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Dataset Samples To Keep	500
Scale	
Time window	500
Time	years
Vertical scale	Fixed
Chart Vertical Scale: To	100
Appearance	
Labels horizontal position	DEFAULT
Labels vertical position	DEFAULT
Label format	Model time units
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	340.0
Width	400.0
y	350.0
Height	200.0
Legend	
Show legend	false
Chart area	
Chart Area: X Offset	40.0
Chart Area: Width	340.0
Chart Area: Y Offset	10.0
Chart Area: Height	160.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Time window moves	Continuously
Show name	false
Logging	true

Plot Items:

Title	Type	Dataset / Value	Color
Good condition	value	archive.NGood()*100.0/archive.size()	new Color(176, 196, 222, 157)
Fair condition	value	archive.NFair()*100.0/archive.size()	orange
Critical condition	value	archive.NCritical()*100.0/archive.size()	orangeRed

Histogram: chart

Name	Value
General	
Show mean	false
Show CDF	false

Name	Value
Show PDF	true
Lock	false
Public	true
Data update	
Analysis auto update	false
Appearance	
Bars relative width	0.8
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	390.0
Width	290.0
y	50.0
Height	220.0
Legend	
Show legend	true
Legend size	30.0
Legend text color	black
Chart area	
Chart Area: X Offset	50.0
Chart Area: Width	210.0
Chart Area: Y Offset	30.0
Chart Area: Height	130.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Show name	false

Histogram Data Items:

Title	Dataset	PDF Color	CDF Color	Mean Color	Width	Low % Color	High % Color
Histogram DP	histogramdpdata	dodgerBlue	lightSlateBlue	darkKhaki	1	coral	yellowGreen

Histogram: chart1

Name	Value
General	
Show mean	false
Show CDF	false
Show PDF	true
Lock	false
Public	true
Data update	
Analysis auto update	false
Appearance	
Bars relative width	0.8
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray

Name	Value
Position and size	
x	760.0
Width	270.0
y	50.0
Height	220.0
Legend	
Show legend	true
Legend size	30.0
Legend text color	black
Chart area	
Chart Area: X Offset	50.0
Chart Area: Width	190.0
Chart Area: Y Offset	30.0
Chart Area: Height	130.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Show name	false

Histogram Data Items:

Title	Dataset	PDF Color	CDF Color	Mean Color	Width	Low % Color	High % Color
Histogram pH	histogramphdata	mediumSeaGreen	deepPink	darkKhaki	1	darkMagenta	violetRed

Data Set: datasetGood

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NGood()*100.0/archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetFair

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NFair()*100.0/archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true

Name	Value
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetCritical

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NCritical()*100.0/archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Histogram Data: histogramdpdata

Name	Value
General	
Logging	true
Calculate percentiles	false
Calculate CDF	true
Number of intervals	10
Show at runtime	true
Show name	true
Values range	
Data range	false
Range minimum	0
Range maximum	3000
Data update	
Analysis auto update	false

Histogram Data: histogramphdata

Name	Value
General	
Logging	true
Calculate percentiles	false
Calculate CDF	true
Number of intervals	10
Show at runtime	true
Show name	true
Values range	
Data range	false
Range minimum	2
Range maximum	9

Name	Value
Data update	
Analysis auto update	false

Edit Box: editbox1

Name	Value
General	
Enabled	true
Link	metersArchive
Link	metersArchive
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	360.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox2

Name	Value
General	
Enabled	true
Link	deacidificationRate
Link	deacidificationRate
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	490.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox3

Name	Value
General	

Name	Value
Enabled	true
Link	maxPH
Link	maxPH
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	570.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox4

Name	Value
General	
Enabled	true
Link	minDP
Link	minDP
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	600.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox5

Name	Value
General	
Enabled	true
Link	temperature
Link	temperature
Link to	true
Lock	false
Public	true
Appearance	

Name	Value
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	410.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox6

Name	Value
General	
Enabled	true
Link	relativeHumidity
Link	relativeHumidity
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	440.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox7

Name	Value
General	
Enabled	true
Link	weigth
Link	weigth
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	650.0
Height	20.0
Advanced	

Name	Value
Show name	false
Embedded icon	false

Edit Box: editbox8

Name	Value
General	
Enabled	true
Link	price
Link	price
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	680.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox9

Name	Value
General	
Enabled	true
Link	budget
Link	budget
Link to	true
Lock	false
Public	true
Action	
Action	deacidificationRate = budget/ (weigth * price);
Appearance	
Text color	steelBlue
Position and size	
x	230.0
Width	70.0
y	520.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Excel File: excelFile

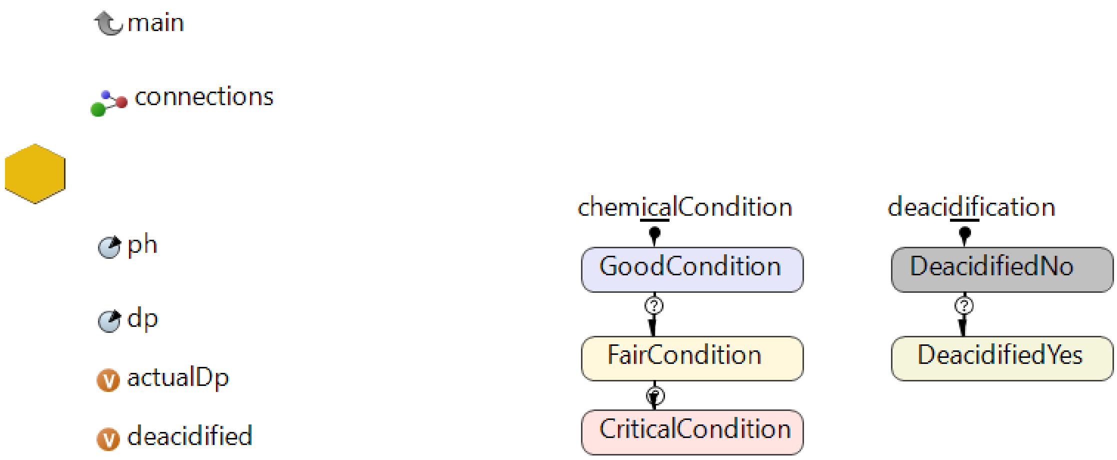
Name	Value
General	
File name	ResourceReference: chemicalpreservationmodel01.outputChemicalPreservation.xlsx
Show at runtime	true
Show name	true
Advanced	
Load on model startup	true
Save on model termination	true
Save in snapshot	false

Link to agents: connections

Name	Value
General	
Show at runtime	true
Show name	true
Communication	
Message type	Object
Animation	
Draw line	false

Agent Type: InventoryNumber

Name	Value
Agent in flowcharts	
Use in flowcharts as	Agent
Dimensions and movement	
Speed	(10 : MPS)
Rotate animation towards movement	true
Rotate vertically as well (along Z-axis)	false
Space and network	
Space Type	Continuous
Advanced Java	
Generic	false
Advanced	
Logging	true
Auto-create datasets	true
AOC_DATASETS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Limit the number of data samples	false



Parameter: ph

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: dp

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Variable: actualDp

Name	Value
General	
Initial value	dp
Type	double
Show at runtime	true
Show name	true

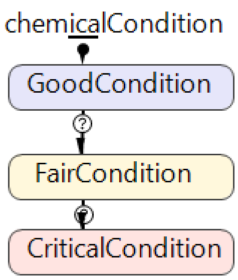
Name	Value
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: deacidified

Name	Value
General	
Initial value	false
Type	boolean
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Statechart Entry Point: chemicalCondition

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: transition

Name	Value
General	
Condition	actualDp < 800
Trigger type	Condition
Show name	false

Transition: transition1

Name	Value
General	

Name	Value
Condition	actualDp < 300
Trigger type	Condition
Show name	false

State: GoodCondition

Name	Value
General	
Entry action	shapeBox.setFillColor(lavender);
Fill color	lavender
Show name	true

State: FairCondition

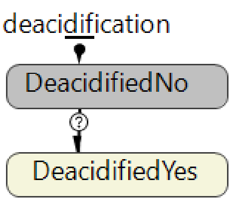
Name	Value
General	
Entry action	shapeBox.setFillColor(cornsilk);
Fill color	cornsilk
Show name	true

State: CriticalCondition

Name	Value
General	
Entry action	shapeBox.setFillColor(mistyRose);
Fill color	mistyRose
Show name	true

Statechart Entry Point: deacidification

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: transition2

Name	Value
General	
Condition	deacidified == true
Trigger type	Condition
Show name	false

State: DeacidifiedNo

Name	Value
General	
Entry action	shapeBox.setLineColor(midnightBlue);
Fill color	silver
Show name	true

State: DeacidifiedYes

Name	Value
General	
Entry action	shapeBox.setLineColor(red);
Fill color	beige
Show name	true

Link to agents: connections

Name	Value
General	
Show at runtime	true
Show name	true
Communication	
Message type	Object
Animation	
Draw line	false

Simulation Experiment: Simulation

Name	Value
General	
Bypass Initial Simulation Screen	false
Maximum available memory	512
Agent type	Main
Model time	
Execution mode	Virtual time (as fast as possible)
Stop option	Stop at specified time
Initial time	0.0
Final time	500.0
Initial date	Thu Oct 17 00:00:00 GMT 2019
Randomness	
Random Number Generation Type	Random seed (unique simulation runs)
Window	
Title	ChemicalPreservationModel01 : Simulation
Enable zoom and panning	true
Enable developer panel	true
Show developer panel on start	false
Advanced	
Load root from snapshot	false

ChemicalPreservationModel01

Variable: db

Name	Value
General	
Initial value	"acidic_collection_2000"
Type	String
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Radio Buttons: radio

Name	Value
General	
Enabled	true
Link to	false
String values	acidic_collection, modern_collection, mixed_collection, acidic_library, acidic_RoyalLibrary, niod, utrechtsarchief
Orientation	Vertical
Lock	false
Action	

Action	if (radio.getValue() == 0)
	db = "acidic_collection_2000";
	if (radio.getValue() == 1)
	db = "modern_collection_2000";
	if (radio.getValue() == 2)
	db = "mixed_collection_2000";
	if (radio.getValue() == 3)
	db = "acidic_library_2000";
	if (radio.getValue() == 4)
	db = "acidic_royallibrary_2000";
	if (radio.getValue() == 5)
	db = "niod";
	if (radio.getValue() == 6)
	db = "utrechtsarchief";

Position and size	

Name	Value
x	160.0
Width	150.0
y	140.0
Height	200.0
Advanced	
Show name	false

Database: Database

Name	Value
General	
Database shutdown compacr	false
Log	
Logging	false

Database Table: acidic_collection_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: mixed_collection_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: modern_collection_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: acidic_library_2000_res

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: acidic_library_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: acidic_royallibrary_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: niod

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: utrechtsarchief

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	

null	null	null	null	null
ph	DOUBLE	-	-	

Appendix F

APM-Access model in system dynamics

Screenshots showing the Excel formulae and layout of the model.

A	B	C	D	E	F	G
Year	**requestsRR (Rr)**	**decrease Rr**	**requestsSoD (Rs)**	**requestsBlock (Rb)**	**matchSoD (Ms)**	**matchBlock (Mb)**
2005	24932	=B2*(R2+S2)	0	0	1	1
2006	=B2-C3	=B2*(R3+S3)	0	2747	1	0,225700764470331
2007	=B3-C4	=B3*(R4+S4)	288	3801	0,524305555555556	0,312812417784793
2008	=B4-C5	=B4*(R5+S5)	2793	1359	0,42141066953097	0,36644591611479
2009	=B5-C6	=B5*(R6+S6)	3371	694	0,427469593592406	0,39193083573487
2010	=B6-C7	=B6*(R7+S7)	4251	9154	0,36485532815808	0,222744155560411
2011	=B7-C8	=B7*(R8+S8)	3873	9513	0,394009811515621	0,101755492483969
2012	=B8-C9	=B8*(R9+S9)	3794	2646	0,423036373220875	0,215041572184429
2013	=B9-C10	=B9*(R10+S10)	3860	5121	0,421502590673575	0,128099980472564
2014	=B10-C11	=B10*(R11+S11)	3018	2682	0,450629555997349	0,397837434750186
2015	=B11-C12	=B11*(R12+S12)	3739	1843	0,442364268520995	0,285946825827455
2016	=B12-C13	=B12*(R13+S13)	4036	3245	0,441030723488603	0,416103257529195
2017	=B13-C14	=B13*(R14+S14)	4994	4973	0,4221065278334	0,542127488437563
2018	=B14-C15	=B14*(R15+S15)	6023	12005	0,416901876141458	0,471137026239067

H	I	J	K	L	M	N
increase Cad	**increase Cnd**	**collection**	**proportion Ca**	**newRequested (P)**	**increase Ca**	**requested (Ca)**
=(D2*F2)+(E2*G2)	=D2*(1-F2)+E2*(1-G2)	1000000	0,1	0	=B2*L2	=J2*K2
=(D3*F3)+(E3*G3)	=D3*(1-F3)+E3*(1-G3)	=O3+P3+Q3+N3		0,2	=B3*L3	=N2+M3-H3
=(D4*F4)+(E4*G4)	=D4*(1-F4)+E4*(1-G4)	=O4+P4+Q4+N4		0,2	=B4*L4	=N3+M4-H4
=(D5*F5)+(E5*G5)	=D5*(1-F5)+E5*(1-G5)	=O5+P5+Q5+N5		0,19	=B5*L5	=N4+M5-H5
=(D6*F6)+(E6*G6)	=D6*(1-F6)+E6*(1-G6)	=O6+P6+Q6+N6		0,19	=B6*L6	=N5+M6-H6
=(D7*F7)+(E7*G7)	=D7*(1-F7)+E7*(1-G7)	=O7+P7+Q7+N7		0,17	=B7*L7	=N6+M7-H7
=(D8*F8)+(E8*G8)	=D8*(1-F8)+E8*(1-G8)	=O8+P8+Q8+N8		0,2	=B8*L8	=N7+M8-H8
=(D9*F9)+(E9*G9)	=D9*(1-F9)+E9*(1-G9)	=O9+P9+Q9+N9		0,18	=B9*L9	=N8+M9-H9
=(D10*F10)+(E10*G10)	=D10*(1-F10)+E10*(1-G10)	=O10+P10+Q10+N10		0,2	=B10*L10	=N9+M10-H10
=(D11*F11)+(E11*G11)	=D11*(1-F11)+E11*(1-G11)	=O11+P11+Q11+N11		0,25	=B11*L11	=N10+M11-H11
=(D12*F12)+(E12*G12)	=D12*(1-F12)+E12*(1-G12)	=O12+P12+Q12+N12		0,26	=B12*L12	=N11+M12-H12
=(D13*F13)+(E13*G13)	=D13*(1-F13)+E13*(1-G13)	=O13+P13+Q13+N13		0,26	=B13*L13	=N12+M13-H13
=(D14*F14)+(E14*G14)	=D14*(1-F14)+E14*(1-G14)	=O14+P14+Q14+N14		0,3	=B14*L14	=N13+M14-H14
=(D15*F15)+(E15*G15)	=D15*(1-F15)+E15*(1-G15)	=O15+P15+Q15+N15		0,37	=B15*L15	=N14+M15-H15

O	P	Q	R	S	T
neverRequested (Cn)	**requestedDigi (Cad)**	**neverRequestedDigi (Cnd)**	**matchRequested (Ma)**	**matchNonRequested (Mn)**	**weightMn (W)**
=J2-N2	0	0	=H2/(N2+P2)	=T2*(Q2/(O2+Q2))	0,3
=O2-I3-M3	=P2+H3	=Q2+I3	=H3/(N3+P3)	=T3*(Q3/(O3+Q3))	0,3
=O3-I4-M4	=P3+H4	=Q3+I4	=H4/(N4+P4)	=T4*(Q4/(O4+Q4))	0,3
=O4-I5-M5	=P4+H5	=Q4+I5	=H5/(N5+P5)	=T5*(Q5/(O5+Q5))	0,3
=O5-I6-M6	=P5+H6	=Q5+I6	=H6/(N6+P6)	=T6*(Q6/(O6+Q6))	0,3
=O6-I7-M7	=P6+H7	=Q6+I7	=H7/(N7+P7)	=T7*(Q7/(O7+Q7))	0,3
=O7-I8-M8	=P7+H8	=Q7+I8	=H8/(N8+P8)	=T8*(Q8/(O8+Q8))	0,3
=O8-I9-M9	=P8+H9	=Q8+I9	=H9/(N9+P9)	=T9*(Q9/(O9+Q9))	0,3
=O9-I10-M10	=P9+H10	=Q9+I10	=H10/(N10+P10)	=T10*(Q10/(O10+Q10))	0,3
=O10-I11-M11	=P10+H11	=Q10+I11	=H11/(N11+P11)	=T11*(Q11/(O11+Q11))	0,3
=O11-I12-M12	=P11+H12	=Q11+I12	=H12/(N12+P12)	=T12*(Q12/(O12+Q12))	0,3
=O12-I13-M13	=P12+H13	=Q12+I13	=H13/(N13+P13)	=T13*(Q13/(O13+Q13))	0,3
=O13-I14-M14	=P13+H14	=Q13+I14	=H14/(N14+P14)	=T14*(Q14/(O14+Q14))	0,3
=O14-I15-M15	=P14+H15	=Q14+I15	=H15/(N15+P15)	=T15*(Q15/(O15+Q15))	0,3

Appendix G

APM model

Documentation created by the software AnyLogic 8.6.

Model: ReadingRoom4MoreAgentsPreser

Name	Value
General	
Model time units	years
System Dynamics solver	
Differention Equations Method	Euler
Algebraic Equations Method	Modified Newton
Mixed Equations Method	RK45+Newton
Absolute accuracy	1.0E-5
Time accuracy	1.0E-5
Relative accuracy	1.0E-5
Fixed time step	0.001
Advanced	
Java package name	readingroom
File Name	C:\Users\crist\Dropbox\Cristina\UCL\PhD\AnyLogic\ReadingRoom4MoreAgents.02 - kopie\ReadingRoom4MoreAgentsPreser.alp

Agent Type: Main

Description: In Main the number of popular agents is set. In InventoryNumber the number of requested once as well teh number of requests for the popular once is chosen.

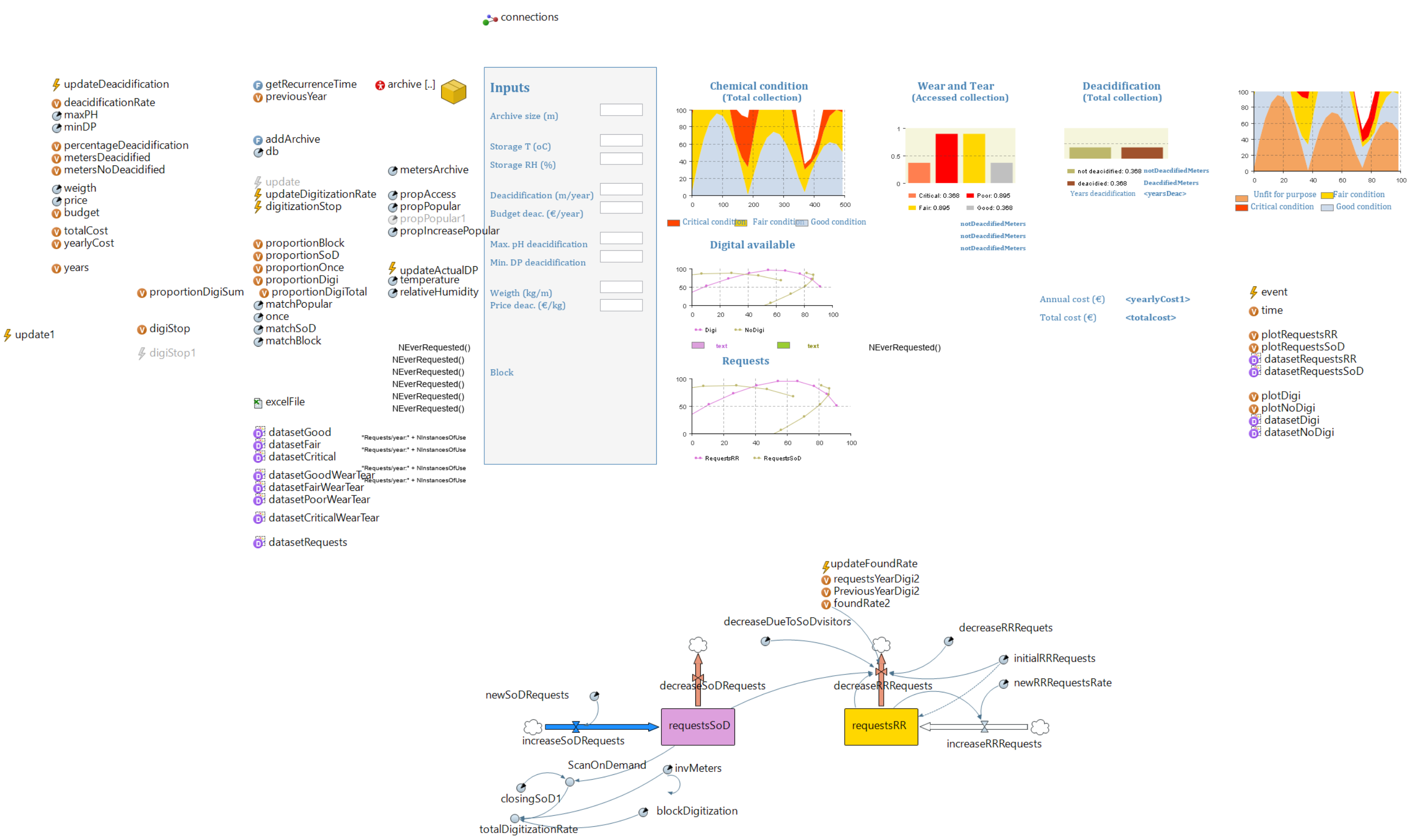
Name	Value
Agent actions	

Startup code	addArchive(((Simulation)getExperiment()).db);
	applyLayout();
	
	//On startup number of popular archives is chosen//
	List<InventoryNumber> tmpArchives = new ArrayList<InventoryNumber>();
	for(InventoryNumber a: archive)
	tmpArchives.add(a);
	double n = archive.size()* propAccess*propPopular;
	for(int i = 0; i < n; i++){
	InventoryNumber randomArchive = tmpArchives.remove(uniform_discr(tmpArchives.size() - 1));
	randomArchive.popular = true;
	randomArchive.requestsTotal = (int) exponential(0.3, 1);
	//randomArchive.requestsTotal = (int) 20000;
	randomArchive.everRequested = true;
	}
	List<InventoryNumber> requestedArchive = filter(archive, a -> a.popular != true);
	double m = archive.size()* 0.035;
	for(int i = 0; i < m; i++){
	InventoryNumber randomRequestedArchive = randomFrom(requestedArchive);
	randomRequestedArchive.requestsTotal = (int) 1;
	requestedArchive.remove(randomRequestedArchive);
	}

Destroy code	excelFile.writeDataSet(datasetRequests,1,1,1);
	excelFile.writeDataSet(datasetCritical,1,1,3);
	excelFile.writeDataSet(datasetFair,1,1,5);
	excelFile.writeDataSet(datasetCriticalWearTear,1,1,7);
	excelFile.writeDataSet(datasetFairWearTear,1,1,9);

Agent in flowcharts	
Use in flowcharts as	Agent

Name	Value
Dimensions and movement	
Speed	(10 : MPS)
Rotate animation towards movement	true
Rotate vertically as well (along Z-axis)	false
Space and network	
Space Type	Continuous
Dynamic: Width	500
Dynamic: Height	500
Dynamic: z Height	0
Layout Type	User-defined
Layout Type Apply On Startup	true
Network type	User-defined
Network Type Apply On Startup	true
Enable steps	false
Advanced Java	
Generic	false
Advanced	
Logging	true
Auto-create datasets	true
AOC_DATASETS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Limit the number of data samples	false
Description	
Description	In Main the number of popular agents is set. In InventoryNumber the number of requested once as well teh number of requests for the popular once is chosen.



Parameter: propAccess

Description: Percentage of archives accessed in 20 years

Name	Value
General	
Array	false
Default value	0.15
Type	double
Show at runtime	true
Show name	true
Value editor	
Label	Accessed archives (%)
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true
Description	
Description	Percentage of archives accessed in 20 years

Parameter: propPopular

Description: Percentage of archives that are popular (requested more than the once)

Name	Value
General	
Array	false
Default value	0.3
Type	double
Show at runtime	true
Show name	true
Value editor	
Label	Popular archives (%)
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true
Description	
Description	Percentage of archives that are popular (requested more than the once)

Parameter: invMeters

Name	Value
General	
Array	false
Default value	25
Type	double
Show at runtime	true
Show name	true
Value editor	

Name	Value
Label	InvNr per meter (invNr/m)
Editor control	Text
Advanced	
System dynamics units	true
Unit	invnummer/m
Save in snapshot	true

Parameter: blockDigitization

Name	Value
General	
Array	false
Default value	0
Type	double
Show at runtime	true
Show name	true
Value editor	
Label	Digitization rate (m/year)
Editor control	Text
Advanced	
System dynamics units	true
Unit	m/time
Save in snapshot	true

Parameter: newSoDRequests

Name	Value
General	
Array	false
Default value	0
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: closingSoD1

Name	Value
General	
Array	false
Default value	1
Type	double
Show at runtime	true
Show name	true
Value editor	

Name	Value
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: newRRRequestsRate

Name	Value
General	
Array	false
Default value	(0 : PER_YEAR)
Unit	per year
Show at runtime	true
Show name	true
Value editor	
Editor control	Unit editor
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: decreaseRRRequets

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: decreaseDueToSoDvisitors

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: initialRRRequests

Name	Value
General	
Array	false
Default value	24782
Type	double
Show at runtime	true
Show name	true
Value editor	
Label	Nr requests rr
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: matchSoD

Name	Value
General	
Array	false
Default value	0.42
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: matchBlock

Name	Value
General	
Array	false
Default value	0.33
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: metersArchive

Name	Value
General	
Array	false
Default value	40000

Name	Value
Type	double
Show at runtime	true
Show name	true
Value editor	
Label	Total meter archive (m)
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: matchPopular

Name	Value
General	
Array	false
Default value	1
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: propIncreasePopular

Name	Value
General	
Array	false
Default value	0.002
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: once

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	

Name	Value
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: db

Name	Value
General	
Array	false
Default value	"acidic_collection_2000"
Type	String
Show at runtime	true
Show name	true
Value editor	
Label	Type of collection
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: temperature

Name	Value
General	
Array	false
Default value	18
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: relativeHumidity

Name	Value
General	
Array	false
Default value	50
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false

Name	Value
Save in snapshot	true

Parameter: maxPH

Name	Value
General	
Array	false
Default value	6
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: minDP

Name	Value
General	
Array	false
Default value	0
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: weigth

Name	Value
General	
Array	false
Default value	32
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: price

Name	Value
General	
Array	false
Default value	45
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Function: getRecurrenceTime

Name	Value
General	
Return type	double
Return type:	Returns value
Show at runtime	true
Show name	true
Function body	
Body	Date nextDate = toDate((getYear()+1), getMonth(), getDayOfMonth(), getHourOfDay(), getMinute(), getSecond()); Date currentDate = date(); double difference = differenceInCalendarUnits(DAY, currentDate, nextDate); return difference;
Advanced	
Access type	default
System dynamics units	false

Function: addArchive

Name	Value
General	
Return type:	Just action (returns nothing)
Show at runtime	true
Show name	true
Function body	
Body	selectAndDoForEach(rs -> { add_archive(rs.getDouble("ph"), rs.getDouble("dp")); }, "SELECT * FROM " + db + ";");
Advanced	
Access type	default
System dynamics units	false

Arguments:

Name	Type
db	String

Event: updateFoundRate

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	int currentYear = getYear() +1;
	if(previousYear != (currentYear)){
	
	//variable = archive.NPopularEverDigital();
	requestsYearDigi2 = archive.NRequestsTotalDigital()-
	PreviousYearDigi2;
	PreviousYearDigi2 = archive.NRequestsTotalDigital();
	foundRate2 = requestsYearDigi2/archive.NRequestsTotal();
	
	previousYear = currentYear;
	}

Event: updateDigitizationRate

Name	Value
General	
Logging	true
Condition	archive.NDigitalAvailable()*100.0/archive.size() == 100
Trigger type	Condition
Show at runtime	true
Show name	true
Action	
Action	blockDigitization = 0;

Event: event

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	time = time();
	plotRequestsSoD = requestsSoD + blockDigitization;
	plotRequestsRR = requestsRR;

Name	Value

	datasetRequestsRR.add(time, plotRequestsRR);
	datasetRequestsSoD.add(time, plotRequestsSoD);
	
	//plotDigi = archive.NInUse()*100.0/archive.size();
	//plotNoDigi = archive.NDigitalAvailable()*100.0/archive.size() ;
	plotDigi = proportionDigiTotal*100;
	plotNoDigi = (1-proportionDigiTotal)*100;
	
	datasetDigi.add(time, plotDigi);
	datasetNoDigi.add(time, plotNoDigi);
	
	if(time() < 100){
	plotRequests.updateData();
	plotDigitalAvailable1.updateData();
	
	}

Event: digitizationStop

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Occurs once
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	
Action	update1.reset();

Event: updateActualDP

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	
	//histogramdpdata.reset();
	for(InventoryNumber in: archive){
	    double H = pow(10, -in.ph);
	    double LnH = log(H);
	    double H20 = 100*pow((-log(1-relativeHumidity/100)/(285.655-1.67*temperature)),(1/(2.491-0.012*temperature)));
	    double LnK = 36.98+0.367*H20+0.244*LnH-14299.49/(temperature+273.15);
	    double K = exp(LnK);
	    double Years = (1/300.-1/in.dp)/K;
	    in.actualDp -= (in.dp-299)/Years;
	    in.onChange();
	    //histogramdpdata.add(in.actualDp);
	}
	//histogramphdata.reset();
	//for(InventoryNumber in : archive){

Name	Value
	//histogramphdata.add(in.ph); //}

Event: updateDeacidification

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	//int currentYear = getYear() +1;
	//if(previousYear != (currentYear))
	int currentYear = getYear() +1;
	
	//{
	
	//update percentage deacidification//
	percentageDeacidification
	=(100*deacidificationRate)/metersArchive;
	
	//update deacidification action//
	List<InventoryNumber> acidicArchives = filter(archive, a -> a.ph < maxPH && a.dp > minDP);
	double n = (int)(archive.size()*percentageDeacidification)/100 > acidicArchives.size() ? acidicArchives.size() :
	(int)(archive.size()*percentageDeacidification)/100;
	for(int i = 0; i < n; i++){
	  InventoryNumber randonAcidicArchive =
	randomFrom(acidicArchives);
	  randonAcidicArchive.ph = uniform(7, 9.5);
	  randonAcidicArchive.deacidified = true;
	  acidicArchives.remove(randonAcidicArchive);
	 }
	
	//update yearly cost deacidification//
	yearlyCost = weigth * price * deacidificationRate;
	
	//update years and cost deacidification//
	List<InventoryNumber> deacidifiedArchives = filter(archive, a ->
	a.ph < maxPH && a.dp > minDP && a.deacidified == false);
	if (deacidificationRate != 0 && deacidifiedArchives.size() > 0)
	{years = years + 1;
	totalCost = totalCost + yearlyCost;
	}
	
	//update meters deacidified//
	//metersNoDeacidified =
	((double)archive.NDeacidifiedNo()/archive.size())*metersArchive;
	//metersDeacidified =
	((double)archive.NDeacidifiedYes()/archive.size())*metersArchive;
	
	//previousYear = currentYear;
	//}

Event: update1

Name	Value

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	int currentYear = getYear() +1;
	
	blockDigitization= (double) selectFrom(digitization)
	.where(digitization.db_year.eq(currentYear))
	.firstResult(false, digitization.db_blockdigitization);
	
	requestsSoD= (double) selectFrom(digitization)
	.where(digitization.db_year.eq(currentYear))
	.firstResult(false, digitization.db_sod);
	
	matchSoD= (double) selectFrom(digitization)
	.where(digitization.db_year.eq(currentYear))
	.firstResult(false, digitization.db_matchsod);
	
	matchBlock= (double) selectFrom(digitization)
	.where(digitization.db_year.eq(currentYear))
	.firstResult(false, digitization.db_matchblock);
	
	once= (double) selectFrom(digitization)
	.where(digitization.db_year.eq(currentYear))
	.firstResult(false, digitization.db_once);
	
	//Add new popular inventoryNumbers//
	List<InventoryNumber> nonPopularArchive = filter(archive, a ->
	a.popular == false && a.inState(InventoryNumber.InUse));
	double m = archive.size()* propIncreasePopular;
	for(int i = 0; i < m; i++){
	  InventoryNumber randomNonPopularArchive =
	randomFrom(nonPopularArchive);
	  randomNonPopularArchive.popular = true;
	  randomNonPopularArchive.requestsTotal = (int) exponential(0.3,
	1);
	  //randomNonPopularArchive.requestsTotal = 20000;
	  nonPopularArchive.remove(randomNonPopularArchive);
	}
	
	if (digiStop == 1)
	{
	blockDigitization =0;
	requestsSoD = 0;
	}
	
	//Update percentages//
	proportionBlock =
	(double)(100*(blockDigitization/invMeters))/metersArchive;
	proportionSoD =
	(double)(100*requestsSoD/invMeters)/metersArchive;
	//proportionDigi= proportionBlock + proportionSoD;
	//proportionSoD = ScanOnDemand*matchSoD;
	//proportionBlock = blockDigitization*matchBlock;
	//proportionDigi = (proportionSoD +
	proportionBlock)/metersArchive;
	proportionDigi= (blockDigitization + requestsSoD)/invMeters;
	proportionDigiSum += proportionDigi;
	proportionDigiTotal = proportionDigiSum/metersArchive;
	proportionOnce = proportionDigi * once;
	
	if (archive.NEverRequested() != archive.NEverRequestedDigi())
	{
	//Upadate digi once
	List<InventoryNumber> nonDigitizedArchives = filter(archive, a ->
	a.inState(InventoryNumber.InUse) && a.popular == false && a.

Name	Value

	requestsTotal == 1);
	double n = (int) (archive.size()*proportionBlock)/100 > nonDigitizedArchives.size() ? nonDigitizedArchives.size() :
	(int)
	(archive.size()*((proportionBlock*matchBlock*once)+(proportionSoD*matchSoD*once)))/100;
	for(int i = 0; i < n; i++){
	  InventoryNumber randonNonDigitizedArchive = randomFrom(nonDigitizedArchives);
	   randonNonDigitizedArchive.digitized = true;
	  nonDigitizedArchives.remove(randonNonDigitizedArchive);
	}
	
	//Update digi popular
	List<InventoryNumber> nonDigitizedArchives2 = filter(archive, a -> a.inState(InventoryNumber.InUse) && a.popular == true && a.requestsTotal >= 2);
	double p = (int) (archive.size()*proportionSoD)/100 > nonDigitizedArchives2.size() ? nonDigitizedArchives2.size() :
	(int)(archive.size()*((proportionBlock*matchBlock*(1-once))+(proportionSoD*matchSoD*(1-once))))/100;
	for(int i = 0; i< p; i++){
	   InventoryNumber randonNonDigitizedArchive2 = randomFrom(nonDigitizedArchives2);
	  randonNonDigitizedArchive2.digitized = true;
	   nonDigitizedArchives2.remove(randonNonDigitizedArchive2);
	}
	//Update digi non accessed
	List<InventoryNumber> nonDigitizedArchives4 = filter(archive, a -> a.inState(InventoryNumber.InUse) && a.popular == false && a.requestsTotal == 0);
	double a = (int) (archive.size()*(proportionSoD + proportionBlock))/100 > nonDigitizedArchives4.size() ? nonDigitizedArchives4.size() :
	(int)(archive.size()*((proportionBlock*(1-matchBlock))+(proportionSoD*(1-matchSoD))))/100;
	for(int i = 0; i< a; i++){
	   InventoryNumber randonNonDigitizedArchive4 = randomFrom(nonDigitizedArchives4);
	  randonNonDigitizedArchive4.digitized = true;
	  nonDigitizedArchives4.remove(randonNonDigitizedArchive4);
	}
	}
	else
	{
	//Update digi non accessed
	List<InventoryNumber> nonDigitizedArchives5 = filter(archive, a -> a.inState(InventoryNumber.InUse) && a.popular == false && a.requestsTotal == 0);
	double b = (int) (archive.size()*(proportionSoD + proportionBlock))/100 > nonDigitizedArchives5.size() ? nonDigitizedArchives5.size() :
	(int)(archive.size()*(proportionBlock + proportionSoD))/100;
	for(int i = 0; i< b; i++){
	   InventoryNumber randonNonDigitizedArchive5 = randomFrom(nonDigitizedArchives5);
	  randonNonDigitizedArchive5.digitized = true;
	 nonDigitizedArchives5.remove(randonNonDigitizedArchive5);

Variable: foundRate2

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public

Name	Value
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: requestsYearDigi2

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: PreviousYearDigi2

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionSoD

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionBlock

Name	Value
General	
Type	double
Show at runtime	true

Name	Value
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: plotRequestsSoD

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: time

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: plotRequestsRR

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: previousYear

Name	Value

Name	Value
General	
Initial value	-1
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: years

Name	Value
General	
Initial value	0
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionDigi

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionOnce

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true

Name	Value
System dynamics units	false

Variable: percentageDeacidification

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: metersDeacidified

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: metersNoDeacidified

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: totalCost

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	

Name	Value
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: yearlyCost

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: budget

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: deacidificationRate

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: plotDigi

Name	Value
General	
Type	double

Name	Value
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: plotNoDigi

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionDigiTotal

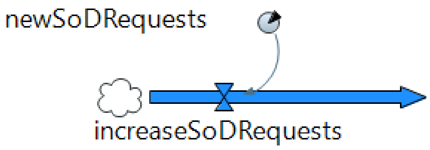
Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: proportionDigiSum

Name	Value
General	
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

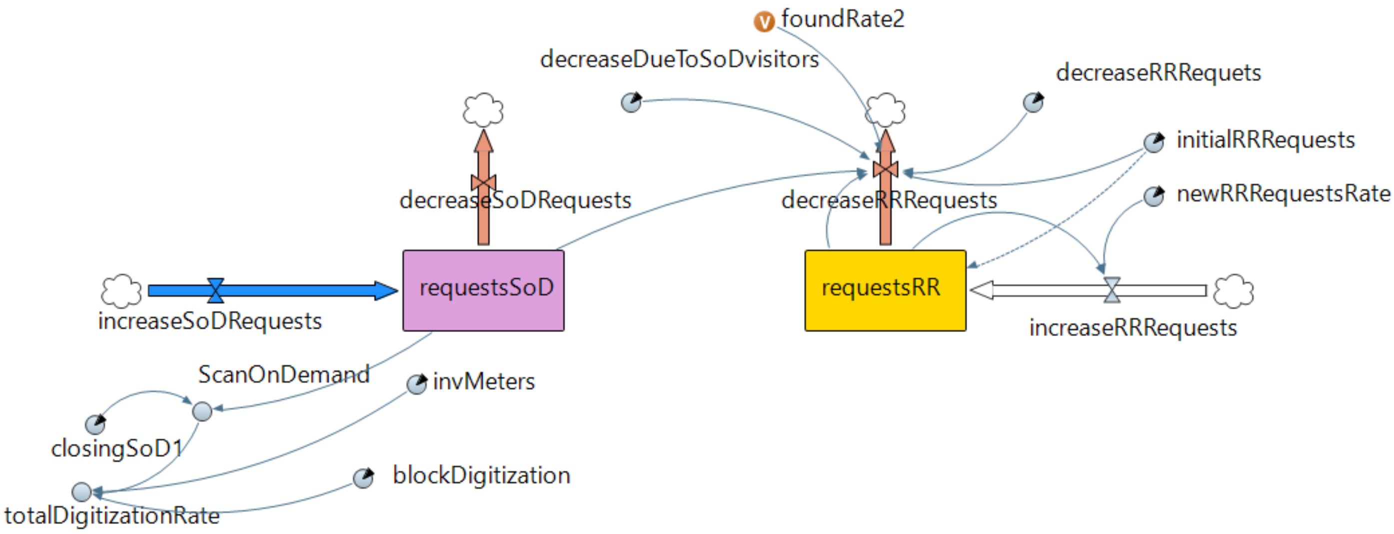
Variable: digiStop

Name	Value
General	
Initial value	0
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false



Flow: increaseSoDRequests

Name	Value
General	
Formula	newSoDRequests
Constant	false
External	false
Array	false
Color	dodgerBlue
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false



Stock: requestsSoD

Name	Value
General	
Equation mode	Classic
Initial value	0
Array	false
Color	plum
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false

Dynamic Variable: ScanOnDemand

Description: 20 average number of inventory numbers/meter

Name	Value
General	
Formula	requestsSoD > 0 ? requestsSoD * closingSoD1 : 0
Constant	false
External	false
Array	false
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	true
Unit	m/time
Description	
Description	20 average number of inventory numbers/meter

Dynamic Variable: totalDigitizationRate

Name	Value
General	
Formula	(ScanOnDemand + blockDigitization)/ + invMeters
Constant	false
External	false
Array	false
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	true
Unit	m/time

Flow: decreaseRRRequests

Name	Value
General	
Formula	requestsRR > 0 ? ((initialRRRequests * foundRate2) +

Name	Value
	decreaseRRRequets + (requestsSoD*decreaseDueToSoDvisitors)): 0
Constant	false
External	false
Array	false
Color	darkSalmon
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	true
Unit	invnummer/time

Stock: requestsRR

Name	Value
General	
Equation mode	Classic
Initial value	initialRRRequests
Array	false
Color	gold
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false

Flow: increaseRRRequests

Name	Value
General	
Formula	newRRRequestsRate * requestsRR
Constant	false
External	false
Array	false
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false

Flow: decreaseSoDRequests

Name	Value
General	
Formula	0
Constant	false
External	false
Array	false
Color	darkSalmon

Name	Value
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false

Flow: increaseSoDRequests

Name	Value
General	
Formula	newSoDRequests
Constant	false
External	false
Array	false
Color	dodgerBlue
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	false

Flow: decreaseRRRequests

Name	Value
General	
Formula	requestsRR > 0 ? ((initialRRRequests * foundRate2) + decreaseRRRequets + (requestsSoD*decreaseDueToSoDvisitors)): 0
Constant	false
External	false
Array	false
Color	darkSalmon
Show at runtime	true
Public	false
Show name	true
Advanced	
System dynamics units	true
Unit	invnummer/time

Flow: increaseRRRequests

Name	Value
General	
Formula	newRRRequestsRate * requestsRR
Constant	false
External	false
Array	false
Show at runtime	true
Public	false
Show name	true

Name	Value
Advanced	
System dynamics units	false

Bar Chart: WearTear

Name	Value
General	
Chart Scale: To	1
Chart Scale: From	0
Scale type	Fixed
Lock	false
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Appearance	
Bars relative width	0.8
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	670.0
Width	220.0
y	70.0
Height	180.0
Legend	
Show legend	true
Legend size	60.0
Legend text color	black
Chart area	
Chart Area: X Offset	30.0
Chart Area: Width	180.0
Chart Area: Y Offset	20.0
Chart Area: Height	90.0
Chart Area: Background Color	beige
Advanced	
Show name	false

Chart Items:

Title	Color	Value
Critical	coral	zidz(archive.NCriticalWearTear(), archive.NEverRequested())
Poor	red	zidz(archive.NPoorWearTear(),archive.NEverRequested())
Fair	gold	zidz(archive.NFairWearTear(), archive.

Chart Items:

Title	Color	Value
		NEverRequested())
Good	silver	zidz(archive.NGoodWearTear(), archive.NEverRequested())

Bar Chart: PopularDigital1

Name	Value
General	
Chart Scale: To	1
Chart Scale: From	0
Scale type	Fixed
Lock	false
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Appearance	
Bars relative width	0.8
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	960.0
Width	180.0
y	70.0
Height	140.0
Legend	
Show legend	true
Legend size	60.0
Legend text color	black
Chart area	
Chart Area: X Offset	0.0
Chart Area: Width	170.0
Chart Area: Y Offset	20.0
Chart Area: Height	50.0
Chart Area: Background Color	beige
Advanced	
Show name	false

Chart Items:

Title	Color	Value
not deacidified	darkKhaki	zidz(archive.NDeacidifiedNo(), archive.size())
deacidied	sienna	zidz(archive.NDeacidifiedYes(), archive.size())

Plot: plotRequests

Name	Value
General	

Name	Value
Lock	false
Public	true
Data update	
Analysis auto update	false
Dataset Samples To Keep	100
Scale	
Horizontal scale	Fixed
Chart Horizontal Scale: From	0
Chart Horizontal Scale: To	100
Vertical scale	Fixed
Chart Vertical Scale: From	0
Chart Vertical Scale: To	initialRRRequests
Appearance	
Labels horizontal position	DEFAULT
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Draw line	true
Interpolation	Linear
Position and size	
x	300.0
Width	340.0
y	470.0
Height	180.0
Legend	
Show legend	true
Legend size	30.0
Legend text color	black
Chart area	
Chart Area: X Offset	50.0
Chart Area: Width	260.0
Chart Area: Y Offset	30.0
Chart Area: Height	90.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Show name	false
Logging	true

Plot Items:

Title	Type	Dataset		Point Style	Color	Line	Width	Interpolation
		X Axis Value	Y Axis Value					
RequestsRR	value	time	plotRequests RR	CIRCLE	orchid	true	1.0	LINEAR
RequestsSoD	value	time	plotRequests SoD	CIRCLE	darkKhaki	true	1.0	LINEAR

Plot: plotDigitalAvailable1

Name	Value

Name	Value
General	
Lock	false
Public	true
Data update	
Analysis auto update	false
Dataset Samples To Keep	100
Scale	
Horizontal scale	Fixed
Chart Horizontal Scale: From	0
Chart Horizontal Scale: To	100
Vertical scale	Fixed
Chart Vertical Scale: From	0
Chart Vertical Scale: To	100
Appearance	
Labels horizontal position	DEFAULT
Labels vertical position	DEFAULT
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Draw line	true
Interpolation	Linear
Position and size	
x	300.0
Width	310.0
y	290.0
Height	150.0
Legend	
Show legend	true
Legend size	30.0
Legend text color	black
Chart area	
Chart Area: X Offset	50.0
Chart Area: Width	230.0
Chart Area: Y Offset	30.0
Chart Area: Height	60.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Show name	false
Logging	true

Plot Items:

Title	Type	Dataset		Point Style	Color	Line	Width	Interpolation
		X Axis Value	Y Axis Value					
Digi	value	time	plotDigi	CIRCLE	orchid	true	1.0	LINEAR
NoDigi	value	time	plotNoDigi	CIRCLE	darkKhaki	true	1.0	LINEAR

Time Stack Chart: chartWearTear

Name	Value

Name	Value
General	
Lock	false
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Dataset Samples To Keep	100
Scale	
Time window	100
Time	years
Vertical scale	Fixed
Chart Vertical Scale: To	100
Appearance	
Labels horizontal position	DEFAULT
Labels vertical position	DEFAULT
Label format	Model time units
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	1230.0
Width	300.0
y	20.0
Height	180.0
Legend	
Show legend	false
Chart area	
Chart Area: X Offset	40.0
Chart Area: Width	240.0
Chart Area: Y Offset	10.0
Chart Area: Height	130.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Time window moves	Continuously
Show name	false
Logging	true

Plot Items:

Title	Type	Dataset / Value	Color
Unfit for purpose	value	archive.NCriticalWearTear()*100.0/archive.NEverRequested()	sandyBrown
Good condition wear tear	value	archive.NGoodWearTear()*100.0/archive.NEverRequested()	new Color(176, 196, 222, 157)
Fair condition wear tear	value	archive.NFairWearTear()*100.0/archive.NEverRequested()	gold
Critical condition wear tear	value	archive.NPoorWearTear()*100.0/archive.NEverRequested()	red

Time Stack Chart: chartChemical

Name	Value
General	
Lock	false
Public	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Dataset Samples To Keep	500
Scale	
Time window	500
Time	years
Vertical scale	Fixed
Chart Vertical Scale: To	100
Appearance	
Labels horizontal position	DEFAULT
Labels vertical position	DEFAULT
Label format	Model time units
Labels Text Color	darkGray
Chart Area Grid Color	darkGray
Position and size	
x	310.0
Width	310.0
y	50.0
Height	180.0
Legend	
Show legend	false
Chart area	
Chart Area: X Offset	40.0
Chart Area: Width	250.0
Chart Area: Y Offset	10.0
Chart Area: Height	140.0
Chart Area: Background Color	white
Chart area border color	black
Advanced	
Time window moves	Continuously
Show name	false
Logging	true

Plot Items:

Title	Type	Dataset / Value	Color
Good condition	value	archive.NGood()*100.0/archive.size()	new Color(176, 196, 222, 157)
Fair condition	value	archive.NFair()*100.0/archive.size()	gold
Critical condition	value	archive.NPoor()*100.0/archive.size()	orangeRed

Data Set: datasetRequestsRR

Name	Value
General	
Dataset Samples To Keep	100
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	false
Logging	true

Data Set: datasetRequestsSoD

Name	Value
General	
Dataset Samples To Keep	100
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	false
Logging	true

Data Set: datasetRequests

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	requestsRR
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetDigi

Name	Value
General	
Dataset Samples To Keep	100
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	false
Logging	true

Data Set: datasetNoDigi

Name	Value
General	
Dataset Samples To Keep	100
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	false
Logging	true

Data Set: datasetGood

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NGood()*100/archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetFair

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NFair()*100/ archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetFairWearTear

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	zidz(archive.NFairWearTear(), archive.NEverRequested())
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true

Name	Value
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetCritical

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	archive.NPoor()*100/archive.size()
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetCriticalWearTear

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	zidz(archive.NCriticalWearTear(), archive.NEverRequested())
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetGoodWearTear

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	zidz (archive.NGoodWearTear(), archive.NEverRequested())
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Data Set: datasetPoorWearTear

Name	Value
General	
Dataset Samples To Keep	500
Axis Data Vertical Y Axis	zidz(archive.NPoorWearTear(), archive.NEverRequested())
Axis Data Freeze X Axis	true
Show at runtime	true
Show name	true
Data update	
Analysis auto update	true
ANALYSIS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Logging	true

Edit Box: editbox1

Name	Value
General	
Enabled	true
Link	metersArchive
Link	metersArchive
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	50.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox2

Name	Value
General	
Enabled	true
Link	deacidificationRate
Link	deacidificationRate
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	180.0

Name	Value
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox3

Name	Value
General	
Enabled	true
Link	maxPH
Link	maxPH
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	260.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox4

Name	Value
General	
Enabled	true
Link	minDP
Link	minDP
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	290.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox5

Name	Value
General	
Enabled	true
Link	temperature
Link	temperature
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	100.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox6

Name	Value
General	
Enabled	true
Link	relativeHumidity
Link	relativeHumidity
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	130.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox7

Name	Value
General	
Enabled	true
Link	weigth
Link	weigth
Link to	true
Lock	false
Public	true

Name	Value
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	340.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox8

Name	Value
General	
Enabled	true
Link	price
Link	price
Link to	true
Lock	false
Public	true
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0
y	370.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Edit Box: editbox9

Name	Value
General	
Enabled	true
Link	budget
Link	budget
Link to	true
Lock	false
Public	true
Action	
Action	deacidificationRate = budget/ (weigth * price);
Appearance	
Text color	steelBlue
Position and size	
x	200.0
Width	70.0

Name	Value
y	210.0
Height	20.0
Advanced	
Show name	false
Embedded icon	false

Excel File: excelFile

Name	Value
General	
File name	ResourceReference: readingroom.outputRequests.xlsx (Resolved: true)
Show at runtime	true
Show name	true
Advanced	
Load on model startup	true
Save on model termination	true
Save in snapshot	false

Link to agents: connections

Name	Value
General	
Show at runtime	true
Show name	true
Communication	
Message type	Object
Animation	
Draw line	false

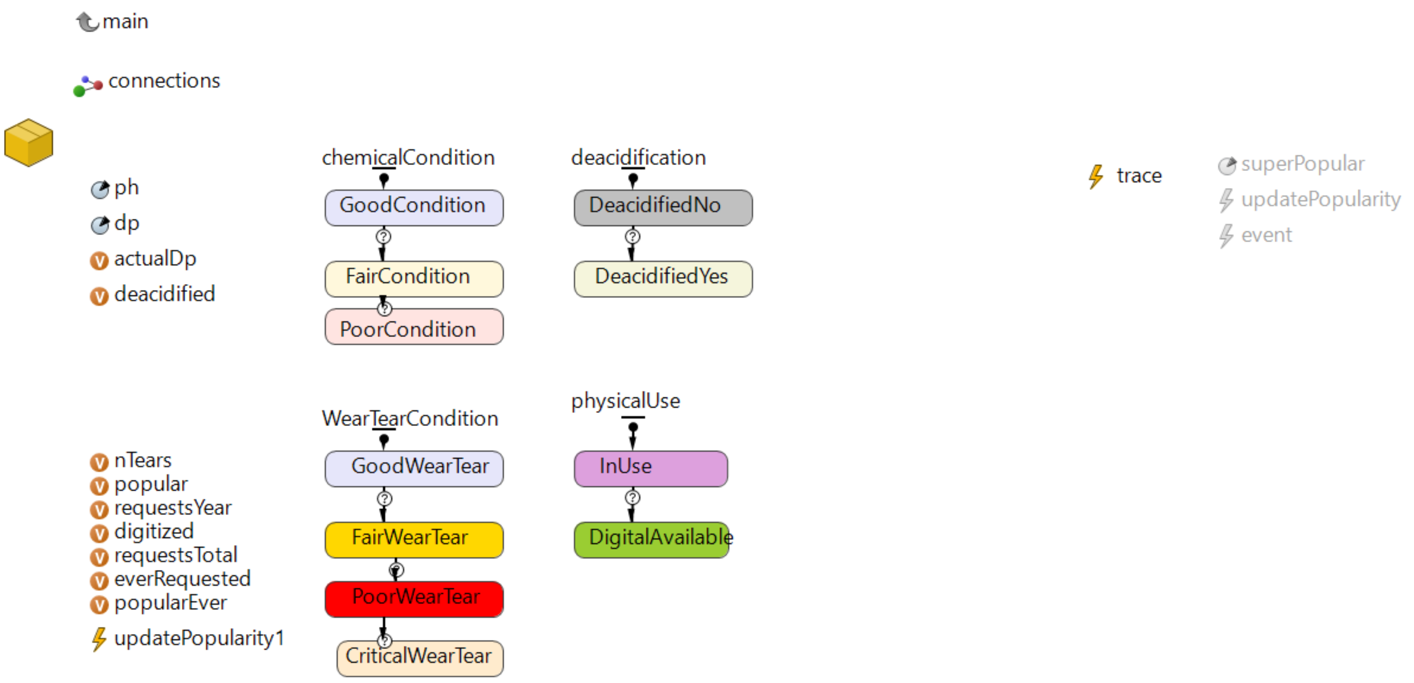
Agent Type: InventoryNumber

Name	Value
Agent actions	

Startup code	//On startup number of requests for popular is set//
	//if (popular == true)
	 //requestsYear = (int) uniform(0, 3);
	 //requestsTotal= (int) exponential(2, 10, 2, 1.5);
	//number of inventory requested once is chosen//
	//else requestsTotal = randomTrue(0.04) ? 1 : 0;
	
	//if (popular!= true)
	//requestsTotal = randomTrue(0.04) ? 1 : 0;
	
	if (requestsTotal >=1)
	{everRequested = true;
	popularEver = requestsYear;
	requestsTotal = requestsYear;
	}
	
	//if (popular == true)
	//{superPopular = randomTrue (0.8);
	//everRequested = true;
	//}

Agent in flowcharts	
Use in flowcharts as	Agent

Name	Value
Dimensions and movement	
Speed	(10 : MPS)
Rotate animation towards movement	true
Rotate vertically as well (along Z-axis)	false
Space and network	
Space Type	Continuous
Advanced Java	
Generic	false
Advanced	
Logging	true
Auto-create datasets	true
AOC_DATASETS_UPDATE_TIME_PROPERTIES	- Recurring Event Properties
Limit the number of data samples	false



Parameter: ph

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Parameter: dp

Name	Value
General	
Array	false
Type	double
Show at runtime	true
Show name	true
Value editor	
Editor control	Text
Advanced	
System dynamics units	false
Save in snapshot	true

Event: trace

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Occurs once
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	
Action	traceln(actualDp + " " + popular + " " + nTears + " "+requestsYear+ " " + requestsTotal + " " + digitized);

Event: updatePopularity1

Name	Value
General	
Logging	true
EVENT_TIMEOUT_PROPERTIES	- Recurring Event Properties
Mode	Cyclic
Trigger type	Timeout
Show at runtime	true
Show name	true
Action	

Action	//non-popular are requested just once,
	//therefore every year requestYear are 0 at the beginning of the year
	//if (popular != true)
	//requestsYear = 0;
	
	//set number of requests for popular agents//
	//if (inState (InUse) && popular == true)
	//requestsYear = (int) uniform (3,15); pert(1, 20, 3)binomial(0.2, 3)
	
	//number of uses for those that are popular
	//if (popular == true & superPopular==false){
	//requestsYear = (int) pert(0,2,1);
	//}
	
	//number of uses for those that are superpopular
	//if (popular == true & superPopular==true){

Name	Value

```
//requestsYear = (int) binomial(0.2, 3);
//}
requestsYear = 0;

if (main.requestsRR > 0)
{
if (inState (InUse) && popular == true)
requestsYear = (int) normal(1, -1);
//requestsYear =(int) exponential(1, 1);
//exponential(0.3, 1);
//requestsYear = 20000;

if (inState (InUse) && popular == true && requestsYear > 0)
requestsTotal += requestsYear;
else
requestsTotal += 0;

//chose those agents that are requested once//
if (inState (InUse) && popular != true && requestsTotal == 0)
requestsTotal = randomTrue(0.003) ? 1 : 0;


//requestsTotal are update//
if (requestsYear > 0)
{//requestsTotal += requestsYear;
popularEver = requestsYear;
everRequested = true;
}

if (inState (InUse))
{
if (actualDp < 800)
{double Intears = (-0.0096*actualDp+7.0236);
double tears = exp(Intears);
nTears = roundToInt (tears*requestsTotal);
}
if (actualDp > 800 && requestsYear >0)
{
    nTears+=requestsTotal*0.1;
     }
}

}
if (nTears < 0.0)
nTears = 0;
```

Variable: digitized

Name	Value
General	
Type	boolean
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: everRequested

Name	Value
General	

Name	Value
Type	boolean
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: actualDp

Name	Value
General	
Initial value	dp
Type	double
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: deacidified

Name	Value
General	
Initial value	false
Type	boolean
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: popular

Name	Value
General	
Type	boolean
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: requestsYear

Name	Value
General	
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: requestsTotal

Name	Value
General	
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Variable: popularEver

Name	Value
General	
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

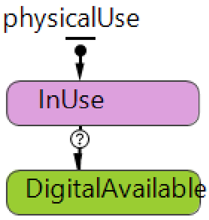
Variable: nTears

Name	Value
General	
Initial value	0
Type	int
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false

Name	Value
Save in snapshot	true
System dynamics units	false

Statechart Entry Point: physicalUse

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: Digitize

Name	Value
General	
Condition	digitized == true
Trigger type	Condition
Show name	false

State: InUse

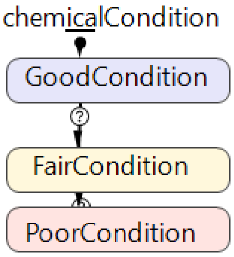
Name	Value
General	
Fill color	plum
Show name	true

State: DigitalAvailable

Name	Value
General	
Fill color	yellowGreen
Show name	true

Statechart Entry Point: chemicalCondition

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: transition

Name	Value
General	
Condition	actualDp < 800
Trigger type	Condition
Show name	false

Transition: transition1

Name	Value
General	
Condition	actualDp < 300
Trigger type	Condition
Show name	false

State: GoodCondition

Name	Value
General	
Entry action	shapeBox.setFillColor(lavender);
Fill color	lavender
Show name	true

State: FairCondition

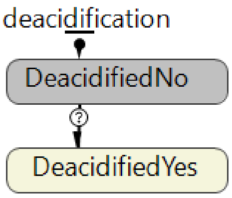
Name	Value
General	
Entry action	shapeBox.setFillColor(cornsilk);
Fill color	cornsilk
Show name	true

State: PoorCondition

Name	Value
General	
Entry action	shapeBox.setFillColor(mistyRose);
Fill color	mistyRose
Show name	true

Statechart Entry Point: deacidification

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: transition2

Name	Value
General	
Condition	deacidified == true
Trigger type	Condition
Show name	false

State: DeacidifiedNo

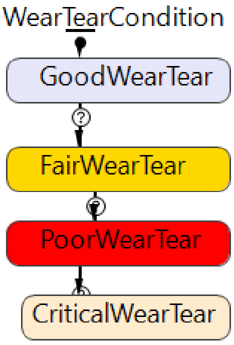
Name	Value
General	
Entry action	shapeBox.setLineColor(midnightBlue);
Fill color	silver
Show name	true

State: DeacidifiedYes

Name	Value
General	
Entry action	shapeBox.setLineColor(red);
Fill color	beige
Show name	true

Statechart Entry Point: WearTearCondition

Name	Value
General	
Logging	true
Show at runtime	true
Show name	true



Transition: transition3

Name	Value
General	
Condition	nTears > 20
Trigger type	Condition
Show name	false

Transition: transition4

Name	Value
General	
Condition	nTears > 60
Trigger type	Condition
Show name	false

Transition: transition5

Name	Value
General	
Condition	nTears > 100
Trigger type	Condition
Show name	false

State: GoodWearTear

Name	Value
General	
Entry action	shapeBox.setFillColor(lavender);
Fill color	lavender
Show name	true

State: FairWearTear

Name	Value
General	
Entry action	shapeBox.setFillColor(gold);
Fill color	gold
Show name	true

State: PoorWearTear

Name	Value
General	
Entry action	shapeBox.setFillColor(red);
Fill color	red
Show name	true

State: CriticalWearTear

Name	Value
General	
Entry action	shapeBox.setFillColor(blanchedAlmond);
Fill color	blanchedAlmond
Show name	true

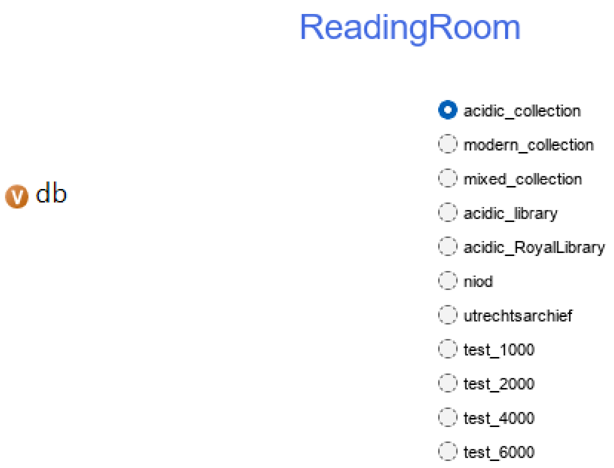
Link to agents: connections

Name	Value
General	
Show at runtime	true
Show name	true
Communication	
Message type	Object
Animation	
Draw line	false

Simulation Experiment: Simulation

Name	Value
General	
Bypass Initial Simulation Screen	false
Maximum available memory	512
Agent type	Main
Model time	
Execution mode	Virtual time (as fast as possible)
Stop option	Stop at specified time
Initial time	0.0
Final time	500.0
Initial date	Thu Jan 01 00:00:00 GMT 2004
Randomness	
Random Number Generation Type	Random seed (unique simulation runs)
Window	
Title	ReadingRoom : Simulation
Enable zoom and panning	true
Enable developer panel	true
Show developer panel on start	false
Java actions	
Before simulation run	getEngine().setSimultaneousEventsSelectionMode(Engine.EVENT_SELECTION_FIFO);

Name	Value
Advanced	
Load root from snapshot	false



Variable: db

Name	Value
General	
Initial value	"acidic_collection_4000"
Type	String
Show at runtime	true
Show name	true
Advanced	
Access type	public
Constant	false
Save in snapshot	true
System dynamics units	false

Radio Buttons: radio

Name	Value
General	
Enabled	true
Link to	false
String values	acidic_collection, modern_collection, mixed_collection, acidic_library, acidic_RoyalLibrary, niod, utrechtsarchief, test_1000, test_2000, test_4000, test_6000
Orientation	Vertical
Lock	false
Action	
Action	if (radio.getValue() == 0) db = "acidic_collection_4000";

Name	Value

	if (radio.getValue() == 1)
	db = "modern_collection_4000";
	if (radio.getValue() == 2)
	db = "mixed_collection_4000";
	if (radio.getValue() == 3)
	db = "acidic_library_4000";
	if (radio.getValue() == 4)
	db = "acidic_royallibrary_4000";
	if (radio.getValue() == 5)
	db = "niod_4000";
	if (radio.getValue() == 6)
	db = "utrechtsarchief_4000";
	if (radio.getValue() == 7)
	db = "test_1000";
	if (radio.getValue() == 8)
	db = "test_2000";
	if (radio.getValue() == 9)
	db = "test_4000";
	if (radio.getValue() == 10)
	db = "test_6000";

Position and size	
x	140.0
Width	150.0
y	90.0
Height	310.0
Advanced	
Show name	false

Database: Database

Name	Value
General	
Database shutdown compacr	false
Log	
Logging	false

Database Table: digitization

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
db_blockdigitization	DOUBLE	-	-	
db_year	INTEGER	-	-	
db_blockdigitization 2	DOUBLE	-	-	
db_sod	DOUBLE	-	-	
db_matchsod	DOUBLE	-	-	
db_matchblock	DOUBLE	-	-	
db_once	DOUBLE	-	-	

Database Table: digitization_nh

Name	Value

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
db_year	INTEGER	-	-	
db_blockdigitization	DOUBLE	-	-	
db_sod	DOUBLE	-	-	
db_matchsod	DOUBLE	-	-	
db_matchblock	DOUBLE	-	-	
db_once	DOUBLE	-	-	
db_once2	DOUBLE	-	-	

Database Table: digitization_ua

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
db_year	INTEGER	-	-	
db_blockdigitization	DOUBLE	-	-	
db_sod	DOUBLE	-	-	
db_matchsod	DOUBLE	-	-	
db_matchblock	DOUBLE	-	-	
db_once	DOUBLE	-	-	

Database Table: acidic_collection_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: acidic_library_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: acidic_royallibrary_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: mixed_collection_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: modern_collection_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
ph	DOUBLE	-	-	
dp	DOUBLE	-	-	

Database Table: niod_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: utrechtsarchief_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: digitization_test

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
db_blockdigitization	DOUBLE	-	-	
db_year	INTEGER	-	-	
db_sod	DOUBLE	-	-	
db_matchsod	DOUBLE	-	-	
db_matchblock	DOUBLE	-	-	
db_once	DOUBLE	-	-	

Database Table: test_1000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: test_2000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Database Table: test_4000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	

null	null	null	null	null
ph	DOUBLE	-	-	

Database Table: test_6000

Name	Value
Advanced	
Cached database table	false

null	null	null	null	null
dp	DOUBLE	-	-	
ph	DOUBLE	-	-	

Bibliography

4C Collaboration to clarify the costs of curation (n.d.).
URL: *https://www.4cproject.eu/*

Ahmed, A., Greensmith, J. and Aickelin, U. (2012). Variance in system dynamics and agent based modelling using the SIR model of infectious disease, *Proceedings - 26th European Conference on Modelling and Simulation, ECMS 2012* .

Ahn, K., Henniges, U., Blüher, A., Banik, G. and Potthast, A. (2011). Sustainability of mass deacidification. Part I: Concept, selection of sample books and pH-determination, *Restaurator* **32**(3): 193–222.

American Institute for Conservation (n.d.). Collection Care Network.
URL: *https://www.culturalheritage.org/membership/groups-and-networks/collection-care*

Andres, H., Blüher, A., Grossenbacher, G., Reist, M. and Wälchli, M. (2008). The Papersave Swiss-Process quality control and efficacy, *Restaurator* **28**: 3–28.

Andro, M. (2018). *Digital libraries and crowd sourcing*, John Wiley & Sons, Hoboken, NJ, USA.

ARA (2019). Survey of Visitors to UK Archives 2018, *Technical report*, ARA National Surveys Group.
URL: *https://www.archives.org.uk/images/Public_Services_Quality_Grp/Survey_of_Visitors_to_UK_Archives_2018_-_National_Headline_Report_.pdf*

ARA (n.d.). Archive Surveys.
URL: *https://www.archives.org.uk/what-we-do/archive-surveys.html*

Ashley, K. (1999). Digital archive costs. Facts and fallacies., *Proceedings of the DLM-Forum on Electronic Records (DLM '99)*.

Ashley-Smith, J. (1999). *Risk Assessment for Object Conservation*, Butterworth-Heinemann.

Ashley-Smith, J. (2013). British Publicly Available Specification (PAS) 198, "Specification for managing environmental conditions for cultural collections", *Summit on the Museum Preservation Management*, Smithsonian Institution, pp. 29–32.

Astle, P. J. and Muir, A. (2002). Digitization and preservation in public libraries and archives, *Journal of Librarianship and Information Science* **34**(2): 67–76.

Auchincloss, A. H. and Roux, A. V. D. (2008). A new tool for epidemiology. The usefulness of dynamic-agent models in understanding place effects on health, *American Journal of Epidemiology* **168**(1): 1–8.

Ayris, P., Wheatley, P., Aitken, B., Hole, B., McCann, P., Peach, C. and Lin, L. (2010). The LIFE3 Project: Bringing digital preservation to LIFE, *Technical report*.
URL: *http://www.life.ac.uk/3/docs/life3_report.pdf*

Baird, B. J., Krentz, J. and Schaffner, B. (1997). Findings from the condition surveys conducted by the University of Kansas Libraries, *College & Research Libraries* **58**(2): 115–126.

Bajzíková, M., Hanus, J. and Bukovský, V. (2008). Characterisation of the Slovak national library collections, *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, pp. 51–52.

Baker, N. (2001). *Double Fold: Libraries and the Assault on Paper*, Random House, New York.

Balazic, A., Habicht, S., Smodis, M., Kolar, J. and Strlic, M. (2007). Extending the useful life of paper. Evaluation of the effect of various preservation actions., *in* T. Padfield and K. Borchersen (eds), *Museum Microclimates*, National Museum of Denmark, pp. 39–41.
URL: *http://www.conservationphysics.org/mm/balazic/balazic.pdf*

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study, *Annals of Operations Research* **53**: 121–173.

Bansa, H. (1998). Aqueous deacidification with calcium or with magnesium?, *Restaurator* **19**: 1–40.

Bantin, J. and Agne, L. (2010). Digitizing for value: A user-based strategy for university archives, *Journal of Archival Organization* **8**(3-4): 244–250.

Barański, A., Łagan, M. and Łojewski, T. (2005). Acid-catalysed degradation, *in* M. Strlič and J. Kolar (eds), *Ageing and stabilisation of paper*, National and University Library, Ljubljana, chapter 6, pp. 85–100.

Barnes, D. J. and Chu, D. (2015). *Guide to simulation and modeling for biosciences*, second edn, Springer International Publishing, London.

Barrow, W. (1974). *Permanence/Durability of the book—VII. Physical and chemical properties of book papers, 1507-1949.*, W.J. Barrow Research Laboratory Inc., Richmond, Virginia.

Barton, P., Bryan, S. and Robinson, S. (2004). Modelling in the economic evaluation of health care : selecting the appropriate approach, *Journal of Health Services Research and Policy* **9**(2): 110–118.

Baty, J. W., Maitland, C. L., Minter, W., Hubbe, M. A. and Jordan-Mowery, S. K. (2010). Deacidification for the conservation and preservation of paper-based works. A review., *BioResources* **5**(3): 1955–2023.

Beagrie, N. and Daphne, C. (2017). D4.9 Cost-benefit advocacy toolkit deliverable report, *Technical report*, CESSDA.

Beagrie, N., Lavoie, B. and Woollard, M. (2010). Keeping research data safe 2, final report, *Technical report*, JISC.
URL: *http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf*

Beentjes, G. (2016). Ontzuren . . . of niet.
URL: *https://www.metamorfoze.nl/nieuws-agenda/ontzuren-niet*

Bell, N., Cassar, M. and Strlič, M. (2018). Evidence for informed preservation planning and advocacy. A synoptic view, *Studies in Conservation* **63**(sup1): 8–14.

Bilcke, J., Beutels, P., Brisson, M. and Jit, M. (2011). Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: A practical guide, *Medical Decision Making* **31**(4): 675–692.

Blüher, A. and Großenbacher, G. (eds) (2006). *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern.

Bogvad, K., Hougaard, E., Thirifays, A. and Bo, N. (2014). D3.1 Evaluation of cost models and needs & gaps analysis., *Technical report*.
URL: *cproject.eu/summary-of-cost-models/*

Bonabeau, E. (2002). Agent-based modeling : Methods and techniques for simulating human systems, *PNAS* **99**(Suppl. 3): 7280–7287.

Borshchev, A. (2013). Multi-method modeling, *in* R. Pasupathy, S. H. Kim, A. Tolk, R. R. Hill and M. Kuhl (eds), *Proceedings of the 2013 Winter Simulation Conference*, pp. 4089–4100.

Borshchev, A. and Filippov, A. (2004). From system dynamics and discret event to practical agent based modeling: reasons, technques, tools, *The 22nd International Conference of the System Dynamics Society*.

Brailsford, S., Churilov, L. and Dangerfield, B. (eds) (2014). *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd.

Brailsford, S. and Hilton, N. (2001). A comparison of discrete event simulation and system dynamics for modelling healthcare systems, *in* J. Riley (ed.), *Proceedings from ORAHS 2000*, Caledonian University, Glasgow, pp. 18–39.

Brancolini, K. R. (2000). Selecting research collections for digitization: Applying the Harvard model, *Library Trends* **48**(4): 783–799.

Brandis, L. (1993). A comparision of methods used for measuring the pH of paper., *AICCM BUlletin* **18**: 7–17.

Brennan, A., Chick, S. E. and Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies, *Health Economics* **15**: 1295–1310.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J. and Paltiel, A. D. (2012). Model parameter estimation and uncertainty. A report of the ISPOR-SMDM modeling good research practices task force-6, *Value in Health* **15**(6): 835–842.

British Library (2004). Future life of collections. Report on a meeting on Applied Conservation Research, held at The British Library on 27-28 September 2004, *Technical report*, British Library, London.

British Standards Institution (2012). *PAS 198:2012 Specification for managing environmental conditions for cultural collections*, BSI Standards Limited.

Brokerhof, A. W., Ankersmit, B. and Ligterink, F. (2017). *Risk management for collections*, Cultural Heritage Agency of the Netherlands, Amesfoort.

Brokerhof, A. W. and Bülow, A. E. (2016). The QuiskScan-a quick risk scan to identify value and hazards in a collection, *Journal of the Institute of Conservation* **39**(1): 18–28.

Brown, L. (2003). Useful or useless use statistics? A summary of conference presentations on usage data from the 22nd Annual Charleston Conference, Issues in Book and Serial Acquisition, *Serials Review* **29**(2): 145–150.

Brown, N., Coppola, F., Modelli, A., Amicucci, F., Lichtblau, D. and Strlič, M. (2020). Non-destructive collection survey of the historical Classense Library. Part I: Paper characterisation, *Heritage Science* **8**(1): 1–11.

Brown, N., Lichtblau, D., Fearn, T. and Strlič, M. (2017). Characterisation of 19th and 20th century Chinese paper, *Heritage Science* **5**(47).

Bülow, A. E. (2010). Collection management using preservation risk assessment, *Journal of the Institute of Conservation* **33**(1): 65–78.

Bülow, A. E. and Brokerhof, A. W. (2011). The QALY in collection care: A cost effectiveness approach to collection management, *in* J. Bridgland (ed.), *ICOM-CC 16th Triennial Conference Lisbon 19-23 September 2011: Preprints*, Critério, Almada.

Bülow, A. E., Stitt, J. and Brokerhof, A. W. (2018). I can see further now. Preventive conservation in a changing heritage world, *Studies in Conservation* **63**(sup1): 35–42.

Calvini, P. (2005). The influence of levelling-off degree of polymerisation on the kinetics of cellulose degradation, *Cellulose* **12**(4): 445–447.

Calvini, P. (2012). The role of the Ekenstam equation on the kinetics of cellulose hydrolytic degradation, *Cellulose* **19**: 313–318.

Campagnolo, A. (ed.) (2020). *Book conservation and digitization: The challenges of dialogue and collaboration*, ARC Humanities Press.

Cassar, M. (1998). *Cost / benefits appraisals for collection care*, Bigwood and Staple Ltd., Bridgwater, Somerset.

Cassidy, R., Singh, N. S., Schiratti, P. R., Semwanga, A., Binyaruka, P., Sachingongu, N., Chama-Chiliba, C. M., Chalabi, Z., Borghi, J. and Blanchet, K. (2019). Mathematical modelling for health systems research: A systematic review of system dynamics and agent-based models, *BMC Health Services Research* **19**(1): 15–17.

CCSDS (2012). Reference model for an open archival information system (OAIS). **URL:** *http://public.ccsds.org/publications/archive/650x0m2.pdf*

Cedzová, M., Gállová, I. and Katuščák, S. (2006). Patents for paper deacidification, *Restaurator* **27**(1): 35–45.

Cerdá, M., Tracy, M., Keyes, K. M. and Galea, S. (2015). To Treat or to Prevent? Reducing the population burden of violence-related post-traumatic stress disorder, *Epidemiology* **26**(5): 681–689.

Chahal, K. and Eldabi, T. (2010). A multi-perspective comparison for selection between system dynamics and discrete event simulation, *International Journal of Business Information Systems* **6**(1): 4–17.

Chapman, J. and Yakel, E. (2012). Data-driven management and interoperable metrics for special collections and archives user services, *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* **13**(2): 129–151.

Chapman, S. (2006). Counting the costs of digital preservation. Is repository storage affordable?, *Journal of Digital Information* **4**(2): 1–15.

Chung, C. a. (2003). *Simulation modeling handbook. A practical approach*, CRC Press, Boca Raton.

Cimler, R., Tomaskova, H., Kuhnova, J., Dolezal, O., Pscheidl, P. and Kuca, K. (2018). Numeric, agent-based or system dynamics model? Which modeling approach is the best for vast population simulation?, *Current Alzheimer Research* **15**(8): 789–797.

Cooper, A. K., Brailsford, S. C. and Davies, R. (2007). Choice of modelling technique for evaluating health care interventions, *The Journal of the Operational Research Society* **58**(2): 168–176.

Coppola, F., Brown, N., Amicucci, F., Strlič, M. and Modelli, A. (2020). Nondestructive collection survey of the historical Classense Library. Part II: Conservation scenarios, *Heritage Science* **8**(1): 1–10.

Cosaert, A. and Beltran, V. (2021). Comparison of temperature and relative humidity analysis tools to address practitioner needs and improve decisionmaking., *in* J. Bridgland (ed.), *Transcending Boundaries: Integrated Approaches to Conservation. ICOM-CC 19th Triennial Conference Preprints, Beijing, 17–21 May 2021*, International Council of Museums, Paris.

COUNTER (2021). COUNTER Code of Practice Release 5.0.2.
URL: *https://cop5.projectcounter.org/en/5.0.2/*

Coyle, K. (2006). Mass digitization of books, *Journal of Academic Librarianship* **32**(6): 641–645.

Cséfalvayová, L., Strlič, M. and Karjalainen, H. (2011). Quantitative NIR chemical imaging in heritage science, *Analytical Chemistry* **83**(13): 5101–5106.

Currie, C. S., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A. and Tako, A. A. (2020). How simulation modelling can help reduce the impact of COVID-19, *Journal of Simulation* **14**(2): 83–97.

Dalgleish, P. (2011). The thorniest area: making collections accessible online while respecting individual and community sensitivities, *Archives and Manuscripts* **39**(1): 67–84.

Daly, L. E. and Bourke, G. (2007). *Statestical analysis, interpretation and uses of medical statistics*, 5th edn, Blackwell Science Ltd, Oxford.

Dangerfield, B. (2014). Systems thinking and system dynamics. A primer, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discrete-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., chapter 3, pp. 26–51.

Daniels, V. (2006). Paper, *in* E. May and M. Jones (eds), *Conservation Science*, RSC Publishing, Cambridge, chapter 3.

Darabi, N. and Hosseinichimeh, N. (2020). System dynamics modeling in health and medicine: a systematic literature review, *System Dynamics Review* **36**(1): 29–73.

Davahli, M. R., Karwowski, W. and Taiar, R. (2020). A system dynamics simulation applied to healthcare: A systematic review, *International Journal of Environmental Research and Public Health* **17**(16): 1–27.

de Pinho, H. (2015). Participant guidelines systems tools for complex health systems. A guide to creating causal loop diagrams, *Technical report*, Mailman School of Public Health, Columbia University, New York City.

de Wit, M. (2006). *Hambase. Heat, air and moisture model for building and systems evaluation*, Techinische Universiteit Eindhoven, Eindhoven.

Di Pietro, G., Ligterink, F., Porck, H. and de Bruin, G. (2016). Chemical air filtration in archives and libraries reconsidered, *Studies in Conservation* **61**(5): 245–254.

Diaz, R., Behr, J. G. and Britton, B. S. (2015). Estimating cost adjustments required to accomplish target savings in chronic disease management interventions. A simulation study, *Simulation Transactions of the Society for Modeling and Simulation International* **91**(7): 599–614.

Ding, H. (2009). Degradation phenomenology and life modelling of paper., *in* S. I'Anson (ed.), *Advances in Pulp and Paper Research, Trans. of the XIVth Fund. Res. Symp. Oxford*, FRC, Manchester, pp. 947–1000.

Ding, H. and Wang, Z. D. (2008). On the degradation evolution equations of cellulose, *Cellulose* **15**(August 2007): 205–224.

Djanatliev, A. and German, R. (2013). Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation, *in* R. Pasupathy, S. H. Kim, A. Tolk, R. R. Hill and M. Kuhl (eds), *Proceedings of the 2013 Winter Simulation Conference*, pp. 270–281.

Djanatliev, A., German, R., Kolominsky-rabas, P. and Hofmann, B. M. (2012). Hybrid simulation with loosely coupled system dynamics and agent based models for prospective health technology assessments, *in* C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A. Uhrmacher (eds), *Proceedings of the 2012 Winter Simulation Conference*, pp. 770–781.

Dobreva, M., O'Dwyer, A. and Konstantelos, L. (2011). User needs in digitization, *in* L. Hugues (ed.), *Evaluating and Measuring the Value, use and Impact of Digital Collections*, Facet, chapter 6.

Dong, Y. (2020). Multivariable linear regression analysis of degree of polymerization and paper composition, *Master thesis*, University College London.

Dörner, D. (1990). The logic of failure, *Philosophical Transactions of the Royal Society B* **327**(1241): 463–473.

Duff, W. M., Yakel, E., Tibbo, H. R., Cherry, J. M., Mckay, A., Krause, M. G. and Sheffield, R. (2010). The development, testing, and evaluation of the archival metrics toolkits, *The American Archivist* **73**: 569–599.

Dupont, A. and Cheradame, H. (2008). Simultaneous deacidification and reinforcement of paper by treatment with aminoalkylalkoxysilanes, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, p. 37.

Duran, C., Grau-Bové, J. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part I: exploring the concepts of reliability and epidemiology, *Heritage Science* **7**(1).

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2017). Data mining in collections: From epidemiology to demography, *ICOM-CC 18th Triennial Conference, 2017 Copenhagen*.

Duran-Casablancas, C., Grau-Bové, J., Fearn, T. and Strlič, M. (2019). Accumulation of wear and tear in archival and library collections. Part II: an epidemiological study, *Heritage Science* **7**(11).

Dutch Admin. (1995). Act of the Minister of Education, Culture and Science, of 28 April 1995, regarding the sustainability and proper and accessible condition of archive records and the construction and setup of archive rooms and depositories (Archiefwet).
URL: *https://wetten.overheid.nl/BWBR0007376/2020-01-01*

Dutch Admin. (2014). Regulation of the Minister of Education, Culture and Science, of 17 December 2009, no. WJZ/178205 (8189), regarding the sustainability and proper and accessible condition of archive records and the construction and setup of archive rooms and depositories.
URL: *https://wetten.overheid.nl/BWBR0027041/2014-01-01*

Eden, P., Dungworth, N., Bell, N. and Matthews, G. (1998). *A model for assessing preservation needs in libraries*, British Library Research and Innovation Centre, Boston Spa.

Emsley, A. M. and Stevens, G. C. (1994). Kinetics and mechanisms of the low-temperature degradation of cellulose, *Cellulose* **1**(1): 26–56.

Erway, R. and Schaffner, J. (2017). *Shifting gears. Gearing up to get into the flow*, 2nd edn, OCLC Research, Dublin, OH.

European Commission (n.d.). Shaping Europe's digital future. Digital cultural heritage.
URL: *https://digital-strategy.ec.europa.eu/en/policies/cultural-heritage*

Feather, J. (1996). *Preservation and the management of library collections*, second edn, Library Association Publishing, London.

Feather, J. (2004). Introduction. principles and policies, *in* J. Feather (ed.), *Managing preservation for libraries and archives*, Ashgate, Aldershot, chapter 1.

Fenech, A., Dillon, C., Ntanos, K., Bell, N., Barrett, M. and Strlič, M. (2013). Modelling the lifetime of colour photographs in archival collections, *Studies in Conservation* **58**(2): 107–117.

Fenech, A., Fearn, T. and Strlič, M. (2012). Use of Design-of-Experiment principles to develop a dose-response function for colour photographs, *Polymer Degradation and Stability* **97**(4): 621–625.

Foot, M. (2006). Preservation policy and planning, *in* G. Gorman and S. Shep (eds), *Preservation management for libraries, archives and museums*, Facet, London, chapter 2, pp. 19–41.

Foot, M. M. (2013). *Building a preservation policy*, Preservation Advisory Centre, British Library.

Forrester, J. W. (1961). *Industrial dynamics*, M.I.T. Press, Cambridge.

Forrester, J. W. (1971). Counterintuitive behavior of social systems, *Theory and Decision* **2**: 109–140.

Frieder, R. (1993). CIC Cooperative project. Logistics and managements issues, *Proceedings of the New York state seminar on mass deacification*, Columbia University Libraries, New York.

Galea, S., Hall, C. and Kaplan, G. A. (2009). Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research, *International Journal Drug Policy* **20**(3): 209–216.

Galea, S., Riddle, M. and Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology, *International Journal of Epidemiology* **39**(2009): 97–106.

Gao, A., Osgood, N. D. and Dyck, R. F. (2014). A tripartite hybrid model architecture for investigating health and cost impacts and intervention tradeoffs for diabetic end-stage renal disease, *in* A. Tolk, S. Y. Diall, I. O. Ryzhov, L. Yilmaz, S. Buckley and J. Miller (eds), *Proceedings of the 2014 Winter Simulation Conference*, pp. 1676–1687.

Garmendia, J. (2017). Digital cataloguing practices at The National Archives, *Technical report*, The National Archives.
URL: *https://www.nationalarchives.gov.uk/documents /digital-cataloguing-practices-march-2017.pdf*

Garside, P., Bradford, K. and Hamlyn, S. (2018). The use of risk management to support preventive conservation, *Studies in Conservation* **63**(sup1): 94–100.

Gess, J. M. and Rodriguez, J. M. (eds) (2005). *The sizing of paper*, third edit edn, Tappi Press, Atlanta.

Ghaffarzadegan, N., Lyneis, J. and Richardson, G. P. (2011). How small system dynamics models can help the public policy process, *System Dynamics Review* **27**(1): 22–44.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making, *Annu. Rev. Psychol.* **62**: 451–482.

Gilbert, N. (2011). *Agent-based models*, SAGE Publications, Inc.

Gillesse, R., Jochems, R. and Maris, M. (2010). Handleiding rekenmodel digitaliseringskosten, *Technical report*, Digitaal Erfgoed Nederland.

Glouberman, S. and Zimmerman, B. (2002). Complicated and complex system. What would successful reform of Medicare look like?, *Discussion Paper N° 8*, Commission on the Future of Health Care in Canada, Ottawa, Ontario.

Goldenfeld, N. and Kadanoff, L. P. (1999). Simple lessons from complexity, *Science* **284**(5411): 87–89.

Greene, B. (2021). *Until the end of time*, Penguin Books.

Grossenbacher, G. (2006). Paper deacidification as a measure for preserving originals, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern, pp. 7–20.

Guerrero, C. N., Schwarz, P. and Slinger, J. H. (2016). A recent overview of the integration of system dynamics and agent-based modelling and simulation, *Proceedings of the 34th International Conference of the System Dynamics Society, July 17-21, 2016 Delft, Netherlands*, p. 1153.

Gunal, M. M. (2012). A guide for building hospital simulation models, *Health Systems* **1**: 17–25.

Gurnagul, N., Howard, R., Zou, X., Uesaka, T. and Page, D. (1993). The mechanical permanence of paper: a literature review, *J. Pulp Pap. Sci.* **19**: 160–166.

Hanus, J., Katuščák, S., Čeppan, M. and Bukovský, V. (2006). Research on paper deacidification in Slovakia, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, pp. 75–91.

Hanus, J. and Mináriková, J. (2002). Deacidification without equipment and money – dream or reality ?, *13th ICOM Triennial Meeting Rio de Janeiro Preprints*, pp. 603–608.

Heath, S. K., Brailsford, S., Buss, A. and Macal, C. M. (2011). Cross-paradigm imulation modeling. Challenges and successes, *in* S. Jain, R. Creasey, J. Himmelspach, K. White and M. Fu (eds), *Proceedings of the 2011 Winter Simulation Conference*, pp. 2788–2802.

Henderson, J. and Waller, R. (2016). Effective preservation decision strategies, *Studies in Conservation* **61**(6): 308–323.

Henry, M. (2013). Conservation environments, museum buildings, and sustainability, *Summit on the Museum Preservation Management*, Smithsonian Institution, pp. 20–24.

Hirsch, G., Homer, J., Evans, E. and Zielinski, A. (2010). System dynamics model for planning cardiovascular disease interventions, *American Journal of Public Health* **100**(4): 616–622.

Hirsch, G., Homer, J., Trogdon, J., Wile, K. and Orenstein, D. (2014). Using simulation to compare 4 categories of intervention for reducing cardiovascular disease risks, *American Journal of Public Health* **104**(7): 1187–1195.

Hirtle, P. B. (2002). The impact of digitization on special collections in libraries, *Libraries & Culture* **27**(1).

Holtman, M. (2017). Digitisation based on industrial principles. A new vision for the quality management of mass digitisation projects, *Technical report*, Amsterdam City Archives, Amsterdam.

Homer, J. B. and Hirsch, G. B. (2006). System dynamics modeling for public health: Background and opportunities, *American Journal of Public Health* **96**(3): 452–458.

Homer, J., Hirsch, G., Minniti, M. and Pierson, M. (2004). Models for collaboration. How system dynamics helped a community organize cost-effective care for chronic illness, *System Dynamics Review* **20**(3): 199–222.

Homer, J., Wile, K., Yarnoff, B., Trogdon, J. G., Hirsch, G., Cooper, L., Soler, R. and Orenstein, D. (2014). Using simulation to compare established and emerging interventions to reduce cardiovascular disease risk in the United States, *Preventing Chronic Disease* **11**(140130).

Hong, S. H., Strlič, M., Ridley, I., Ntanos, K., Bell, N. and Cassar, M. (2012). Climate change mitigation strategies for mechanically controlled repositories. The case of The National Archives, Kew, *Atmospheric Environment* **49**: 163–170.

Hubbe, M. A. (2006). Paper's resistance to wetting. A Review of internal sizing chemicals and their effects, *BioResources* **2**(1): 106–145.

Hughes, L. (2003). Why digitize? The costs and benefits of digitization, *in* L. Hughes (ed.), *Digitizing Collections. Strategic Issues for the Information Manager*, Facet Publishing, London, chapter 1.

Hughes, M. (2016). A long-term study of collection use based on detailed Library of Congress classification. A statistical tool for collection management decisions, *Collection Management* **41**(3): 152–167.

ICN (2000). De steekproef alshulpmiddel bij collectiebeheer, *ICN Informatie* **4**.
URL: *https://kennis.cultureelerfgoed.nl/index.php/Steekproef_-_hulpmiddel_bij_collectiebeheer*

IFLA (2002). Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives, *Technical report*, International Federation of Library Associations and Institutions (IFLA) and International Council on Archives (ICA).
URL: *https://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf*

Image Permanence Institute (IPI) (2018). eClimateNotebook. Fundamentals. Preservation metrics.
URL: *https://www.eclimatenotebook.com/fundamentals_nl.php*

Jerzy Haber Institute of Catalysis and Surface Chemistry (2020). HERIe. Quantitative assessment of risks to heritage assets.
URL: *https://www.herie.pl/*

Jeurgens, C. (2013). The scent of the digital archive: Dilemmas with archive digitisation, *BMGN - Low Countries Historical Review* **128**(4): 30–54.

Joel, A., Indictor, N., Hanlan, J. and Baer, N. (1972). The measurement and significance of pH in paper conservation, *Bulletin of the American Group. International Institute for Conservation of Historic and Artistic Works* **12**(2): 119–125.

Jones, A. P., Homer, J., Murphy, D. and Essien, J. (2006). Understanding diabetes population dynamics through simulation modeling and experimentation, *American Journal of Public Health* **96**(3): 488–494.

Jones, D. (2008). Mass digitization of historical records, *Serials* **21**(2).

Kahn, R., Robertson, R., Smith, R. and Eddy, D. M. (2008). The impact of prevention on reducing the burden of cardiovascular disease, *Diabetes Care* **31**(8).

Kelly, E. J. (2014). Assessment of digitized library and archives materials: A literature eview, *Journal of Web Librarianship* **8**(4): 384–403.

Kemp, J. (2016). How digitisation integrates in the world of archives preservation, *Journal of the Institute of Conservation* **39**(1): 57–63.

Kes, M. and Christensen, B. E. (2013). A re-investigation of the Mark-Houwink-Sakurada parameters for cellulose in Cuen: A study based on size-exclusion chromatography combined with multi-angle light scattering and viscometry, *Journal of Chromatography A* **1281**: 32–37.

Ketelaar, E., Huysmans, F. and van Mensch, P. (2016). Netherlands: archives, libraries and museums, *in* M. J. Bates and M. N. Maack (eds), *Encyclopedia of Library and Information Sciences*, third edition edn, CRC Press, pp. 3874–3900.

Kiesling, K. and Stroud, J. (1995). Mass deacidification of archives and manuscript collections, *Advances in preservation and access. Volume 2*, Learned Information, Medford, pp. 77–100.

Kim, D. (1992). Guidelines for drawing causal loop diagrams, *The Systems Thinker* **3**(1): 5–6.

Kingma, B. (2000). The costs of print, fiche, and digital access. The Early Canadiana Online Project., *D-Lib Magazine* **6**(2).

Kleijnen, J. P. (1995). Verification and validation of simulations models, *European Journal of Operational Research* **82**: 145–162.

Klein, G. A. (1999). *Sources of power. How people make decisions*, MIT Press, Cambridge, Massachusetts.

KNMI (n.d.). Maand- en jaarwaarden van de temperatuur, neerslag en luchtdruk.
URL: *https://www.knmi.nl/nederland-nu/klimatologie/maandgegevens*

Kolar, J. (2009). SSPI-006584. PaperTreat. Evaluation of mass deacidification processes. 3rd report, *Technical report*.

Kompatscher, K., Kramer, R. P., Ankersmit, B. and Schellen, H. L. (2019). Intermittent conditioning of library archives. Microclimate analysis and energy impact, *Building and Environment* **147**(October 2018): 50–66.

Konsa, K. (2007). Condition survey for the Estonian national preservation policy, *Restaurator* **28**(4): 239–255.

Kramer, R. P., Schellen, H. L. and van Schijndel, A. W. (2016). Impact of ASHRAE's museum climate classes on energy consumption and indoor climate fluctuations. Full-scale measurements in museum Hermitage Amsterdam, *Energy and Buildings* **130**.

Kramer, R., Schellen, H. and Van Schijndel, J. (2015). Energy impact of ASHRAE's museum climate classes. A simulation study on four museums with different quality of envelopes, *Energy Procedia* **78**: 1317–1322.

Kyujin, A. (2013). *Sustainability of mass deacidification of library objects*, PhD thesis, Universität für Bodenkultur Wien.

Lane, D. (2000). You just don't understand me. Modes of failure and success in the discourse between system dynamics and discrete event simulation, *Operational Research working papers, LSEOR 00.34*, Department of Operational Research, London School of Economics and Political Science, London.

Lättilä, L., Hilletofth, P. and Lin, B. (2010). Expert systems with applications hybrid simulation models – When , Why , How ?, *Expert Systems with Applications* **37**: 7969–7975.

Law, A. (2007). *Simulation modeling and analysis*, fourth edn, McGraw-Hill, New York.

Lejeune, B. (2007). The effects of online catalogues in London and other museums: A Study of an alternative way of access, *Papers from the Institute of Archaeology* **1**: 79–97.

Lesk, M. (1996). Substituting images for books. The economics for libraries., *Symposium on document analysis and information retrieval, April 15-17, 1996, Las Vegas, Nevad*.

Lichtblau, D. and Dobrusskin, S. (2009). Das SurveNIR – System zur Bewertung des Zustands und des Alterungsverhaltens von Papier, *Technical report*.

Lichtblau, D., Strlič, M., Trafela, T., Kolar, J. and Anders, M. (2008). Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics – a new instrument, *Applied Physics A* **92**(1): 191–195.

LIFE (2010). Life Cycle Information for E-Literature.
URL: *http://www.life.ac.uk/*

Ligterink, F. and Di Pietro, G. (2018). The limited impact of acetic acid in archives and libraries, *Heritage Science* **6**(1).

Lindsay, H. (2003). Preservation microfilming and digitization at London Metropolitan Archives: Surveying and conservation preparation prior to image capture, *Paper Conservator: Journal of the Institute of Paper Conservation* **27**: 47–57.

Ling, T. and McLean, A. (2004). Taking it to the people: Why the National Archives of Australia embraced digitisation on demand, *Australian Academic and Research Libraries* **35**(1): 2–15.

Lipsitz, L. A. (2012). Understanding health care as a complex system: The foundation for unintended consequences, *JAMA - Journal of the American Medical Association* **308**(3): 243–244.

Liu, S., Xue, H., Li, Y., Xu, J. and Wang, Y. (2018). Investigating the diffusion of agent-based modelling and system dynamics modelling in population health and healthcare research, *Systems Research and Behavioral Science* **35**(2): 203–215.

Łojewski, T., Zieba, K. and Łojewska, J. (2010). Size exclusion chromatography and viscometry in paper degradation studies. New Mark-Houwink coefficients for cellulose in cupri-ethylenediamine, *Journal of Chromatography A* **1217**(42): 6462–6468.

Lucchi, E. (2016). Multidisciplinary risk-based analysis for supporting the decision making process on conservation, energy efficiency, and human comfort in museum buildings, *Journal of Cultural Heritage* **22**: 1079–1089.

Luke, D. A. and Stamatakis, K. A. (2012). Systems science methods in public health: Dynamics, networks, and agents, *Annual Review of Public Health* **33**: 357–376.

Macal, C. M. (2010). To Agent-based simulation from system dynamics, *in* B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan and E. Yucessan (eds), *Proceedings of the 2010 Winter Simulation Conference*, pp. 371–382.

Macal, C. M. and North, M. J. (2009). Agent-based modeling and simulation, *in* M. Rossetti, R. R. Hill, B. Johansson and R. Ingalls (eds), *Proceedings of the 2009 Winter Simulation Conference*, pp. 86–98.

Macpherson, P. (2010). Building a better horse and buggy: The privileging of access in reading rooms over online access, *Archives and Manuscripts* **38**(2): 61–78.

Mahamoud, A., Roche, B. and Homer, J. (2013). Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto , Canada, *Social Science & Medicine* **93**: 247–255.

Mahgoub, H., Bardon, T., Lichtblau, D., Fearn, T. and Strlič, M. (2016). Material properties of Islamic paper, *Heritage Science* **4**(34).

Małachowska, E., Dubowik, M., Boruszewski, P., Łojewska, J. and Przybysz, P. (2020). Influence of lignin content in cellulose pulp on paper durability, *Scientific Reports* **10**(1): 1–12.

Marchal, B., Van Belle, S., De Brouwere, V. and Witter, S. (2013). Studying complex interventions. Reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco, *BMC Health Services Research* **13**(1).

Maroso, A. L. (2003). Digitization as a preservation practice: Look before you leap, *Serials Librarian* **45**(2): 27–37.

Marshall, B. D. L. and Galea, S. (2015). Practice of epidemiology formalizing the role of agent-based modeling in causal inference and epidemiology, *American Journal of Epidemiology* **181**(2): 92–99.

Martell, C. (2008). The absent user. Physical use of academic library collections and services continues to decline 1995-2006, *Journal of Academic Librarianship* **34**(5): 400–407.

Martens, M. (2012). *Climate risk assessment in museums. Degradation risks determined from temperature and relative humidity data*, PhD thesis, TU Eindhoven.

Mcavena, K. (2017). Information seeking: Taking a closer look at archival user studies, *Master thesis*, Universiteit Leiden.

McCausland, S. (2011). A future without mediation? Online access, archivists, and the future of archival research, *Australian Academic and Research Libraries* **42**(4): 309–319.

Meadows, D. H. (1999). *Leverage points. Places to intervene in a system*, The Sustainability Institute, Hartland.

Meadows, D. H. (2009). *Thinking in systems*, Earthscan, London.

Meadows, D. H. and Robinson, J. M. (1985). *The electronic oracle. Computer models and social decisions*, John Wiley & Sons, Inc., Chichester.

Menart, E. (2013). *Indoor pollution in archival collections in the context of a changing environment*, PhD thesis, University College London.

Menart, E., De Bruin, G. and Strlič, M. (2011). Dose–response functions for historic paper, *Polymer Degradation and Stability* **96**(12): 2029–2039.

Menart, E., De Bruin, G. and Strlič, M. (2014). Effects of NO_2 and acetic acid on the stability of historic paper, *Cellulose* **21**: 3701–3713.

Meyer, E. T., Eccles, K., Thelwall, M. and Madsen, C. (2009). Usage and impact study of JISC-funded. Phase 1 Digitisation projects & the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR), *Technical report*, University of Oxford.
URL: *http://repository.jisc.ac.uk/6997/1/TIDSR_FinalReport_20July2009.pdf*

Michalski, S. (2003). Double the life for each five-degree drop, more than double the life for each halving of relative humidity, *ICOM committee for conservation, 13th triennial meeting Rio de Janeiro preprints vol. 1*.

Michalski, S. (2018). Sharing conservation decisions. Tools, tactics, and ideas, *in* A. Heritage and J. Copithorne (eds), *Sharing conservation decisions. Current issues and future strategies*, ICCROM, Rome, pp. 183–204.

Michalski, S. and Karsten, I. (2018). The cost-effectiveness of preventive conservation actions, *Studies in Conservation* **63**(sup1): 187–194.

Michalski, S. and Pedersoli, J. L. J. (2016). *The ABC Method. A risk management approach to the preservation of cultural heritage*, Canadian Conservation Institute, Ottawa.
URL: *https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf*

MIDAS. Models of infectious disease agent study (n.d.).
URL: *midasnetwork.us*

Miller, L. K. (2013). All text considered: A perspective on mass digitizing and archival processing, *Source: The American Archivist* **76**(2): 521–541.

Mills, A. (2015). User impact on selection, digitization, and the development of digital special collections, *New Review of Academic Librarianship* **21**(2): 160–169.

Milstein, B., Homer, J., Consulting, H. and Hirsch, G. (2009). The “ HealthBound ” policy simulation game. An adventure in US health reform, *International System Dynamics Conference, July 26-30, 2009, Albuquerque, New Mexico*.

Milstein, B., Homer, J. and Hirsch, G. (2010). Analyzing national health reform strategies with a dynamic simulation model, *American Journal of Public Health* **100**(5): 811–820.

Morgan, J. S., Howick, S. and Belton, V. (2017). A toolkit of designs for mixing Discrete Event Simulation and System Dynamics, *European Journal of Operational Research* **257**(3): 907–918.

Morrisey, L. (2010). Data-driven decision making in electronic collection development, *Journal of Library Administration* **50**(3): 283–290.

Nadal, J. (2016). Comprehensive preservation environments: Site-wide resource management and conservation outcomes for recap, *Up, Up and Away: High Density Storage for Library Materials 10-11 August, 2016 Library of Congress, Washington, D.C.*, ReCAP: The Research Collections and Preservation Consortium, Princeton, NJ.

NARA (2018). 2018-2022 Strategic Plan, *Technical report*, National Archives and Records Administration (NARA).
URL: *https://www.archives.gov/files/about/plans-reports/strategic-plan/2018/strategic-plan-2018-2022.pdf*

NARA (2021). Climate action plan, *Technical report*, National Archives and Records Administration.
URL: *https://www.archives.gov/files/about/plans-reports/sustainability/climate-action-plan.pdf*

Nauta, G. J., van den Heuvel, W. and Teunisse, S. (2017). Europeana DSI 2-Access to digital resources of European heritage deliverable D4.4. Report on ENUMERATE Core Survey 4, *Technical report*, DEN Foundation.

NCDD (2017). Onderzoek naar de kosten digitale duurzaamheid, *Technical report*, Nationale Coalitie Digitale Duurzaamheid (NCDD).

Ness, R. B., Koopman, J. S. and Roberts, M. S. (2007). Causal system modeling in chronic disease epidemiology. A proposal, *Annals of Epidemiology* **17**(7): 564–568.

Nianogo, R. A. and Arah, O. A. (2015). Agent-based modeling of noncommunicable diseases: A systematic review, *American Journal of Public Health* **105**(3): e20–e31.

NVAO (2020). Besluit vervanging archiefbescheiden NVAO 2019.
URL: *https://wetten.overheid.nl/BWBR0043168/2020-02-14*

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V. and Alvin, K. F. (2002). Error and uncertainty in modeling and simulation, *Reliability Engineering and System Safety* **75**(3): 333–357.

O'Brien Miller, L. and Tedone, M. (2011). Archives conservation discussion group 2011. Digitization and its effect on conservation treatment decisions. How has wide-spread digitizing of collections changed our approach to treatment?, *The Book and Paper Group Annual* **30**: 115–127.

Oliver, G. (2011). The digital archive, *in* L. Hugues (ed.), *Evaluating and Measuring the Value, Use and Impact of Digital Collections*, Facet, chapter 4.

Ooghe, B. and Moreels, D. (2009). Analysing selection for digitisation: Current practices and common incentives Analysing Selection for Digitisation Current Practices and Common Incentives, *D-Lib Magazine* **15**(9/10).

Oriola, M., Campo, G., Strlič, M., Csefalvayova, L., Odlyha, M. and Mozir, A. (2011). Non-destructive condition assessment of painting canvases using near infrared spectroscopy., *in* J. Bridgland (ed.), *ICOM Committee for Conservation, 16th Triennial Meeting, Lisbon, 19-23 September 2011. Preprints.*, Criterio - Produçào Gráfica, Lda,, Almada, Portugal, p. paper 1317.

Orr, S. (2021). Multivariable linear regression analysis of degree of polymerisation for groundwood paper, *Unpublished - personal communication* .

Padfield, T. (2006). The interaction of water vapour with paper, *Conservation Physics* .
URL: *http://www.conservationphysics.org/vapap/vapap.pdf*

Padfield, T. (2010). Conservation physics. Calculator for energy use in museums.
URL: *http://www.conservationphysics.org/atmcalc/energyusecalc.php*

PaperTreat (2012). Evaluation of mass de-acidification processes. Deliverable 11. Specification of costs. PaperTreat project, *Technical report*.

Parkinson, N. (2005). Digitisation in the UK– the case for a UK Framework, *Technical report*, JISC.
URL: *www.jisc.ac.uk/digitisation*

Pickford, C. (2002). Archives: A statistical overview, *Cultural Trends* **12**(48): 1–36.

Pidd, M. (2014). The ways forward : A personal view of system dynamics and discrete-event simulation, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, first edit edn, John Wiley & Sons, Ltd., Chichester, chapter 15, pp. 318–336.

Pilette, R. (2004). Mass Deacidification : preservation option for libraries, *IFLA Journal* **30**(1): 31–36.

Plsek, P. E. and Greenhalgh, T. (2001). The challenge of complexity in health care, *British Medical Journal* **323**(7313): 625–628.

Poli, R. (2013). A note on the difference between complicated and complex social systems, *Cadmus* **2**(1): 142–147.

Poole, N. (2010). The cost of digitising Europe's cultural heritage, *Technical report*, Collection Trust.
URL: *https://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf*

Porck, H. J. (2006). The Bookkeeper process and its application at the National Library of the Netherlands, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences -tomorrow's perspectives*, Swiss National Library, Bern, pp. 37–42.

Potthast, A. and Kyujin, A. (2017). Critical evaluation of approaches toward mass deacidification of paper by dispersed particles, *Cellulose* **24**: 323–332.

Preiser, R. (2019). Identifying general trends and patterns in complex systems research: An overview of theoretical and practical implications, *Systems Research and Behavioral Science* **36**: 706–714.

Prescott, A. and Hughes, L. (2018). Why do we digitize? The case for slow digitization, *Archive Journal*.
URL: *https://www.archivejournal.net/essays/why-do-we-digitize-the-case-for-slow-digitization/#footnote_20_7962*

Preston White, K. and Ingalls, R. (2016). The basics of simulation, *in* T. Roeder, P. Frazier, E. Szechtman, E. Zhou, T. Huschka and S. E. Chick (eds), *Proceedings of the 2016 Winter Simulation Conference*, pp. 38–52.

Pruyt, E. (2013). *Small system dynamics models for big issues*, TU Delft Library, Delft.

Radon, J. (2018). Energy use optimization in the building of National Library, *IOP Conference Series: Materials Science and Engineering* **415**(1).

Rahmandad, H. and Sterman, J. D. (2008). Heterogeneity and network structure in the dynamics of diffusion. Comparing agent-based and differential equation models, *Management Science* **54**(5): 998–1014.

Ramin, M., Blüher, A., Reist, M. and Waelchli, M. (2009). Paper De-acidification, *Journal of Paper Conservation* **10**: 17–25.

Reilly, J. (2013). Choosing standards and best practices for environmental design and operation, *Summit on the Museum Preservation Management*, Smithsonian Institute, pp. 25–28.

Reilly, J.and Nishimura, D. and Zin, E. (1995). *New tools for preservation assessing long-term environmental effects*, Commission on Preservation and Access, Washington, D.C.

Rhee, H. L. (2015). Reflections on archival user studies, *Reference and User Services Quarterly* **54**(4): 29–42.

Richardson, G. P. (1986). Problems with causal-loop diagrams, *System Dynamics Review* **2**(2): 158–170.

Richardson, G. P. (2011). Problems with causal-loop diagrams, *System Dynamics Review* **27**(3): 219–243.

Rikowski, R. (2011). Digital libraries and digitisation. An overvieuw and critique, *in* R. Rikowski (ed.), *Digitisation Perspectives*, Sense Publishers, Rotterdam, chapter 2, pp. 21–42.

Riley-Reid, T. D. (2015). The hidden cost of digitization – things to consider, *Collection Building* **34**(3): 89–93.

Roberts, J. (1996). *The chemistry of paper*, The Royal Society of Chemistry, Cambridge.

Robinson, S. (2014). Discrete-event simulation : A primer, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., Chichester, chapter 2, pp. 10–25.

Rotaru, K., Churilov, L. and Flitman, A. (2014). Philosophical positioning of discrete-event simulation and system dynamics as management science tools for process systems. A critical realist perspective, *in* S. Brailsford, L. Churilov and B. Dangerfield (eds), *Discret-event simulation and system dynamics for management decision making*, John Wiley & Sons, Ltd., Chichester, chapter 5, pp. 76–104.

Roungas, B., Meijer, S. and Verbraeck, A. (2017). A framework for simulation validation and verification method selection, *in* A. Ramezani and M. Williams, E. Bauer (eds), *Proceedings of the 9th International Conference on Advances in System Simulation, SIMUL 2017*, pp. 35–40.

Rouse, W. B. (2015). *Modeling and visualiztion of complex systems and enterprises. Explorations of physical, human, economic, and social phenomena*, John Wiley & Sons, Inc., New Jersey.

Salleh, S., Thokala, P., Brennan, A., Hughes, R. and Booth, A. (2017). Simulation modelling in healthcare: An umbrella review of systematic literature reviews, *PharmacoEconomics* **35**(9): 937–949.

Sargent, R. G. (2020). Verification and validation of simulation models: an advanced tutorial, *Proceedings of the 2020 Winter Simulation Conference*, pp. 16–29.

Sargent, R. G., Goldsman, D. M. and Yaacoub, T. (2016). A tutorial on the operational validation of simulation models, *Proceedings of the 2016 Winter Simulation Conference*, pp. 163–177.

Saverwyns, S., Sizaire, V. and Wouters, J. (2002). The acidity of paper: evaluation of methods to measure the pH of paper samples, *13th Triennial Meeting, Rio de Janeiro, 22-27 September 2002: preprints* pp. 628–634.

Schaffner, J., Snyder, F. and Supple, S. (2011). *Scan and deliver: Managing user-initiated digitization in special collections and archives*, OCLC Research, Dublin, OH.

Schieritz, N. and Milling, P. (2003). Modeling the forest or modeling the trees: A Comparison of system dynamics and agent-based simulation, *in* R. Eberlein, V. Diker, R. Langer and J. Rowe (eds), *Proceedings of the 21st international conference of the System Dynamics Society : July 20 - 24, 2003, New York City, USA*, pp. 1–15.
URL: *http://www.systemdynamics.org/conferences/2003/proceed/PAPERS/140.pdf*

Schlessinger, L. and Eddy, D. M. (2002). Archimedes. A new model for simulating health care systems — the mathematical formulation, *Journal of Biomedical Informatics* **35**: 37–50.

Scholl, H. J. J. (2001). Agent-based and system dynamics modeling. A call for cross study and joint research, *Proceedings of the 34th Hawaii International Conference on System Sciences*.

Senge, P. (1990). *The fifth discipline. The art and practice of learning organization*, Doubleday/Currency, New York.

Shannon, R. E. (1998). Introduction to the art and science of simulation, *Proceedings of the 1998 Winter Simulation Conference*, pp. 7–14.
URL: *http://cecs.wright.edu/∼fciarall/ISE195/Readings/ShannonSimulationART.pdf*

Shenton, H. (2003). Life cycle collection management, *LIBER Quarterly* **13**(3-4): 254–272.

Shenton, H. (2006). Strategies for mass preservation treatment, *in* A. Blüher and G. Grossenbacher (eds), *Save paper! Mass deacidification: today's experiences - tomorrow's perspectives*, Swiss National Library, Bern, pp. 63–74.

Shepherd, E. (2009). *Archives and archivists in 20th century England*, Ashgate.

Smithsonian Institution (2013). Summit on the Museum Preservation Environment, *Technical report*, Smithsonian Institution.
URL: *https://repository.si.edu/bitstream/handle/10088/34611/13.03.EnvironPreservationSummit.Final.pdf?sequence=1&isAllowed=y*

Smulders, H. and Martens, M. (2014). Physics of Monuments.
URL: *http://www.monumenten.bwk.tue.nldefault.aspx*

Sobucki, W. (2003). Survey of the preservation status of the 19th and 20th Century collections at the National Library in Warsaw, *Restaurator* **24**(3): 189–201.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*, Mcgraw-Hill Education, Boston.

Sterman, J. D. (2006). Learning from evidence in a complex world, *American Journal of Public Health* **96**(3): 505–514.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and Bruin, G. D. (2015). Damage function for historic paper . Part I : Fitness for use, *Heritage Science* **3**(33).

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G.

(2015a). Damage function for historic paper. Part II: Wear and tear, *Heritage Science* **3**(1): 36.

Strlič, M., Grossi, C. M., Dillon, C., Bell, N., Fouseki, K., Brimblecombe, P., Menart, E., Ntanos, K., Lindsay, W., Thickett, D., France, F. and De Bruin, G. (2015b). Damage function for historic paper. Part III: Isochrones and demography of collections, *Heritage Science* **3**(1): 40.

Strlič, M. and Kolar, J. (eds) (2005). *Ageing and stabilisation of paper*, National and University Library, Ljubljana.

Strlič, M., Kolar, J., Koèar, D., Drnovšek, T., Šelih, V. S., Susiè, R. and Pihlar, B. (2004). What is the pH of alkaline paper?, *e-Preservation Science* **1**: 35–47.

Strlič, M., Kolar, J. and Lichtblau, D. (2007). The SurveNIR project—a dedicated near infrared instrument for paper characterization, *in* T. Padfield, K. B. Scientific, C. Mads, C. Christensen, B. Eshøj, M. Ryhl-Svendsen and D. Thickett (eds), *Museum microclimates*, National Museum of Denmark, Copenhagen, pp. 81–84.
URL: *http://www.conservationphysics.org/mm/musmic/musmic150.pdf*

Strlič, M., Kolar, J., Pihlar, B., Rychly, J. and Matisová-Ryvhlá, L. (2001). Initial degradation processes of cellulose at elevated temperatures revisited - chemiluminecence evidence, *Polymer degradation and stability* **72**: 157–162.

Strlič, M., Liu, Y., Lichtblau, D. A., De Bruin, G., Knight, B., Winther, T., Kralj Cigić, I. and Brereton, R. G. (2020). Development and mining of a database of historic European paper properties, *Cellulose* **27**(14): 8287–8299.

Strlič, M., Thickett, D., Taylor, J. and Cassar, M. (2013). Damage functions in heritage science, *Studies in Conservation* **58**(2): 80–88.

Świątkowska, B., Skoczeń-Rąpała, Ł., Okrągła, D., Jędrychowski, M. and Czop, J. (2018). Chemical degradation and physical failure: Risk analysis for a paper collection, *Studies in Conservation* **63**(sup1): 251–256.

Swinerd, C. and McNaught, K. R. (2012). Design classes for hybrid simulations involving agent-based and system dynamics models, *Simulation modelling practice and theory* **25**: 118–133.

Tanner, S. and Deegan, M. (2011). Inspiring research, inspiring scholarship. The value and benefits of digitised resources for learning, teaching, research and enjoyment, *Technical report*, JISC and KDCL, London.
URL: *https://www.kdl.kcl.ac.uk/fileadmin/documents/Inspiring_Research_Inspiring_Scholarship_2011_SimonTanner.pdf*

Taylor, J. (2005). An integrated approach to risk assessments and condition surveys, *Journal of the American Institute for Conservation* **44**(2): 127–141.

Terras, M. M. (2011). The rise of digitization. An Overview, *in* R. Rikowski (ed.), *Digitisation Perspectives*, Sense Publishers, Rotterdam, chapter 1, pp. 3–20.

Tétreault, J., Bégin, P., Paris-Lacombe, S. and Dupont, A. L. (2019). Modelling considerations for the degradation of cellulosic paper, *Cellulose* **26**(3): 2013–2033.

Tétreault, J., Dupont, A. L., Bégin, P. and Paris, S. (2013). The impact of volatile compounds released by paper on cellulose degradation in ambient hygrothermal conditions, *Polymer Degradation and Stability* **98**(9): 1827–1837.

Thylstrup Bonde, N. (2018). *The Politics of mass digitization*, The MIT Press, Cambridge, MA.

Trafela, T., Strlič, M., Kolar, J., Lichtblau, D. A., Anders, M., Mencigar, D. P. and Pihlar, B. (2007). Non-destructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation, *Analytical Chemistry* **79**: 6319–6323.

Tucker, S. (2007). Doors opening wider: Library and archival services to family history, *Archivaria* **62**: 128–158.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty. Heuristics and biases, *Science* **New Series, 185**(4157): 1124–1131.

Twinn, C. (1997). Passive control of relative humidity to ±5%, *UK, Chartered Institution of Building Services Engineers, 1997, proceedings of CIBSE National Conference held Alexandra Palace, London, UK, 5-7 October 1997*, Vol. 2, pp. 77–89.

Tyler, D. C., Falci, C., Melvin, J. C., Epp, M. L. and Kreps, A. M. (2013). Patron-driven acquisition and circulation at an academic library. Interaction effects and circulation performance of print books acquired via librarians' orders, approval plans, and patrons' interlibrary loan requests, *Collection Management* **38**(1): 3–32.

Valtysson, B. (2017). From policy to platform: the digitization of danish cultural heritage, *International Journal of Cultural Policy* **23**(5): 545–561.

van de Velde, M., Karelse, J., Kiens, P. and Windhorst, M. (2016). Handreiking vervanging archiefbescheiden, versie 2.0, *Technical report*, Groninger Archieven. **URL:** *https://www.nationaalarchief.nl/sites/default/files/field-file /handreiking_vervanging_archiefbescheiden_v2.0.pdf*

VanSnick, S. and Ntanos, K. (2018). On digitisation as a preservation measure, *Studies in Conservation* **63**(sup1): 282–287.

Varlamoff, M. (2004). The future, *in* J. Feather (ed.), *Managing preservation for libraries and archives*, Ashgate, Aldershot, chapter 8, pp. 159–174.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K. and Mazzoleni, S. (2011). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology, *Ecological Modelling* **222**(1): 210–218.

Vinther Hansen, B. and Vest, M. (2008). The lifetime of acid paper in the collection of the Royal Library, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, pp. 38–39.

Vizarova, K., Kischnerova, S., Suty, S., Tino, R. and Katuscak, S. (2008). Strengthening and deacidification of acidic groundwood paper with the ternary system chitosane-methyl-hydroxyethyl-cellulose- cationic starch in Mg(HCO3)2Aqueous solution, *in* M. Strlič and J. Kolar (eds), *Durability of paper and writing 2 : book of abstracts: 2nd international symposium and workshops, Ljubljana, Slovenia, July 5-7, 2008*, Faculty of Chemistry and Chemical Technology, Ljubjana, pp. 53–54.

Waller, R. (1994). Conservation risk assessment. A Strategy for managing resources for preventive conservation, *Preventive conservation: practice, theory and research. Preprints of the contributions to the Ottawa Congress, 12-16 September 1994*, IIC, Los Angeles, pp. 12–16.

Waller, R. (1995). Risk management applied to preventive conservation, *in* C. Rose, C. Hawks and H. Genoways (eds), *Storage of natural history collections. A preventive conservation approach*, Society for the Preservation of Natural History Collections, Iowa, pp. 21–27.

Waller, R. (2003). Cultural property risk analysis model. Development and application to preventive conservation at the Canadian Museum Of Nature, *Goteborg Studies in Conservation* **13**.

Weimer, C. W., Miller, J. O. and Hill, R. R. (2016). Agent-based modeling: an introduction and primer, *in* T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka and S. E. Chick (eds), *Proceedings of the 2016 Winter Simulation Conference*, pp. 65–79.

Whitmore, P. M. and Bogaard, J. (1994). Determination of the cellulose scission route in the hydrolytic and oxidative degradation of paper, *Restaurator* **15**(1): 26–45.

Whitner, R. and Balci, O. (1989). Guidelines for selecting and using simulation model verification techniques, *in* E. MacNair, K. Musselman and P. Heidelberger (eds), *Proceedings of the 1989 Winter Simulation Conference*, pp. 559–568.

Wierzbicki, A. P. (2007). Modelling as a way of organising knowledge, *European Journal of Operational Research* **176**(1): 610–635.

Wilson, H. and VanSnick, S. (2017). The effectiveness of dust mitigation and cleaning strategies at The National Archives, UK, *Journal of Cultural Heritage* **24**: 100–107.

Yuan, W., Van Ballegooie, M. and Robertson, J. L. (2018). Ebooks versus print books: Format preferences in an academic library, *Collection Management* **43**(1): 28–48.

Zervos, S. (2010). Natural accelerated ageing of cellulose and paper: A literature review, *in* A. Lejeune and T. Deprez (eds), *Cellulose: Structure and Properties, Derivatives and Industrial Uses*, Nova Science Publishers, Inc., New York, chapter 5, pp. 155–204.

Zervos, S. and Alexopoulou, I. (2015). Paper conservation methods. A literature review, *Cellulose* **22**: 2859–2897.

Zou, X., Uesaka, T. and Gurnagul, N. (1996). Prediction of paper permanence by accelerated aging. Part I: Kinetic analysis of the aging process, *Cellulose* **3**: 243–267.