

UNIVERSITY OF SZEGED  
DOCTORAL SCHOOL OF EDUCATIONAL SCIENCE

PHD PROGRAMME FOR  
INFORMATION AND COMMUNICATION TECHNOLOGIES IN EDUCATION

Ingo Barkow

**THE CHALLENGES OF METADATA MANAGEMENT  
IN COMPUTER-BASED SURVEYS AND ASSESSMENTS**

Dissertation

Supervisor:  
Prof. Dr. Ben Csapó  
Institute of Education  
University of Szeged



Szeged  
2016

## Table of contents

Table index .....	5
Figure index .....	7
Acknowledgements .....	9
Introduction .....	10
1. Theoretical Background .....	12
1.1 Definitions and scope of the dissertation.....	12
1.1.1 Metadata .....	12
1.1.2 Metadata standard.....	14
1.1.3 Metadata standards in educational sciences vs. social sciences .....	14
1.1.4 Research Data Centre (RDC) / Data Archive .....	15
1.1.5 Computer-based assessment vs. Technology-based assessment .....	16
1.2 Literature Review .....	16
1.2.1 Research literature about generic models describing metadata management.....	17
1.2.1.1 The Generic Statistical Business Process Model (GSBPM).....	17
1.2.2 Research literature about metadata or metadata standards in the educational sciences	25
1.2.3 Research literature about metadata or metadata standards in the social sciences .....	32
1.2.4 Different metadata standards and their capabilities.....	40
1.3 Description of the current situation regarding metadata and data in educational sciences and social sciences .....	42
1.3.1 Overview of the survey and assessment cycle.....	42
1.3.2 The demands of Data Management for computer-based assessment studies .....	44
1.3.3 Examples for Data Management in large-scale studies.....	58
1.3.4 Different software packages in computer-based assessment and their capabilities for metadata management .....	69
1.3.5 Conclusion.....	76
1.3.6 Research questions .....	76
1.3.7 Hypotheses .....	77
2. Survey Design .....	79
2.1 Methodological considerations in questionnaire design.....	79
2.1.1 Web Survey Design.....	80
2.1.2 Target group .....	85
2.1.3 Sampling.....	86
2.1.4 Time frame .....	87
2.2 Content of the questionnaire.....	87

2.2.1	Questions on “Your Research” .....	88
2.2.2	Questions on “Metadata Standards” .....	88
2.2.3	Question on “Data Sources” .....	89
2.2.4	Question on “Research Documentation” .....	89
2.2.5	Questions on Data Collection Instruments .....	90
2.2.6	Questions about Data Analysis.....	90
2.2.7	Questions about Paradata .....	91
2.2.8	Questions about Publication .....	91
2.2.9	Question about Software .....	92
2.2.10	Question on Organisation.....	92
2.2.11	Questions on Information about the Researcher’s background.....	92
2.2.12	Questions on Region.....	93
2.2.13	Questions on Personal Information .....	93
2.2.14	Questionnaire Flow Logic .....	93
2.3	Data Collection.....	94
2.3.1	Feedback from users during Main Study data collection .....	94
3.	Results from the first data collection.....	95
3.1.	Data management practices of the researchers.....	95
3.2.	Usage and satisfaction of metadata standards .....	96
3.3.	Usage and differences of data sources among scientists .....	99
3.4.	Practices of the research documentation .....	100
3.5.	Data documentation usage and differences of data producers.....	101
3.6.	Differences in dataset delivery between data users and producers .....	103
3.7.	Habits and differences of analysing paradata.....	106
3.8.	Background factors in publication processes .....	108
3.9.	Knowledge of the metadata standards’ software products .....	111
3.10.	Conclusions from the first data collection.....	112
4.	Results from the second data collection .....	114
4.1.	Improving sample sizes .....	114
4.2.	Intermediate results .....	115
4.3.	Intermediate conclusion.....	119
5.	Final data analytics on the combined dataset .....	120
5.1	Description of the combined dataset .....	120
5.2	Descriptive statistics on the combined dataset .....	121
5.3	Differences between groups .....	125
5.3.1	Academic status.....	126

5.3.2 Research area.....	128
5.4 Cluster analysis.....	132
5.5 Data Mining.....	138
6. Conclusion.....	142
6.1 Results of the analysis .....	142
6.1.1 Hypotheses revisited.....	142
6.1.2 Answering the research questions .....	143
6.1.3 Impact of the results on current metadata standards.....	145
6.1.3.1 Extending Data Documentation Initiative (DDI) with educational content .....	145
6.1.3.2 Extending Questionnaire and Test Interoperability (QTI) with content from lifecycle models	147
6.1.3.3 Developing a fully new metadata standard for computer-based assessment.....	147
6.2 Development of a model for software development based on the results .....	148
6.2.1 Combining metadata standards for educational research as an interim solution.....	148
6.2.2 Embedding data management personnel into the software development cycle .....	150
6.3 An outlook into the future .....	151
6.3.1 Data management and computer-based assessment .....	151
6.3.2 Recommendations for data management and metadata standards for the Institute for Educational Sciences at the University of Szeged.....	152
References .....	154
Appendices .....	162
I. Main Study Questionnaire in detail.....	162
II. Main Study Questionnaire Flow Logic .....	176
III. Invitation for participation in the study .....	187
IV. Mailing lists used for the data collections .....	188
V. Field Test Questionnaire.....	189
VI. Field Test Questionnaire Flow Logic .....	205
VII. Member list of the DDI Alliance.....	217
VIII. Main Test Questionnaire in re-usable metadata standards and questionnaire formats....	218

## Table index

TABLE 1. DEFINITION OF GSIM OBJECT „QUESTIONNAIRE”. SOURCE: GSIM, 2013, P. 181.....	21
TABLE 2. ATTRIBUTES OF GSIM OBJECT „QUESTIONNAIRE“. SOURCE: GSIM, 2013, P. 181.....	22
TABLE 3. DATACITE METADATA SCHEME SOURCE: DATACITE, 2013 P. 6. ....	39
TABLE 4. COMPARISON BETWEEN GLBPM PROCESSES AND DIFFERENT METADATA STANDARDS.....	40
TABLE 5. COMPARISON BETWEEN GLBPM’S ONGOING PROCESSES AND METADATA STANDARDS .....	41
TABLE 6. COHEN TABLE ON EFFECT STRENGTH .....	87
TABLE 7. FREQUENCY OF THE STORING THE RESEARCH DATA (%).....	96
TABLE 8. FREQUENCY (%) OF METADATA STANDARDS .....	97
TABLE 9. DISTRIBUTION OF THE STATEMENTS REGARDING METADATA STANDARDS.....	97
TABLE 10. SATISFACTION WITH DDI AND QTI (%).....	98
TABLE 11. CROSS TABULATION ON USAGE OF THE DDI STANDARD BETWEEN SOCIAL SCIENTISTS AND EDUCATIONAL SCIENTISTS .....	99
TABLE 12. RANKING ON DATA SOURCES (%).....	100
TABLE 13. RANKING AND FREQUENCY OF RESEARCH DOCUMENTATION (%).....	101
TABLE 14. METADATA PROVIDED WITH ITEMS FROM DATA PRODUCERS.....	102
TABLE 15. IMPORTANCE OF METADATA ON ITEMS FROM THE PERSPECTIVE OF DATA USERS (%).....	103
TABLE 16. METADATA AND PARADATA INFORMATION INCLUDED IN DATASET DELIVERY (DATA USERS’ PERSPECTIVE)	104
TABLE 17. FREQUENCY OF THE ACCESS TO DATASETS (%).....	105
TABLE 18. METADATA AND PARADATA INFORMATION INCLUDED IN DATASET DELIVERY (DATA PRODUCERS’ PERSPECTIVE) .....	106
TABLE 19. PREFERRED FORMATS FOR LOG FILE ANALYSIS.....	107
TABLE 20. CROSS TABULATION ON THE PREFERRED LOG FILE FORMAT – EXAMPLE OF CSV FILES (%).....	107
TABLE 21. FACTORS FOR CHOOSING A JOURNAL (%).....	108
TABLE 22. RECIPIENTS OF DATA .....	109
TABLE 23. REASONS TO NOT SHARE DATA (%).....	110
TABLE 24. MOTIVATING FACTORS TO SHARE DATA (%).....	110
TABLE 25. CROSS TABULATION ON WILLINGNESS FOR DATA SHARING – EXAMPLE OF DATA DELIVERY TO RESEARCH DATA CENTRES (%).....	111
TABLE 26. RANKING OF SOFTWARE PRODUCTS (%).....	112
TABLE 27. SAMPLE BY ACADEMIC STATUS – ONLY FIRST DATASET .....	113
TABLE 28. SAMPLE BY RESEARCH AREA – ALL DATASETS.....	113
TABLE 29. FREQUENCY (%) OF METADATA STANDARDS – SECOND DATA COLLECTION .....	115
TABLE 30. DISTRIBUTION OF THE STATEMENTS REGARDING METADATA STANDARDS – SECOND DATA COLLECTION ....	116
TABLE 31. FREQUENCIES OF DATA SOURCES (%) – SECOND DATA COLLECTION.....	116
TABLE 32. FREQUENCIES OF THE DIFFERENT METADATA PROVIDED (%) – SECOND DATA COLLECTION .....	117
TABLE 33. FREQUENCY OF THE ACCESS TO DATASETS (%) – SECOND DATA COLLECTION .....	118
TABLE 34. RECIPIENTS OF DATA (%) – SECOND DATA COLLECTION .....	119
TABLE 35. REASONS TO NOT SHARE DATA (%) – SECOND DATA COLLECTION.....	119
TABLE 36. SAMPLE BY ACADEMIC STATUS – ALL DATASETS .....	121
TABLE 37. SAMPLE BY RESEARCH AREA – ALL DATASETS.....	121
TABLE 38. FREQUENCY (%) OF METADATA STANDARDS – COMBINED DATASET .....	121
TABLE 39. DISTRIBUTION OF THE STATEMENTS REGARDING METADATA STANDARDS – COMBINED DATASET.....	122
TABLE 40. FREQUENCIES OF DATA SOURCES (%) – COMBINED DATASET .....	122
TABLE 41. FREQUENCIES OF THE DIFFERENT METADATA PROVIDED (%) – COMBINED DATASET.....	123
TABLE 42. FREQUENCY OF THE ACCESS TO DATASETS (%) – COMBINED DATASET .....	124
TABLE 43. RECIPIENTS OF DATA (%) – COMBINED DATASET.....	125
TABLE 44. REASONS TO NOT SHARE DATA (%) – COMBINED DATASET .....	125
TABLE 45. AVERAGES AND STANDARD DEVIATIONS OF THE STATEMENTS REGARDING METADATA STANDARDS.....	126
TABLE 46. AVERAGES AND STANDARD DEVIATIONS OF THE DIFFERENT METADATA PROVIDED .....	127
TABLE 47. AVERAGES AND STANDARD DEVIATIONS OF THE ACCESS TO DATASETS .....	127
TABLE 48 - AVERAGES AND STANDARD DEVIATIONS OF USAGE OF METADATA STANDARDS .....	128

<i>TABLE 49. USAGE OF DIFFERENT KINDS OF METADATA – KRUSKAL-WALLIS TEST .....</i>	128
<i>TABLE 50. USAGE OF DIFFERENT KINDS OF METADATA – MANN-WHITNEY TEST .....</i>	129
<i>TABLE 51 - AVERAGES AND STANDARD DEVIATIONS OF THE STATEMENTS REGARDING METADATA STANDARDS.....</i>	129
<i>TABLE 52. OPINION REGARDING METADATA STANDARDS – MANN-WHITNEY TEST .....</i>	129
<i>TABLE 53 - AVERAGES AND STANDARD DEVIATIONS OF DATA SOURCES .....</i>	130
<i>TABLE 54 - AVERAGES AND STANDARD DEVIATIONS OF THE DIFFERENT METADATA PROVIDED .....</i>	130
<i>TABLE 55. METADATA PROVIDED BY RESEARCHERS – KRUSKAL-WALLIS TEST.....</i>	131
<i>TABLE 56. METADATA PROVIDED BY RESEARCHERS – MANN-WHITNEY TEST.....</i>	131
<i>TABLE 57. AVERAGES AND STANDARD DEVIATIONS OF THE ACCESS TO DATASETS. ....</i>	132
<i>TABLE 58. ACCESS TO DATASETS – MANN-WHITNEY TEST .....</i>	132
<i>TABLE 59. VARIABLES WITH SIGNIFICANT DIFFERENCES BETWEEN CLUSTER 0 AND CLUSTER 1 .....</i>	141

## Figure index

FIGURE 1. THE GENERIC STATISTICAL BUSINESS PROCESS MODEL (GSBPM) .....	18
FIGURE 2. RELATION BETWEEN THE GENERIC STATISTICAL BUSINESS PROCESS MODEL (GSBPM) AND THE GENERIC STATISTICAL INFORMATION MODEL (GSIM) .....	20
FIGURE 3. SNIPPET OF GSIM QUESTIONNAIRE GROUP .....	21
FIGURE 4. THE GENERIC LONGITUDINAL BUSINESS PROCESS MODEL (GLBPM) .....	23
FIGURE 5. THE GENERIC LONGITUDINAL BUSINESS PROCESS MODEL (GLBPM) – CIRCLE VIEW .....	24
FIGURE 6. LEARNING TOOLS INTEROPERABILITY v2.0 (LTI) .....	27
FIGURE 7. THE ROLE OF ASSESSMENT IN QTI .....	28
FIGURE 8. THE APIP v1.0 ARCHITECTURAL MODEL .....	31
FIGURE 9. SNIPPET FROM THE SCORM METADATA MODEL .....	32
FIGURE 10. THE DDI3 COMBINED LIFECYCLE MODEL (VERSION 3.2) .....	34
FIGURE 11. THE SDMX INFORMATION MODEL (VERSION 2.1) .....	36
FIGURE 12. ELEMENTS OF DUBLIN CORE .....	37
FIGURE 13. STRUCTURE OF DATACITE v3.1 .....	39
FIGURE 14. ASSESSMENT CYCLE IN FORMATIVE ASSESSMENT .....	43
FIGURE 15. PIAAC ASSESSMENT DESIGN .....	44
FIGURE 16. GERMAN PIAAC DELIVERY SYSTEM .....	46
FIGURE 17. PIAAC QUESTIONNAIRE ITEM .....	46
FIGURE 18. PIAAC LITERACY ITEM .....	47
FIGURE 19. UK LOM CORE .....	49
FIGURE 20. PIAAC PROBLEM SOLVING ITEM .....	50
FIGURE 21. TAO BLACK LOG FILE FROM PIAAC .....	52
FIGURE 22. LOG FILE FROM CBA ITEM BUILDER EXECUTION ENVIRONMENT (CBA EE) .....	53
FIGURE 23. PEDOCS AT FACHPORTAL PÄDAGOGIK .....	57
FIGURE 24. PISA DATA EXPLORER .....	59
FIGURE 25. PIAAC INTERNATIONAL DATA EXPLORER .....	60
FIGURE 26. DDI LIFECYCLE v3.1 INSTANCE .....	62
FIGURE 27. NEPS STRUCTURE .....	63
FIGURE 28. MULTICOHORT SEQUENCE DESIGN .....	64
FIGURE 29. SNIPPET FROM THE NEPS EXPLORER TOOL .....	65
FIGURE 30. SNIPPET FROM DATABASE ON SCHOOL QUALITY (DAQS) .....	66
FIGURE 31. DAQS CONSTRUCT MODEL .....	67
FIGURE 32. CONCEPTUAL STRUCTURE OF TIPO .....	69
FIGURE 33. STRUCTURE OF TAO .....	70
FIGURE 35. MICROFIN ITEM IN CBA ITEM BUILDER .....	72
FIGURE 36. CONNECTING QUESTIFY TO TAO AND MOODLE .....	74
FIGURE 37. ITEM FROM EDIA .....	75
FIGURE 38. LIMESURVEY V2.05 MAIN ADMINISTRATION WINDOW .....	81
FIGURE 39. SNIPPET FROM LIMESURVEY TEMPLATE EDITOR .....	82
FIGURE 40. ITEM FROM THE DATA MANAGEMENT SURVEY FIELD TEST .....	84
FIGURE 41. ITEM FROM THE DATA MANAGEMENT SURVEY MAIN STUDY .....	85
FIGURE 42. FREQUENCY OF PROVIDING DATA MANAGEMENT PLANS (%) (N=87) .....	95
FIGURE 43. FREQUENCY OF DDI VERSIONS USAGE (%) (N=28) .....	99
FIGURE 44. SATISFACTION WITH METADATA INFORMATION FROM THIRD PARTIES (N=60) .....	104
FIGURE 45. DELIVERY TO EXTERNAL OR INTERNAL USERS (N=60) .....	106
FIGURE 46. SATISFACTION ON PARADATA DOCUMENTATION (N=32) .....	108
FIGURE 47. CLUSTER ANALYSIS OF THE USAGE OF METADATA STANDARDS .....	133
FIGURE 48. CLUSTER ANALYSIS OF STATEMENTS REGARDING METADATA STANDARDS .....	134
FIGURE 49. CLUSTER ANALYSIS OF DATA SOURCES .....	135
FIGURE 50. CLUSTER ANALYSIS OF PROVIDED DIFFERENT METADATA .....	136
FIGURE 51. CLUSTER ANALYSIS OF DATA PRODUCTS .....	137

<i>FIGURE 52. CLUSTER ANALYSIS OF THE REASONS NOT TO SHARE DATA .....</i>	<i>138</i>
<i>FIGURE 53. SCATTER PLOT OF PLANED CLUSTERS .....</i>	<i>140</i>
<i>FIGURE 54. THE MODULARIZED DDI MODEL .....</i>	<i>146</i>
<i>FIGURE 55. HYBRID METADATA MODEL FOR COMPUTER-BASED ASSESSMENT .....</i>	<i>149</i>



## Acknowledgements

Before the actual content of the dissertation starts I first want to thank some persons who helped in finishing it in a prominent position at the very beginning. The dissertation was written between September 1st, 2014 and January 30th, 2015 during a research semester where I left my former data manager position at the German Institute for International Educational Research (DIPF) in Frankfurt am Main to stay as a guest researcher at the University of Szeged for this period. Of course this caused extra work for many of my colleagues and friends.

Therefore I would like to thank:

- My wife Eszter Barkow for supporting my decision to go abroad for half a year with a lot of patience. Not every spouse would have supported this absence.
- My PhD supervisor Ben Csapó for having the idea of me joining the Doctoral School for Education as first international student of the program. Obviously as the dissertation proves it worked out.
- Timea Török who supported me with a lot of hints in data analysis (e.g. proper usage of the Kruskal-Wallis H Test) as this is not my original field of expertise (sorry for all the annoying questions).
- Same goes for István Hegedüs who supported me equally in the parts about data mining.
- My proofreaders Heiko Rölke, Attila Pásztor, Anita Pásztor-Kovacs, David Schiller and Istvan „Jerry” Thékes for providing important feedback and corrections.
- Natalja Menold, Jette Schröder, David Schiller, Alexia Meyermann, Marcel Hebing and David Tobinski for their valuable feedback on the several rounds of Field Test questionnaires until the final Main Study version.
- Heiko Rölke (again), Thorsten Busert, Helge Einspanier and Michael Dahlinger for taking over my project work at DIPF as substitutes.
- Marc Rittberger for supporting my research semester as head of the ICE department at DIPF.
- All colleagues at the TBA21 Group for their patience during the last six months.
- The teaching staff at the Doctoral School for Education at the University of Szeged who made it possible I could do many classes remotely from Germany during the first three years of the PhD process.
- The running track between Szeged and Tápé along the river Tisza for letting me keep my sanity during the stressful times of the writing process.
- The Serbian bakery across the street from my former flat in Szeged which supported this dissertation with huge amounts of „húsos burek” and „túrós burek” even at rather impossible times late at night. Unfortunately as of 2016 under new ownership with different sortiment and reduced opening hours.

*„We're entering a new world in which data may be more important than software.”*  
(Tim O'Reilly)

## **Introduction**

Data management in the educational sciences and especially in the area of assessment of students consisted for many years out of handling different kinds of paper material. To give an example from the home organization of the author. The whole basement of the German Institute of International Educational Research (DIPF) consists of a test archive where standardized cognitive tests from several decades are available. Scientists can visit the archive and look at items, instruments, scales and results in paper form. Data Management in this sense consists only of a register in which box which test from which decade can be found (which should not mean keeping this information is a trivial task). The situation nevertheless gets much more complicated with the introduction of computer-based testing into the Educational Sciences. Computer-based testing has a lot of advantages in comparison to paper&pencil testing procedures like automatic scoring and thus the possibility to develop adaptive tests or the creating of log files which can be analysed to get a richer picture about the participants in a test than a simple „true/false” scenario, but the processes around data management are much more challenging. A modern researcher is basically able to re-use an item from paper&pencil test from the 1950s by copying the material. There might be issues about the layout like old fonts and pictures from this time period which look awkward to the modern student and there are certainly political, cultural or methodological dimensions to be considered, but there is no technological boundary which avoids the usage of the material. Furthermore the information on the paper is sufficient to modernize the item. In computer-based testing we are facing other difficulties when we think of data management and archiving of instruments.

PISA 2009 contained an optional computer-based component called “Electronic Reading Assessment” (ERA) which consisted of items measuring digital reading. Technically the items were produced in 2007 as Adobe Flash items with a tool called Hypertext Builder to be rendered in a delivery platform called TAO 1.0. Though Hypertext Builder and TAO 1.0 use an abstract item model to produce the content it is nowadays in 2014 difficult to reproduce the items without additional programming effort. The Hypertext Builder has been replaced in the meantime by the sixth version of the CBA Itembuilder which uses the JAVA Eclipse framework to produce the items and TAO in the meantime also reached the third major release and nowadays bases on HTML5 technology. Both tools are in their normal version not backward compatible enough to handle items from their previous version which are more than five years old. The situation even looks grimmer if one considers most computer-based assessment tools do not use an abstract item model but the items are developed by programming them individually. In a nutshell this means if technology progresses and frameworks are discontinued those tests are lost as neither the hardware nor the software is running on a modern machine especially if one looks at timeframes of 20 or 30 years. A computer-based test from 1984 would most probably run on a Commodore 64 or Apple II and store its results and logs as sequential files on a Commodore 1541 or Apple Disk II 5.25” floppy drive. Aside from the fact 5.25” floppy disks start to lose their magnetized information after 10-15 years a test from this time period cannot be used seriously in a modern environment by means of emulation as the graphics and input mechanisms are not compatible to modern modes of testing. Therefore computer-based testing contains the risk of losing items and data due to technological changes if no data curation processes are in place.

The question therefore is how do we curate content from computer-based studies so they are also accessible for researcher who want to look at our items, instruments, log files, statistical scripts, data files and codebooks in several decades? Furthermore how far are researchers in the field of educational sciences willing to go to preserve their work? In the earlier description we have only looked at reproducing items, but essentially every step of computer-based assessment from study design, instrument design, item design, data collection, dissemination and publication produces metadata and data in different formats. Which of them are worth preserving as other researchers might need them for secondary research and which are obsolete? How should we handle them – simply store them into our data centre or have regular mechanisms to update them to newer technologies? Where are we currently standing with the processes? Are scientist in the educational sciences aware of these issues?

On the other hand policy makers are keen on pressuring researchers on re-using items or instruments as the costs for item development are significant. For similar reasons secondary analyses of research data are promoted as the willingness of stakeholders to fund individual data collections is shrinking as in many cases useful data from previous studies is available. Furthermore the advances in technology make more and more data available and give possibilities for merging external data e.g. from social media with results from surveys or assessments. In larger projects on international level researchers have to at least provide a plan what they are doing with the data after the study is finished. For the researchers this means they have to invest time into data documentation, but the question is do they really have the necessary understanding in reality and do policy makers check the quality of the processes. Are there structures on national and international level who can support researchers in data management and provide long term archival services? Are there proper standards for those organizations to handle the data and enable enhanced processes like comparison of results from similar studies?

These are the questions this dissertation will discuss. In the first part the current situation of data curation in the educational and social sciences will be analysed in form of a literature review. To get a clearer picture of the needs of data producers (i.e. researchers who conduct surveys or cognitive tests) as well as data users (i.e. researchers who perform secondary analyses of datasets from research data centres or data archives) a survey on data management was conducted from November 2014 to December 2014. The second chapter of the dissertation will show how the instrument was constructed and discuss the data collection process. In the third chapter the results from the data management survey will be analysed with different statistical methods. Lastly the results will be used in the fourth chapter to draw conclusions for the future of data management. In the last step it will also be possible to draw requirements for software development out of this process as the author is heavily involved in software design. For this reason the dissertation will close with some example objects to demonstrate how certain processes of educational assessment can be put into software code. A long-term goal of the research on data management in educational assessment should be to generate a generic model to facilitate all the necessary processes into one tool-chain.

# 1. Theoretical Background

## 1.1 Definitions and scope of the dissertation

This dissertation will focus mainly on processes surrounding computer-based assessment in the educational sciences and how the metadata, paradata and data derived from those surveys are handled in research data facilities or data archives (Leighton, 2002). As these terms carry quite a lot of importance in the topic to be discussed a definition is necessary especially as the concepts behind some of the terms are not trivial.

### 1.1.1 Metadata

Although the term metadata generally means “data about data” or “information about information” (NISO 2004) the understanding what this means differs from discipline to discipline. A computer scientist especially from the area of database administration would think of a repository which stores how table structures, indexes, database files and partitions are handled between the memory and the storage on the hard drive. To give two examples - in Microsoft SQL Server this repository would be the Master Database (Mistry & Misner, 2014) and in Oracle the same instance is called Data Dictionary (Bryla & Loney, 2008). Though this repository is very important for the technical storage of the data the information does not contain any value to a scientist who wants to do secondary analysis of variables derived from a study. Nevertheless for the computer scientist this information is vital for optimizing the *system in regards to performance, reaction time and memory consumption*. Bagley (1968) who coined the term “metadata” for the first time also only referred to the technical sense of organizing data storage as originally this term was only referring to programming concepts. The metadata information the scientist expects for a secondary analysis is rather a description of the variables, very often delivered in a codebook also containing further information about the study (e.g. waves, participants, data collection methods, weights). For the scientist therefore the metadata information is the description of the variables and the data information the files containing the results of the survey within a table while for the computer scientist both kinds of information are simply data within the data structure and the metadata information is the information how the system stored which values in which table and which file. The definition what is metadata and what is data therefore relies heavily on the purpose and the discipline where it is used. From the perspective of one researcher a certain information can be classified as metadata as it describes the next layer of data while another researcher already uses this information to do data analysis and thus classifies it as data.

Kimball (2008) defines three different types of metadata – technical metadata, business metadata and process metadata. Technical metadata in this sense describes like discussed before the structure of the data in the database system and has the function of defining the way how the information is technically stored. The information the scientist needs would be described as business metadata which describes to humans what the data in the storage mean, where they come from and what the relationship to other data in the same storage is. The last type Kimball classifies as process metadata is quite special to storage within a data warehouse and using methods from Business Intelligence to analyse the data. Here information about the data warehouse processes are stored (e.g. CPU time used, rows processes). Analysis of this data enables computer scientists to optimize query performance and to predict which data products might be interesting to users.

Though Kimball's definitions are very useful from the technical perspective they do not fit the purpose of enhancing understanding for scientific users as the term business metadata does not go far enough.

The National Information Standards Organisation (NISO, 2004) also divides metadata into three categories which completely differ as they are more related to archiving of data especially in the field of literature:

- Descriptive metadata describes a resources for discovery and identification (e.g. title, abstract, author)
- Structural metadata describes the composition of objects (e.g. chapters in books)
- Administrative metadata describes the management of a resource (e.g. file creation, file type)

Next to these categories NISO (2004) also lists two additional categories called rights management metadata and preservation metadata which describe subsets of legal considerations and archiving. This literature driven approach again does not fit very well to the subject discussed here. A more useful approach therefore could be looking at the survey processes used in educational sciences and derive metadata descriptions from that.

Here are some survey processes and their related metadata information:

- Information about the survey itself (e.g. name of the survey, involved researchers)
- Technical information about the storing of metadata and data into a repository (e.g. datasets in files or databases, tables, rows, indexes)
- Information about the items (e.g. schemes, questions, missings, stimuli)
- Information about the data collection (e.g. universe, time of collection, happenings during field work, disposition codes, log files)
- Technical information about the structure and rendering of items if the study is computer-based (e.g. rendering code, ontology based model for items, layout)
- Information about the cleaning processes at the data producer (e.g. building new variables, imputation of missing values)
- Information about the dissemination work of the researcher (e.g. creation of dissemination products, creation of new variables, weights)
- Information about archiving and curation of the data (e.g. data access, security, terms)

A researcher analysing log files from interviewer interactions e.g. to analyse behaviour in different types of households needs a different set of metadata and data than a researcher who wants to re-use one of the items of the study in an own context. For the dissertation and the related study it is therefore necessary to have a broader term and a model for different kinds of metadata derived from the survey process. Metadata will therefore be handled as "data describing the next data layer in the survey process".

On the other hand the term metadata was also heavily in the press in the last two years in regards to the National Security Agency (NSA) scandal around Edward Snowden (Reißmann, 2013). Here the term means the tracking of user information while people are using services in the internet (e.g. movement of persons while using the smartphone, which websites have been looked at for how long). While these processes strengthen the notion metadata can be used as an interesting data source itself and thus can be understood as data the dissertation will not be concerned with those issues as though they are very important to matters of privacy and data protection they do not fall into the scope of data management (e.g. Büttner, Hobohm, & Müller, 2011) for the educational sciences. This might change in the future,

but currently it is not state of the art and also conveys heavy ethical questions (e.g. should we monitor the web behaviour of children to predict their success in schools?) (see DFG, 2013).

### **1.1.2 Metadata standard**

Human understanding of data needs metainformation to make the content useful. This approach gets even broader if the information is supposed to be exchanged between different agencies. The process of harmonizing metadata and primary data if the original sources differ from each other is quite extensive and needs complicated import and export procedures. A lot of these procedures can be simplified if the data producer uses from the beginning a metadata standard to describe study information and instruments used in the survey. Metadata standard or standardized metadata means in this sense there is a body which specifies a set of relevant metadata information for a certain discipline to establish a common set of data description. These sets vary from discipline to discipline and also between research methods. There are sets especially for music notation (MEI - Music Encoding Initiative) (Hankinson, Roland, & Fujinaga, 2011) or law enforcement (NIEM – National Information Exchange Model) (NIEM, 2014) which normally are not useful for computer-based assessment in the educational sciences.

Nevertheless, a lack of standardization for computer-based assessment in education leads to several problems, e.g.

- the metadata cannot be processed in research data environments leading in the worst case scenario to data loss as the original data structures cannot be understood anymore
- the items are programmed in propriety formats therefore they cannot be ported from one vendor to another due to high transformation costs

Standardization of metadata for computer-based assessment is the first step for creating item banks, repositories and archives as this is the basis for exchange of storing fully functional computer-based items including statistical parameters. Therefore this work is highly relevant for public or commercial institutions who want to provide software for computer-based assessment as well as any organization who want to curate or archive computer-based educational studies.

The challenge for this dissertation is therefore to identify which metadata standards are valid for this domain. Furthermore also the term “standard” has to be defined. This dissertation will also look at metadata standards which have only the status of de-facto standards or conventions as even prominent and well-known metadata standards like Data Documentation Initiative Lifecycle (DDI, 2009) have not been certified by ISO, DIN, EU or another certification board though they are widely used by research data centres and data archives.

### **1.1.3 Metadata standards in educational sciences vs. social sciences**

As described before metadata standards are crucial for the re-use of instruments and secondary analysis of datasets. Unfortunately the number of metadata standards, which are suitable for the educational sciences (as will be shown in Chapter 2) is quite limited and might not be sufficient for the purpose. Therefore, the author will follow a definition of Punch (2009) and use high-level descriptions of the subject as well as literature from neighbouring disciplines to circle the problem field. Educational sciences in this sense can be defined as an applied social science so metadata standards from this discipline might be re-used for educational sciences as well.

#### 1.1.4 Research Data Centre (RDC) / Data Archive

As described before in the last years there was a strong movement from policy makers towards secondary usage of research data especially if it was generated by internationally renowned studies like Programme for the International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) or Programme for the International Assessment of Adult Competencies (PIAAC). In Germany the German Data Forum (“RatSWD – Rat für Sozial- und Wirtschaftsdaten”) was founded as a steering committee to help creating research infrastructures in all disciplines (e.g. social sciences, educational sciences, economics) for secondary analysis. These infrastructures are called „data service centers” (DSCs) (e.g. Schneider & Wolf, 2008) or „research data centres” (RDCs) or “data archives” (e.g. Solga & Wagner, 2007 or Bambey & Rittberger, 2013). The following information outlines the difference between those institutions through in reality the functions and terms are very often mixed and a lot of data centres fulfill multiple roles at the same time.

Data service centres are metainformation portals where researchers can find information about studies which are available at different institutions. Normally they do not store data on their own, but just provide enough metadata information the data user will be able to identify where a certain dataset is available and the terms of usage. The user is then usually forwarded to a research data centre where the actual data is available under certain terms and condition. A data service centre can therefore be understood as an interactive directory of research data centres and data archives. Unfortunately the number of DSCs in Germany is sinking. As of February 2015 there are only two remaining institutions (DSZ IZA in Bonn and DSZ-BO in Bielefeld).

Research data centres provide metadata, primary data and microdata in a secure environment (see Bender & Heining, 2011; Brandt & Schiller, 2011; Schiller, 2013). Scientists can visit as guest to analyse data for their own research. Sometimes they also offer remote access by terminal services, e.g. MONA at Statistics Sweden (Söderberg, 2005), NORC at the University of Chicago or remote calculations with a queueing system (e.g. the JoSuA system at IZA – Institute for the Study of Labor) to avoid travelling for the scientist. The long term goal is to directly connect the data used in a publication using Digital Object Identifiers (DOIs) or Uniform Resource Name (URNs) as permanent links so a researcher can jump from the text to the connected underlying database. Very often research data centres only offer data storage for a limited time period (e.g. 10 years) as they cannot guarantee the long-term preservation of data due to format changes. Important data is then forwarded to data archives. Policy makers like the European Commission (see EC, 2010 or EC, 2012) demand in their calls for proposals data management plans from researchers which specify data documentation as well as the final storage in research data facilities so the data can be used for secondary analyses.

Data archives are institutions which care for the long-term preservation of data using special archival formats for different kinds of media or preserving hardware components to make old data available for processing on demand. As this process is very cost-intensive data archives offer several different stages of preservation from enrichment and continuous post-processing to storing “as-is” (e.g. UK Data Archive, n.d.) (Bambey, Reinhold, & Rittberger, 2012; Huschka, Oellers, Ott, & Wagner, 2011).

The results of this dissertation are mainly directed to optimize the data management processes in educational sciences to support the workflow from the original researcher to the data archive. The focus is therefore on researchers designing surveys as well as on researchers working in research data centres and data archives who enrich studies for secondary research (Medjedovi , 2011).

### **1.1.5 Computer-based assessment vs. Technology-based assessment**

The dissertation focuses on computer-based assessment in educational sciences. Computer-based assessment describes the process of conducting a cognitive test in a domain like literacy, numeracy or problem solving on desktop computers and notebooks. Therefore examples will be derived mainly from international large scale assessments (e.g. PISA 2009, PIAAC 2011) or national school monitoring. As these tests are often combined in mixed mode with questionnaires (e.g. school, parent or headmaster questionnaire in PISA, background questionnaire in PIAAC) also questionnaires will be taken into account. Very simple computer-based assessment items and questionnaires essentially base on similar technologies (in the case of TAO in PIAAC both are rendered with the same tool). Questionnaires are a vital part of computer-based surveys and therefore will be handled as an additional topic in the text as well as in the accompanying survey.

Out of the scope of this dissertation falls Technology-based assessment (Reeff, 2007) which means assessment on a multitude of technical devices (e.g. mobiles, tablets or other devices) or new design paradigms (e.g. Serious Games) as though they represent the next stage of content where data management needs to happen. This process cannot be started when the foundation in form of computer-based assessment has not been researched yet. Nevertheless this represents more research for the future, but cannot be handled in the scope of a mere dissertation.

## **1.2 Literature Review**

If one considers the project title of the dissertation „the challenges of metadata management in computer-based assessment and surveys” it already contains four different topics which are combined in one very specific subject matter. As the focus is rather small there are not many people who worked on this specific subject which also leads to a problem in finding fitting literature. Basically the focus gets smaller by the following questions, where the sort order has been changed to reflect the underlying analysis process:

- What can be considered metadata in general?
- Which metadata are relevant for educational sciences?
- Which of these metadata are relevant for computer-based assessment?
- Can these metadata be standardized or combined?

There is literature about metadata in general, but already the second limitation – metadata for the educational sciences – limits the amount of literature significantly. Furthermore there is currently almost no scientific literature about standardized metadata in computer-based assessment. Therefore this work can be considered a new research area or a black spot which might lead to the question if these processes or questions are relevant or necessary. Currently in computer-based assessment though there is a high amount of money spent for item development for research studies, high-stakes testing or marketing surveys there is only little money for standardizing metadata or archival of items. This also reflects in the research work about this topic.

As mentioned before a focus of a literature review has therefore to be extended towards neighbouring disciplines like social sciences and data management in statistical archives.



### **1.2.1 Research literature about generic models describing metadata management**

To get a clearer picture which metadata is needed for which process it is valuable to look at different generic models describing the production process in survey organizations and data archives. The handling of statistical data and surveys in the field of social sciences is closely related to similar processes for the educational sciences.

#### **1.2.1.1 The Generic Statistical Business Process Model (GSBPM)**

A first holistic approach to describe the different archival processes is the Generic Statistical Business Process Model (GSBPM) developed by Vale (2010) for the United Nations Economic Commission for Europe (UNECE) and the Conference of European Statisticians Steering Group on Statistical Metadata. Vale (2010, p. 1) describes the purpose of the GSBPM as follows:

*“The original aim of the GSBPM was to provide a basis for statistical organisations to agree on standard terminology to aid their discussions on developing statistical metadata systems. It was conceived as a flexible tool to describe and define the set of business processes needed to produce official statistics. The GSBPM is, however, increasingly being used in other contexts such as harmonising statistical computing infrastructures, facilitating the sharing of software components, and providing a framework for process quality assessment and improvement.”*

The relationship between the goals of the GSBPM and the context of this dissertation is very close though the domains differ. The GSBPM was not developed for survey processes, but it analyses and describes the relationship of data management processes within a National Statistical Institution (NSI) like Destatis (Germany), Australian Bureau of Statistics (Australia) or Központi Statisztikai Hivatal (Hungary). The GSBPM gains its real power from the possibility of exchanging metadata between those organizations. The United Nations, the World Bank and larger statistical organizational bodies like Eurostat therefore endorse the further development of the model (current version: 5.0 as of December 2013). It is currently used by more than 50 statistical organizations worldwide. Figure 1

shows a high level view of the model.

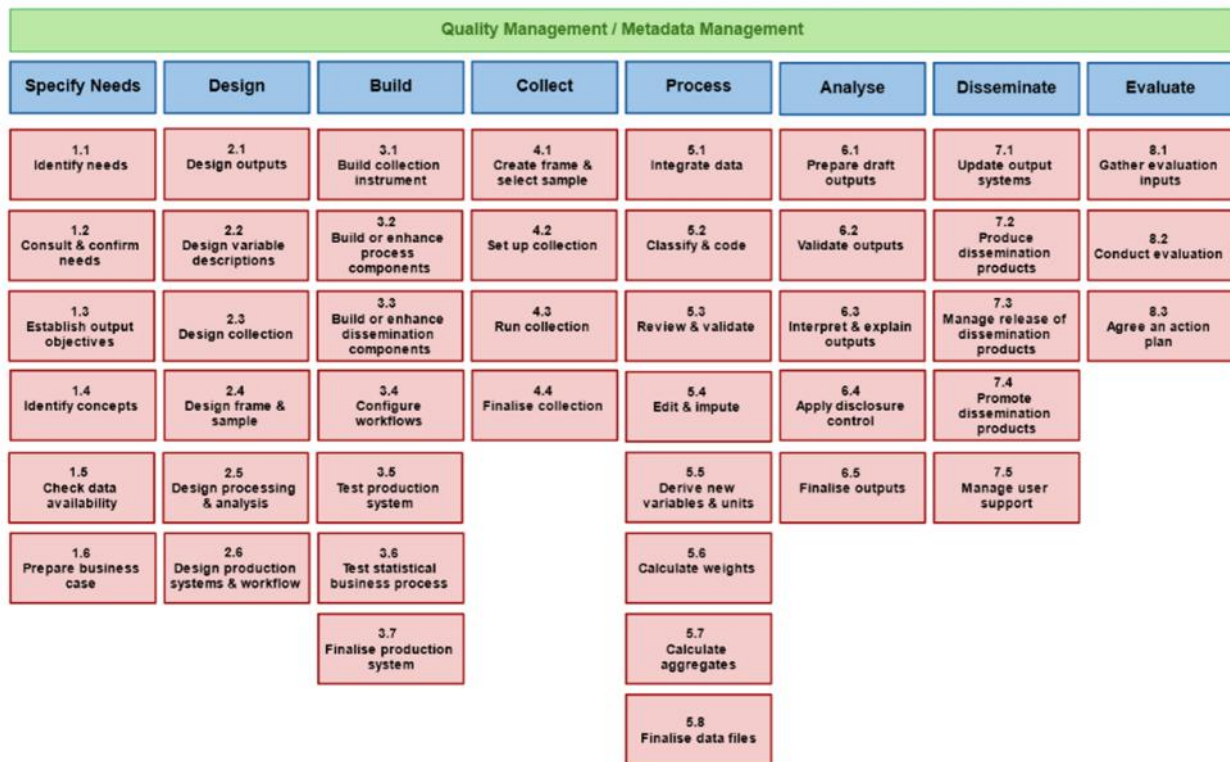


Figure 1. The Generic Statistical Business Process Model (GSBPM).  
Adapted from Vale, 2010 p. 3.

As can be seen from the graphics, the GSBPM contains eight high level processes from “Specify Needs” to “Evaluate” which are then defined into individual work packages to describe the individual data management task. Every task like “2.4 Design frame & sample” leads to metadata information and thus describes also the data production which derives from the process. The whole production process is accompanied by a continuous quality management and metadata management process. GSBPM does not specify any software tools or technical interfaces, but it enables the organizations using it to streamline their production processes as a first step of further collaboration. It is therefore a conceptual model, not a technical implementation blueprint.

Nevertheless, the GSBPM is an excellent example how powerful a common metadata understanding can be between different organizations. For the educational sciences and especially computer-based assessment it cannot be used in this form as it focuses too much on statistical production. The names of the individual steps like “3.1 Build collection instrument” sound promising, but focus on a completely different production process. In this case it means a generic description of questionnaire design. For computer-based assessment a lot of additional metadata information like layout, stimuli or scoring would be necessary even to represent a very basic cognitive item within a test. GSBPM describes this very high-level as can be seen here (GSBPM, 2013, p. 14):

*“3.1. Build collection instrument*

*58. This sub-process describes the activities to build the collection instruments to be used during the "Collect" phase. The collection instrument is generated or built based on the design specifications created during the "Design" phase. A collection may use one or more modes to*

*receive the data, e.g. personal or telephone interviews; paper, electronic or web questionnaires; SDMX hubs. Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets. This sub-process also includes preparing and testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to the statistical metadata system, so that metadata can be more easily captured in the collection phase. Connection of metadata and data at the point of capture can save work in later phases. Capturing the metrics of data collection (paradata) is also an important consideration in this sub-process.”*

The model gives recommendations and forces agencies to use a common vocabulary. Nevertheless, there is not technical definition how the instrument can be constructed. For the purpose of this dissertation the model is therefore too high-level and not specific enough. Also, this standard addresses primarily administrative datasets from statistical offices (Lane, 2009). Still it can serve as a conceptual basis for further design considerations. This has also happened as will be described in the next sections.

### **1.2.1.2 The Generic Statistical Information Model (GSIM)**

The huge success of the Generic Statistical Business Process Model (GSBPM) led the United Nations Economic Commission for Europe (UNECE) to further developments. As described in the section before, the GSBPM does not have any technical specifications, but serves rather as a process map or common vocabulary. As a next step UNECE developed an information model called the Generic Statistical Information Model (GSIM) to describe the tasks more specifically (GSIM, 2013). The model is designed to be complementary to GSBPM. Figure 2 displays the relation between the two standards:

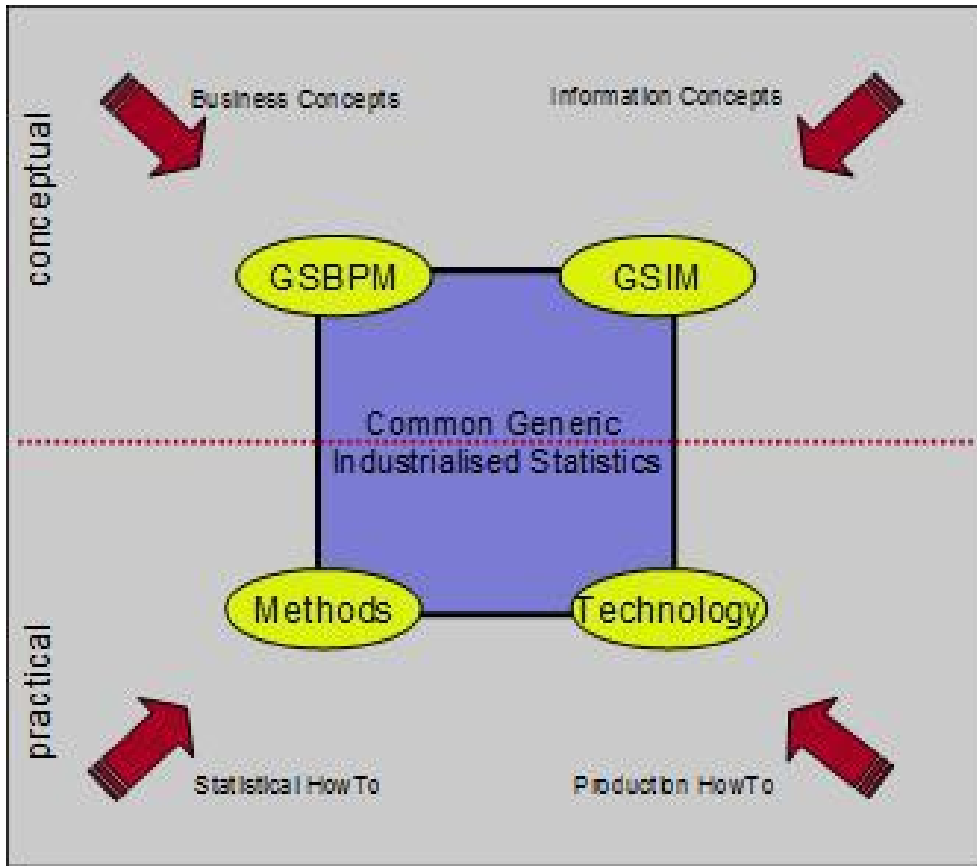


Figure 2. Relation between the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM).

Source: GSIM, 2013, p. 6.

GSIM tries to bridge between the business concepts of GSBPM and technological metadata standards like Data Documentation Initiative (DDI) and Statistical Data and Metadata Exchange (SDMX) which will be discussed in one of the following sections. GSIM offers abstract models (Base Groups, Business Groups, Concepts Groups, Exchange Groups and Structures Groups) which can be transformed by users into own entities. The model is too large to fit it onto a text page therefore the full model can be found in the appendix. Nevertheless, fitting to GSBPM 3.1 Build Collection Instrument Figure 3 shows a snippet from GSIM's user defined questionnaire group.

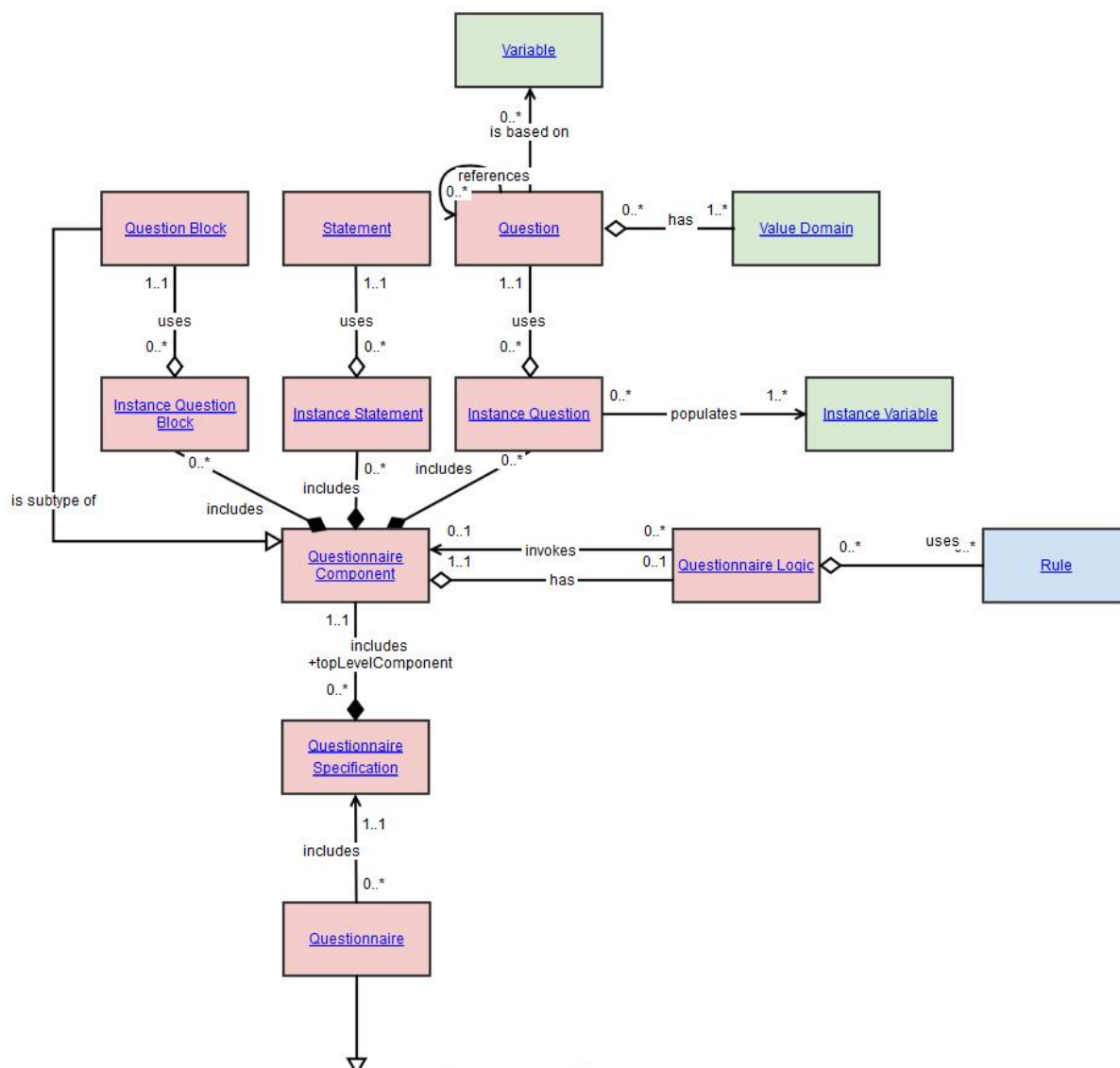


Figure 3. Snippet of GSIM Questionnaire Group.

Source: GSIM, 2013, p. 180.

The information from GSIM is not directly technical as it is not specified how these elements should be implemented into software code. Nevertheless they can be used as requirement analysis for the technical implementation. Furthermore each of the nodes like “Questionnaire” in the graph above contains a definition and sub-elements (GSIM, 2013).

Table 1. Definition of GSIM object „Questionnaire”. Source: GSIM, 2013, p. 181.

Object	Group	Definition	Explanatory Text	Synonyms
Questionnaire Exchange		A concrete and usable tool to elicit information from observation units.	This is an example of a way statistical organizations collect information (an exchange channel). Each mode should be interpreted as a new <i>Questionnaire</i> derived from the <i>Questionnaire Specification</i> .	

---

The *Questionnaire* is a subtype of *Exchange Channel*, as it is a way in which data is obtained.

---

Table 2. Attributes of GSIM object „Questionnaire“. Source: GSIM, 2013, p. 181.

Name	Description	Cardinality	Value Type
Name		1..1	text
Description	Background information about the questionnaire	0..1	text
DateIssued	Date the questionnaire was created.	1..1	date
DateReplacement	Planned date for the replacement of the questionnaire for other version. It can contain the date in which the new version was issued.	1..1	date
Media	Description of the kind of media conceived for the use of the Questionnaire (printed, electronic, etc.)	1..1	Text
Survey	Information on the survey which the questionnaire will be used by.	0..*	Text
SupportArtifacts	A list of devices, software programs, storage media, gadgets or other tools needed to support the use of the Questionnaire.	0..*	Text
DetailDocument	Reference to a document containing details of the implementation of the main elements of the Questionnaire	1..1	Link

As can be seen GSIM as an information model is a step towards a technical implementation. Currently it influences the further development of the metadata standards DDI and SDMX. Nevertheless the main focus of this model is still statistical production within National Statistical Institutions (NSIs). The model mainly describes a questionnaire as an abstract model for documentation or archival purposes. It does not describe an active survey questionnaire designed to go to the field. Furthermore there is not any focus on educational assessment. Nevertheless the model can be used as a blueprint for modelling a similar information standard for this domain as well as models like the GSBPM and GSIM do not exist for metadata in the educational sciences.

### 1.2.1.3 The Generic Longitudinal Business Process Model (GLBPM)

As discussed in the last section the processes of documentation for statistical production lead to the development of business and information models for defining data management processes. This work also influenced other domains like the social sciences. Inspired by the work of the GSBPM a modified version called the Generic Longitudinal Business Process Model (GLBPM) was developed by a group of scientist from different agencies at the DDI Alliance Dagstuhl Workshop in September 2011 (Barkow, Block, Greenfield, Gregory, Hebing, Hoyle and Zenk-Möltgen, 2012). Though the name stands first the author of this dissertation was part of this group, but not the main driver. Here Jay Greenfield from Boozé, Allen, Hamilton (BAH), Larry Hoyle from the University of Kansas and Arofan Gregory from Metadata Technologies Co. have to be rather mentioned.

The main difference between the GSBPM and the GLBPM is the latter model tries to specify the business processes in a longitudinal study in social sciences rather than showing the processes for storing

statistical products. From the pure model the overlap is significant in the horizontal view. Figure 4 shows the horizontal view of the GLBPM:

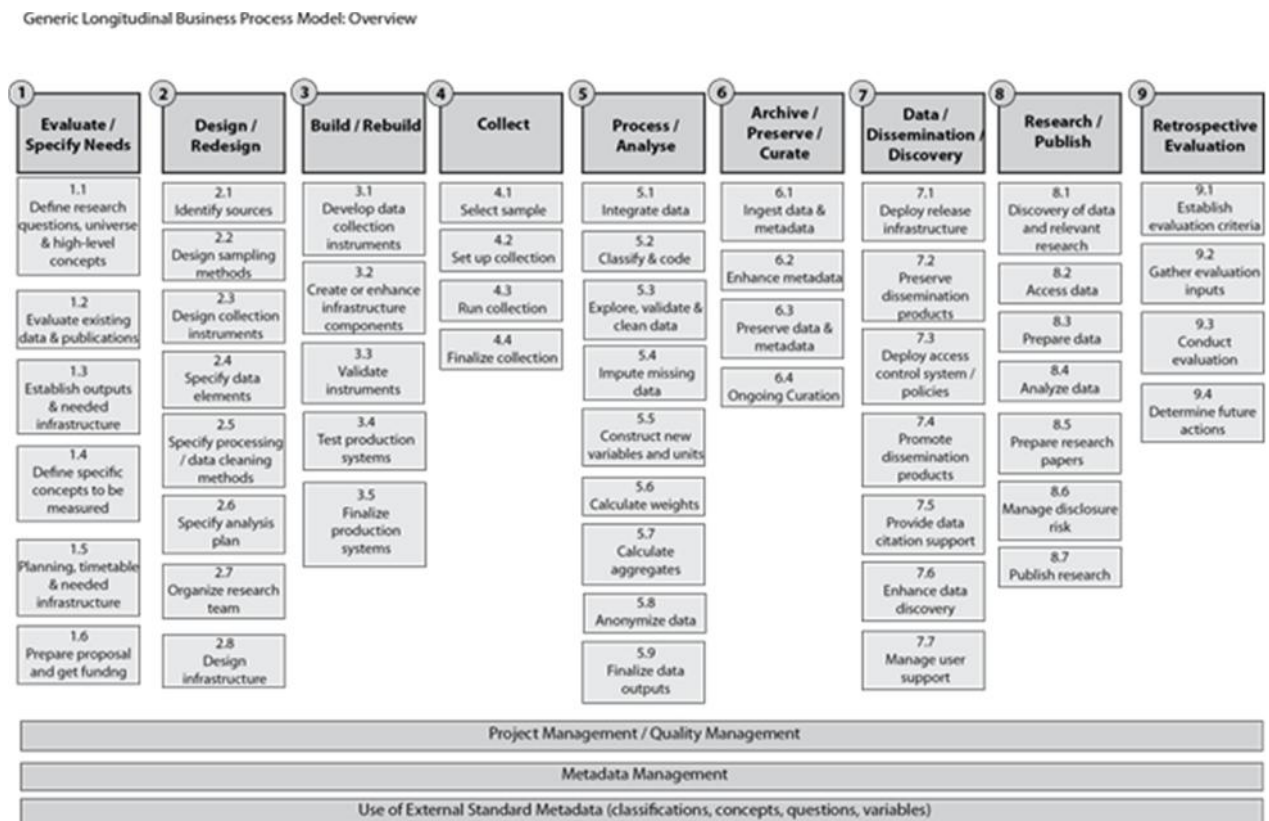


Figure 4. The Generic Longitudinal Business Process Model (GLBPM).  
Source: Barkow, et.al. 2012, p. 7.

The model shows the processes how to setup the instruments and repurpose them in every wave of a longitudinal study. Essentially the proceedings in a longitudinal study are a repeating cycle where the instruments are created before the first wave (processes „Design/Redesign”, „Build/Rebuild”), a data collection is performed (process „Collect”), the scientific dissemination processes are performed (processes „Process/Analyse”, „Archive/Preserve/Curate”, „Data / Dissemination / Discovery”, „Research / Publish” and finally the next wave is prepared (process „Retrospective Evaluation”). The important factor is all the processes are built for re-use which is represented in the terms (e.g. „Redesign” or „Rebuild” imply the next waves) or the permanent accompanying processes „Metadata Management”, „Project Management / Quality Management” and “Use of External Standard Metadata”.

The main difference in the GLBPM lies in the individual definitions of the processes as social sciences use a different terminology than National Statistical Institutes (NSIs). Furthermore there is a huge difference in the logic of the overall model which can be seen in the circle view. The GLBPM focuses heavily on longitudinal processes (though it can be also used to describe a cross-sectional study – there would only be a ‘wave 0’) therefore it specifies its processes in light of permanent repetitions like they happen in waves within a longitudinal study. This process can be seen in the circle view of the model.

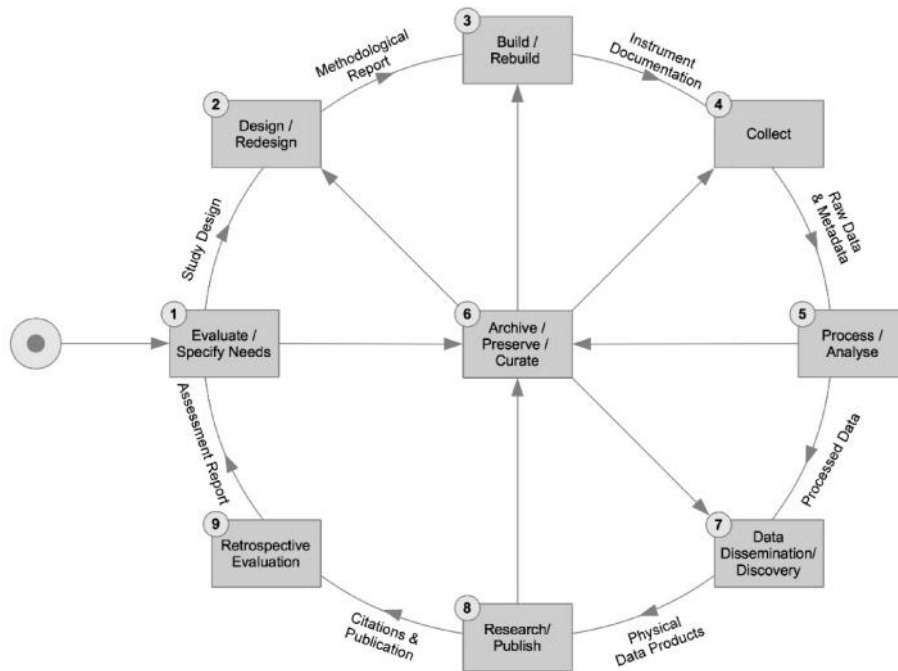


Figure 5. The Generic Longitudinal Business Process Model (GLBPM) – Circle View  
 Source: Barkow, et.al. 2012, p. 10.

Like GSBPM the GLBPM defines each of the sub-processes for its business logic. As an example here the processes “2.3 Design data collections instruments” and “3.1 Develop data collection instruments” from the main processes “Design / Redesign” and “Build / Rebuild” (Barkow, Block, Greenfield, Gregory, Hebing, Hoyle, & Zenk-Möltgen, 2012, p. 13).

### “2.3 DESIGN COLLECTION INSTRUMENTS

*Concepts to be measured may have been chosen, but in this step the specific form of the measurement tool is designed. For surveys this includes specifying categories and codes, the wording of questions, and the flow of the questionnaire. Other concepts may need to be measured through analysis of physical samples, as with biomarkers, or through the action of some mechanical or electronic instrument, as with measuring physical location. Custom sensors may need to be designed.*

*Data may also be obtained by coding something observed, either in real-time, or from recorded material. Observational settings may be structured, as in interviews or focus groups, or unstructured. Data may also be obtained through mining public online sources. This may involve software design.*

*Careful consideration should be given to any methods to improve data quality at the collection point rather than in post-processing.”*

(...)



### “3.1 DEVELOP DATA COLLECTION INSTRUMENTS

*Here the design developed in step 2.3 is actually implemented. This may involve the use of commercial or open source tools or involve custom programming (or hammering and soldering). Best practice is to use tools that develop the instrument automatically from the metadata developed in earlier steps. (see Iverson, Jeremy. Metadata-Driven Survey Design)”*

Here the differences to paragraph 3.1 of the GSBPM can be seen. GLBPM divides the processes more strictly and references to techniques used in the social sciences. It also connects the processes to external metadata standards like DDI.

Again, this model reflects the perspective of social sciences from the view of handling of longitudinal panel study using questionnaires. There are no processes regarding cognitive item construction or the specialities of handling computer-based assessment like it would be needed for a study like PISA or PIAAC. Therefore this model again has to be modified if it should fit the needs of the educational sciences.

For this reason the next section will give an overview of the situation of metadata standards in the educational sciences.

#### **1.2.2 Research literature about metadata or metadata standards in the educational sciences**

As described already in the first paragraph the number of literature on metadata or metadata standards in educational sciences is quite limited. There are the following existing models as of autumn 2014 – IMS Learning Object Model (LOM), IMS Learning Tools Interoperability (LTI), IMS Question & Test Interoperability Specification (QTI), IMS Accessible Portable Item Protocol (APIP) and Shareable Content Object Reference Model (SCORM).

##### **1.2.2.1 Learning Object Metadata (LOM)**

Learning Object Metadata (LOM) is an industry certified metadata standard developed by the IMS Global Learning Consortium and published as IEEE 1484.12.1 – Standard for Learning Object Metadata (2002) to document objects of learning management systems (see Harman & Koohang, 2007). It specifies resources to support learning processes.

The scheme consist of the following main categories:

- 1.) General (e.g. language)
- 2.) Lifecycle
- 3.) Meta-Metadata
- 4.) Technical (e.g. format)
- 5.) Educational (e.g. difficulty)
- 6.) Rights (e.g. copyright)
- 7.) Relation
- 8.) Annotation
- 9.) Classification (e.g. purpose)

LOM is constructed as a very high-level model to accommodate different object types from various sources. In the daily use the model is modified into application profiles where communities or institutions can specify which elements of LOM they will use, e.g. UK LOM Core – a specification for higher learning in the United Kingdom or LOM-CH which is a similar implementation for Switzerland. To make full use of the capabilities LOM has to be adapted to the purpose which has been done in a variety of different countries as the changes in educational structures have to be represented.

As can be seen in the generic model, LOM is able to define attributes like author, language, structure, methodology and also embed different kinds of content (e.g. file format, location, platform), but it is more directed into the area of describing educational content than item descriptions for computer-based assessment. The main application of LOM is to share content between different Learning Management Systems (LMS) like Moodle (Moodle, 2014) or Sakai (Berg & Korcuska, 2009).

### **1.2.2.2 Learning Tools Interoperability (LTI)**

Learning Tools Interoperability (LTI) by the IMS Global Consortium can be rather considered an interface standard than a metadata standard. The idea is to give different learning management systems like Moodle (Moodle, 2014) or Sakai (Berg & Korcuska, 2009) the possibility to embed course material into different environments (see Severance, Hanss, & Hardin, 2010). The learning modules can be used in different systems as plug-ins as long as the interfaces fit to the Service Oriented Architecture (SOA) of LTI (Murray, Blessing, & Ainsworth, 2003). Basically it would be possible to have a Learning Management Platform as the Tool Consumer and an Assessment System as Tool Provider. Currently different assessment platforms like ONYX by BPS Bildungsportal Sachsen (<https://www.bps-system.de/cms/produkte/onyx-testsuite/>), TAO by OAT s.a. (Ras, Swietlik, Plichart, & Latour, 2010) or Questify by Cito (<http://www.questify.eu/>) want to support LTI for this reason. Figure 6 describes how LTI handles different modules.

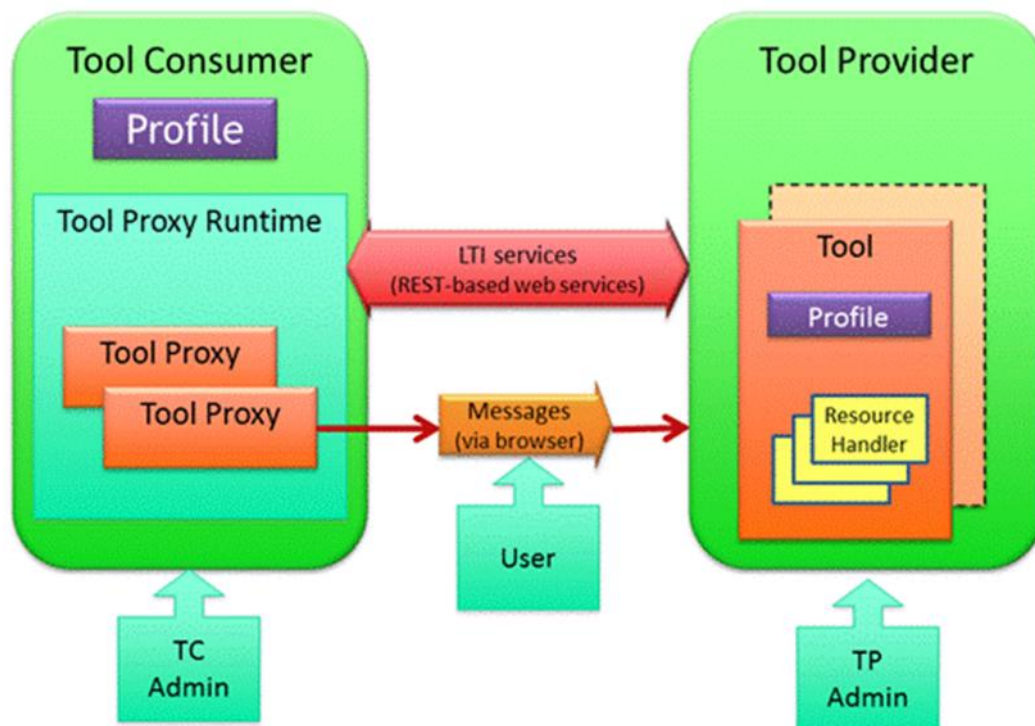


Figure 6. Learning Tools Interoperability v2.0 (LTI)

Source: IMS Global Learning Consortium, 2014, p. 1.

The advantage of LTI is the combination of different platforms by a common set of REST-ful web services. This technical implementation enables software developers to easily combine the systems. Nevertheless LTI is rather an interface specification than a metadata standard. The technical implementation manual describes the communication between two tools (e.g. a user which is registered in one system will be forwarded to the other system so the person does not have to login twice). If LTI is implemented between two assessment systems this does not mean the first system will be able to use items of the other system independently. It only means LTI makes it easier for System A to call functions from System B. Though this is a very useful aspect and strengthens collaboration between different software platforms it does not help in archiving content from educational sciences. A platform can be fully LTI-compliant with interfaces while the content of the platform is fully encrypted so the metadata about items and instruments cannot be extracted. Platforms using LTI also do not have to base on the same programming languages, e.g. TAO is a PHP application using HTML5 while Questify was developed on Microsofts .NET platform in C#.

Questify and TAO are therefore able to call functionalities from the other platform on run-time by using Learning Tools Interoperability (LTI) as an interface. Nevertheless they are also able to share items by using a third LMS Global standard called Question and Test Interoperability (QTI).

### 1.2.2.3 Question and Test Interoperability (QTI)

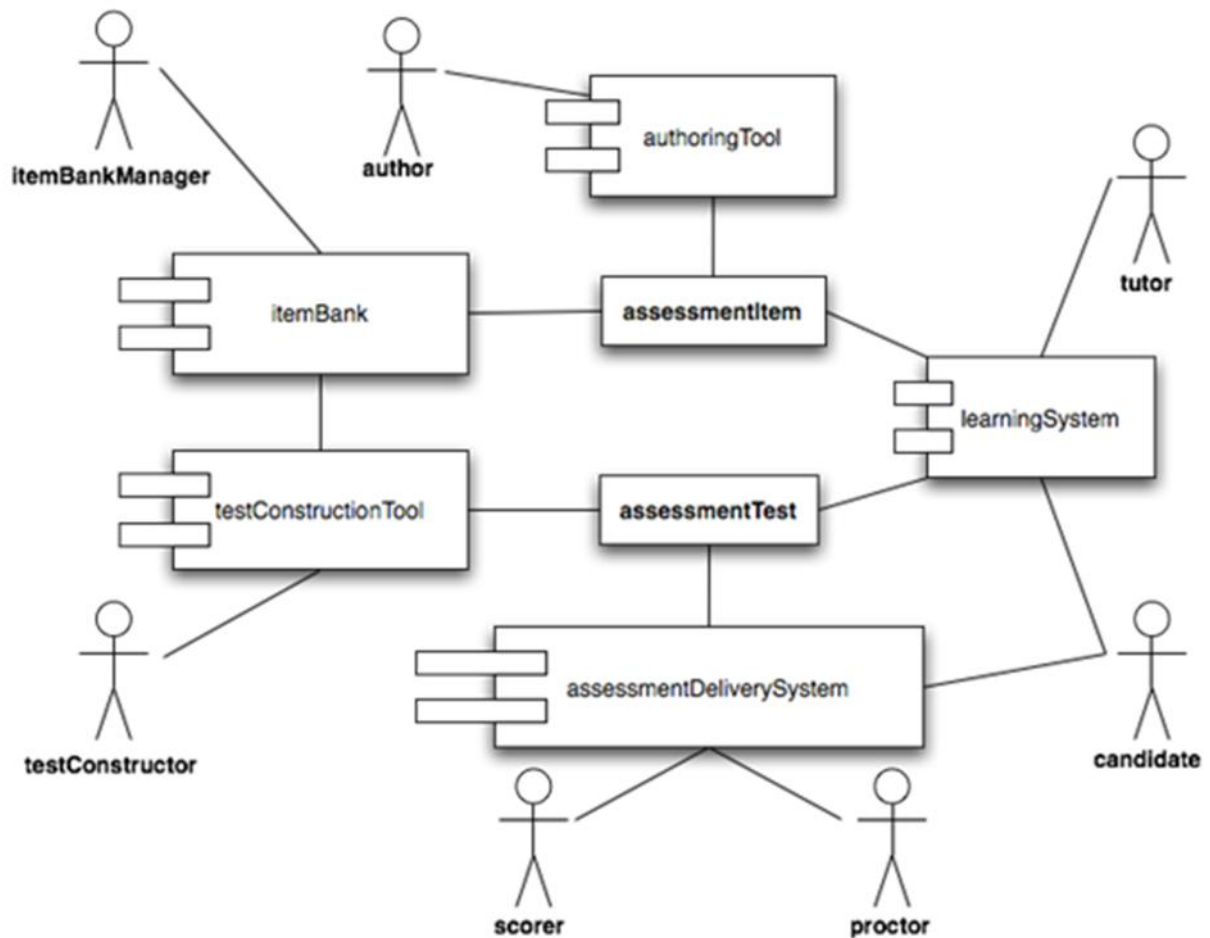


Figure 7. The Role of Assessment in QTI

Source: IMS Global Learning Consortium, 2006 p. 1.

Like LOM and LTI the Question & Test Interoperability Specification (QTI) is a product of the IMS Global Learning Consortium. This standard was specifically developed to specify content for computer-based assessment (see Smythe & Roberts, 2000). The goal was to create an abstract item description independent of technology so content can easily be moved from one computer platform to another without the need for any transformation (Helic, 2006).

As can be seen from the schematics QTI is able to describe the content of assessment items from the authoring tool via delivery to the item bank and therefore matches one of the requirements for a metadata standard for computer-based assessments.

Nevertheless there are drawbacks:

- QTI contains only a limited number of answer categories (e.g. multiple choice, open short text answer). It cannot handle complex scenarios for problem solving like in Csapó (2011) or Greiff and Funke (2010).
- QTI can only specify items and not the surrounding environment (it lacks study information and classifications like in LOM).

- QTI does not have any layout specifications or fixed frame formats. The recommendation from IMS Global is to use Cascading Style Sheets (CSS) for layout purposes.

Currently in Version 2.1 QTI supports the following interactions:

- ChoiceInteraction (multiple choice items with one or more correct answers)
- OrderInteraction (sorting answers in a list)
- AssociateInteraction (building pairs of answers)
- MatchInteraction (assigning correct responses to a matrix)
- GapMatchInteraction (matrix with open answers)
- InlineChoiceInteraction (choice of different answers within a text)
- TextEntryInteraction (open text boxes within different media)
- ExtendedTextInteraction (longer text box for free text passages)
- HotTextInteraction (choices of several text answers within a text)
- HotSpotInteraction (clickable areas in a graphic)
- SelectPointInteraction (marking an area in a graphic)
- GraphicOrderInteraction (marking hotspots with numbers)
- GraphicAssociateInteraction (building relations between hotspots in graphics)
- GraphicGapMatchInteraction (graphics with text boxes to enter data)
- PositionObjectInteraction (similar to HotSpot, but a predefined graphic can be used)
- SliderInteraction (offers a slider bar for a numeric answer)
- DrawingInteraction (offers a sketchboard for drawing)
- UploadInteraction (data created by another tool can be attached as an answer)

Additionally to that QTI has also specifications for scoring rules and supports external libraries like MathML. From an archive perspective QTI is currently the most used standard for computer-based items in educational assessment, but it lacks functionality which goes beyond this. Within QTI there is no handling of any metadata regarding survey information, instrument design, variable handling or context information. The special problem of layout handling will be discussed later in the comparison of different tools for computer-based assessment in Chapter 2.

The further development of QTI from V2.0 to V2.1 has proven to be difficult. QTI 2.1 was supposed to be released in 2009, but was recalled by IMS Global. The final release was published in September 2012 with a delay of three years. The IMS Global Learning Consortium currently tries to develop a new standard which combines elements of LOM, LTI and QTI called Accessible Portable Item Protocol (APIP). The development of the next version of QTI will be combined with the further development of APIP. Therefore it is worth to take a look at this upcoming standard.

#### **1.2.2.4 Accessible Portable Item Protocol (APIP)**

The Accessible Portable Item Protocol (APIP) by IMS Global can be considered to be a combination of three previous developments (IMS, 2012):

- Question and Text Interoperability (currently Version 2.1) – described in the previous section
- The Access for All Personal Needs and Preferences (PNP) – a standard for defining needs and preferences of individual students, especially in the area of accessibility

- IMS Content Packaging (CP) – CP collates the QTI content and accessibility information into a common exchange format

As can be seen APIP's main difference to QTI is its focus on all students including students with disabilities as assessment of these student groups with standard instruments can be challenging.

APIP therefore supports all the interactions of QTI, but adds the following accessibility profiles to it (IMS, 2014):

- Accessibility through Alternate Representations (Spoken Support, Braille, Tactile, Sign Language, Item Translation, Keyword Translation, Simplified Language, Alternative Representation)
- Accessibility through Adapted Presentations (Magnification, Color Adjustment)
- Accessibility through Adapted Interactions (Masking, Auditory, Calming, Additional Test Time, Breaks, Keyword Emphasis, Line Reader, Guidance)
- Accessibility through Adapted Response Modes (Calculator, Rule, Protractor, Reading Passage, Digital Materials, Physical Materials)

APIP specifies which tools are allowed and can be offered as an alternative to the individual needs of students. This can range from alternatives which can be offered to students with a certain disability to the simple question which calculator is allowed for a certain item. As computer-based assessment is supposed to support all test takers the work on APIP is quite important and fulfils in most countries also a legal requirement of including all students into testing. Figure 8 describes the structure of APIP v1.0.

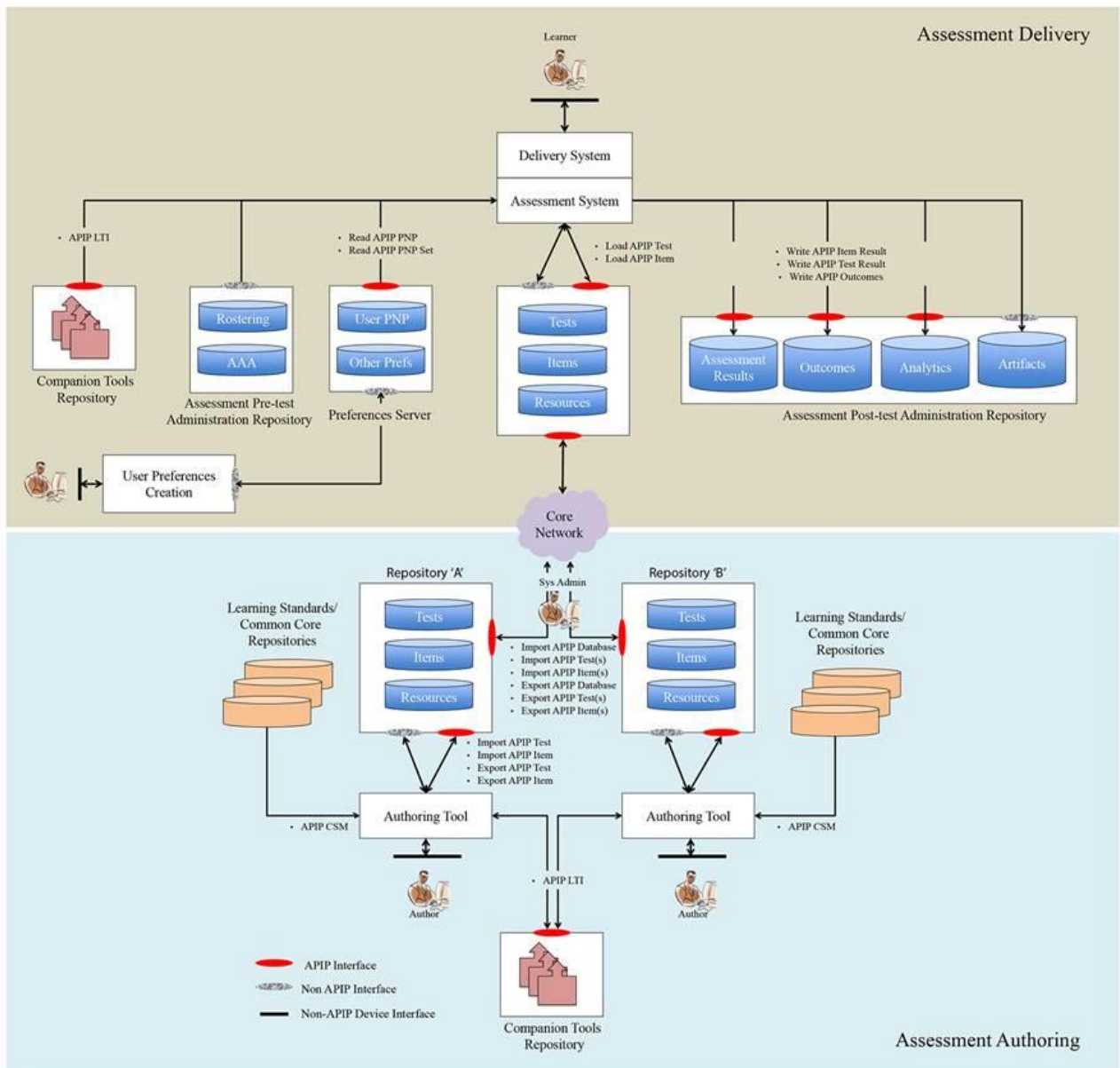


Figure 8. The APIP v1.0 architectural model

Source: IMS Global Learning Consortium, 2012, p. 1.

The benefits of APIP are the integration of different standards and its focus on inclusion. It furthermore seems the development of APIP triggered the further development of QTI which had been stagnating in the years before. Therefore this development certainly has to be welcomed. As of 2014 the first version of the standard became final, so APIP is available, but not widely used as the technical development has now to follow the creation of the standard. There is currently no multitude of tools like it exists with QTI. Nevertheless the compatibility with QTI might lead vendors to add APIP to their systems as well, especially as it solves some legal requirements they are facing with students with disabilities.

From an archival perspective APIP unfortunately opens more new questions than solving them as it like QTI focusses on content and not on archiving results, log files or item parameters for research data infrastructures. The focus on accessibility even adds another layer of complexity as those interactions also have to be stored, but APIP does not have any information on long-term preservation. Actually this

might have to be an addition for later versions of the standard to make it feasible for longer usage periods.

### 1.2.2.5 Shareable Content Object Reference Model (SCORM)

Shareable Content Object Reference Model (SCORM) is a set of interface definitions between different Learning Management Systems like Moodle (Moodle, 2014) or Sakai (Berg & Korcuska, 2009) specified by the Advanced Distributed Learning initiative by the Office of the United States Secretary of Defence. The current version is the fourth edition of SCORM 2004. Similar to LOM described in the previous section SCORM defines the content of learning objects to be exchanged between different platforms by specifying a data model and an application programming interface (API). The specification to define a container in SCORM is described in the SCORM Content Aggregation Model (CAM) (see Jesukiewicz, 2009). It also defines metadata to be used in SCORM-compliant packages. Figure 9 shows a snippet from the SCORM metadata model in XML (ibid, p. 168):

```
<manifest>
<metadata>
<schema>ADL SCORM</schema>
<schemaversion>2004 4th Edition</schemaversion>
<lom xmlns="http://ltsc.ieee.org/xsd/LOM">
<general />
<classification />
<annotation />
<lifeCycle />
<technical />
<metaMetadata />
<educational />
<relation />
<rights />
</lom>
</metadata>
</manifest>
```

Figure 9. Snippet from the SCORM metadata model

Source: Jesukiewicz, 2009, p. 168.

Interestingly SCORM links its metadata to LOM v1.0 which can be seen in the scheme reference (xmlns="http://ltsc.ieee.org/xsd/LOM") (Baldoni et al., 2004) and uses parts of the same model which unfortunately leads to the same problem of the data description for computer-based assessment. Like LOM SCORM was specified to define learning objects between e-learning platforms and therefore does not contain the necessary scheme to specify the workflow of an assessment or a survey (though written for an older version - see Bohl, Scheuhase, Sengler, & Winand, 2002). Nevertheless it might provide the interfaces to embed an assessment platform using internally a different standard into a given Learning Management platform.

### 1.2.3 Research literature about metadata or metadata standards in the social sciences

As it was mentioned in Section 1.1.3 the social sciences can be considered as neighbouring discipline of the educational sciences. Punch (2009) even calls educational sciences an applied social science.



Therefore it is worth considering metadata standards from this discipline as an example of how similar standards can be developed in the educational sciences, especially as the upcoming metadata standards focus on a slightly different content than the standards described in the last section.

### **1.2.3.1 Data Documentation Initiative (DDI)**

Currently two of the most prominent metadata standards in the social sciences are the Data Documentation Initiative (DDI) Codebook (current version 2.5) and DDI Lifecycle (current version: 3.2) standards (see DDI, 2009, Blank & Rasmussen, 2006 or Vardigan et. al., 2008LOM). Both standards are developed in parallel by the DDI Alliance which as of February 2015 consists of 38 member organizations (e.g. universities, research institutes, statistical institutions, software companies). The difference between DDI Codebook and DDI Lifecycle is the newer 3.2 standard focuses on the re-usability of instruments, items and schemes in form of a lifecycle model. DDI Codebook on the other hand focuses more on variable descriptions. As a lot of organizations represented their metadata in DDI Codebook, the older standard has to be kept alive for some time as DDI Lifecycle is not downward compatible to DDI Codebook. This means data has to be transformed to be compliant to the new standard which causes costs. DDI Codebook 2.1 is represented as a Document-Type-Definition (DTD) scheme, while DDI Codebook 2.5 and all versions of DDI Lifecycle are eXtended Markup Language (XML) schemes (XSD).

Furthermore the DDI Alliance has just started a process called “Next Generation DDI” to specify a DDI Lifecycle version (4.0) which does not use an XML representation as a standard anymore but a Resource Description Framework (RDF) ontology (Bosch, Cyganiak, Wackerow, & Zapilko, 2012). This has the advantage the model can be transformed into a variety of different formats like XML, Relational Database Systems (RDB) or analytical databases (Amin Barkow, Kramer, Schiller, & Williams, 2012) which allows data archives and research data centres more flexibility in implementing DDI into their own infrastructure. Figure 10 shows a model of DDI Lifecycle.

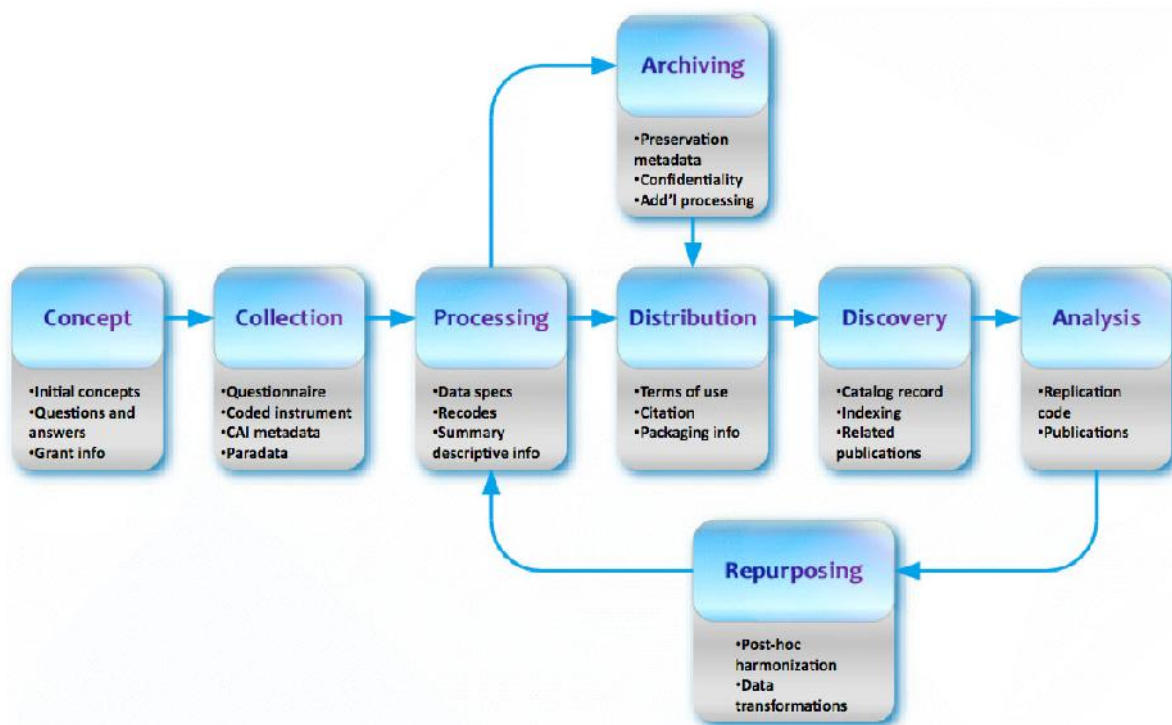


Figure 10. The DDI3 Combined Lifecycle Model (Version 3.2)

Source: DDI, 2014, p. 3.

As can be seen in Figure 10 a strong emphasis is put on „Repurposing“ and „Data Archiving“ which means metadata will be used in a standardized manner in several studies. The goal of the mechanism is to provide repositories with items, schemes and scales for questionnaire designers. A strong emphasis is furthermore put onto describing the “Study Concept” and the “Data Collection”. This feature is different from APIP mentioned in the paragraph before where these descriptions are lacking. Nevertheless DDI is still not sufficient for educational sciences as in its normal form it is designed to describe questionnaires for the social sciences. Unlike QTI or APIP it cannot handle graphical items and for a long time it also did not feature mechanisms to design a workflow in a survey (e.g. filtering, routing and dynamic text). Until Version 3.1 this could only be described in prose, but not in a native machine-readable format. Version 3.2 changed that, but currently it is in comparison to 3.1 not widely used as it was only released middle of 2014. Also the 3.2 release saw some improvements in regards to education assessment as some extended answer modes like QuestionGrid and QuestionBlock were implemented. Nevertheless these rather reflect the demands of paper&pencil educational items than computer-based tests.

DDI was originally developed to store a survey for preservation after it has finished for documentation purposes. This can be proven from the model as the important process of designing the questionnaire is hidden between the bullet points “Study Concept” and “Data Collection”. This notion has changed in 3.2 as more process-oriented tasks have been implemented, but DDI is still bound to its history as a documentation standard. It is in the meantime possible to create a working questionnaire which can be rendered in a computer-based platform, but if the processes are compared to Questionnaire Design

Languages used in systems like Blaise (Statistics Netherlands, 2005), MMIC (Weermann, Oriens, & Colvin, 2011) or Cases (see <http://cases.berkeley.edu/>) the possibilities are still quite limited as e.g. more complex routing and looping options like loops within loops cannot be modeled. Nevertheless this might change in future releases as well. As the development of the new ontology can be considered very time consuming (the step from DDI 3.1 to DDI 3.2 took four years and DDI 4.0 will be a major release with a new data model) it can be expected like in DDI Codebook there will be parallel releases like DDI 3.3 and DDI 3.4 while the work on DDI 4.0 continues.

Currently DDI can be considered only a “de facto” standard as it has not been certified by an international body like ISO, DIN, EN or IEEE. The DDI Alliance is planning to do this step with one of the later releases.

### **1.2.3.2 Statistical Data and Metadata Exchange**

Statistical Data and Metadata Exchange (SDMX) is an initiative of international financial and statistical institutions like the European Central Bank, the World Bank, the International Monetary Fund (IMF), Eurostat, Organisation for Economic Co-operation and Development (OECD), the United Nations Statistics Division and the United Nations Educational, Scientific and Cultural Organization (UNESCO). It is approved as an international standard by the International Standards Organization (ISO) (ISO 17369:2013) and uses messages in XML (SDMX-ML) or EDIFACT (SDMX-EDI) messages for communication between services. The current version as of 2014 is SDMX Version 2.1.

The main goal of SDMX is to enable statistical organizations to exchange datasets. This means the metadata are mainly structural and describe the content of the datasets (see Capadisli, 2015).

Gregory (2009) defines SDMX the following way:

- Describes the structure of aggregate/dimensional data (“structural metadata”)
- Provides formats for the dimensional data
- Provides a model of data reporting/collection and dissemination
- Provides a way of describing the structures of arbitrary metadata sets (“reference metadata”)
- Provides formats for the arbitrary metadata sets
- Provides a set of standard registry interfaces, providing a catalog of resources
- Provides guidelines for deploying standard web services for SDMX resources
- Provides a way of describing statistical processes

Figure 11 describes the information model of SDMX Version 2.1:

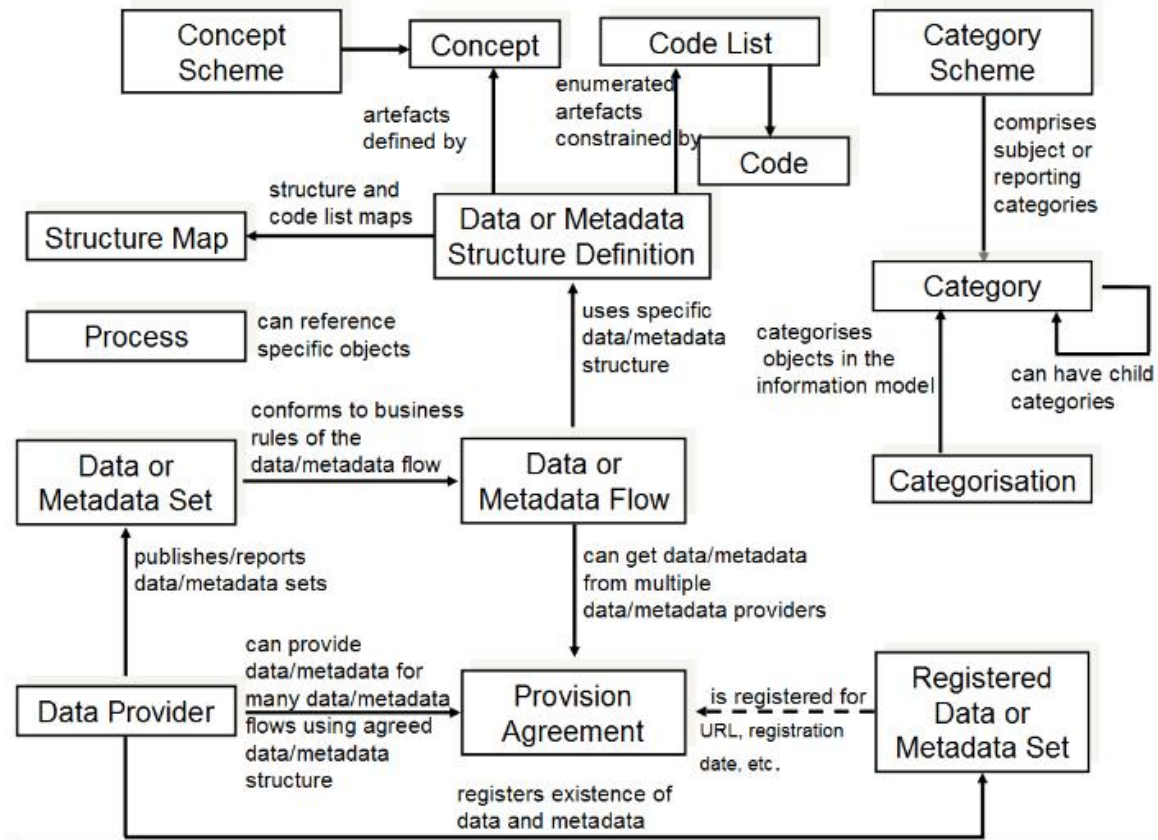


Figure 11. The SDMX Information Model (Version 2.1)

Source: SDMX, 2012, p. 17.

SDMX as shown in Figure 11 has quite a different approach than DDI. It focuses more on the distribution of datasets and their statistical metadata than on documenting surveys and instruments. SDMX can therefore be seen as complimentary to DDI. Nevertheless both base on a similar information model. The DDI Alliance and the SDMX representatives therefore started in 2010 a DDI-SDMX dialogue to describe how both standards can be used in addition. After five sessions there was no further progress reported after 2011.

From an educational sciences point of view SDMX might not be used directly as it cannot represent cognitive items, but its connection to DDI and focus on statistical reproduction make it an interesting business case for further developments.

### 1.2.3.3 Dublin Core

Dublin Core is an older but very popular metadata standard used to describe documents and other objects specified in 1994 by the Dublin Core Metadata Initiative (DCMI, 1994). It consists of 15 main elements and was standardized by the ISO as ISO 15836:2009. Due to its use for referencing objects on websites it also was handed in as Request for Comments (RFC) 2413 at the Internet Engineering Taskforce (IETF) where all current internet protocols are specified (Weibel, Kunze, Lagoze, & Wolf, 1998). Figure 12 presents a graph of the core properties of Dublin Core.

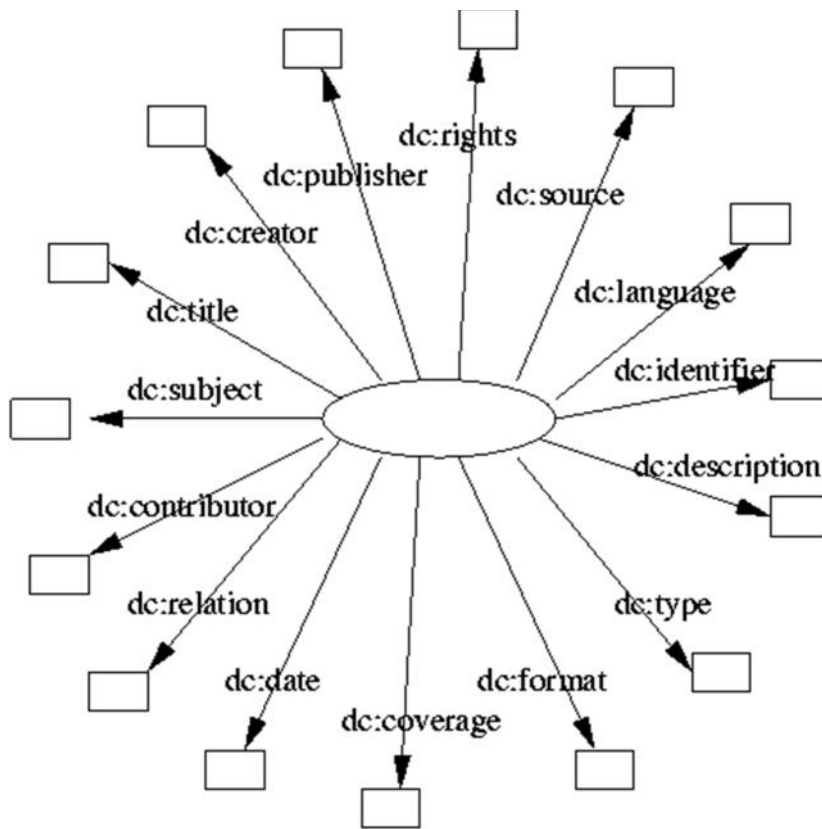


Figure 12. Elements of Dublin Core

Source: DCMI, 1994, p. 2.

Dublin Core is quite extensive used in library science as the core elements were developed to describe different types of media. Furthermore the standard has been in constant development since 1994 and therefore offers a lot of documentation and additional material for implementing documentations for different kinds of media (Baker, 2000). For example there is an extension for educational sciences in form of the property *dcterms:audience* with the sub-properties *dcterms:mediator* and *dcterms:EducationLevel* for describing learning materials.

In case of computer-based assessment Dublin Core could describe a resource like a book or web site containing cognitive tests but currently there are no elements for testing included in the properties. Dublin Core is actually lacking properties which describe the content of an object in detail, which actually could be the reason for its success. Nevertheless Dublin Core is a good example of the power of metadata standards if they are widely used.

Other standards like DDI or SDMX contain mappings to Dublin Core or even implemented the whole core set into their own standard.

### 1.2.3.4 DataCite Metadata Scheme

DataCite (2014) is an interesting mix of an organisation and a metadata scheme and logically related to Dublin Core. The idea behind DataCite is (as the name indicates) to make data citable like publications. This addresses a typical problem why data production is not that attractive to researchers. Currently the only measure of scientific success is publications of articles in peer-reviewed journals with an impact factor (see Allen, Brand, Scott, Altman, & Hlava, 2014). If a researcher is mainly working in data collection this means he or she will normally not get any scientific recognition until they publish their results in papers (see Smit, 2011 or Siegert, 2010). Especially in larger surveys this might not be possible as the project management might eat up the time necessary for research. In a worst case scenario external scientists might publish about the results earlier than the team originally working on the data collection (Borgman, 2012a).

To avoid this from happening there is the possibility to put the research data under an embargo for several years so the original team has the privilege of first research. Nevertheless this slows down research in general. DataCite goes another way and proposes the acceptance of the dataset as a citable medium to benefit the original researcher and data collector as source (Brase, 2010).

DataCite is an organization which handles the administrative processes around these digital quotations and consists of member organizations from currently nine countries which are mostly national libraries, research institutes or academies of sciences. Other research institution can join as associates of their national member institutes. Figure 13 shows the structure of the DataCite organization.

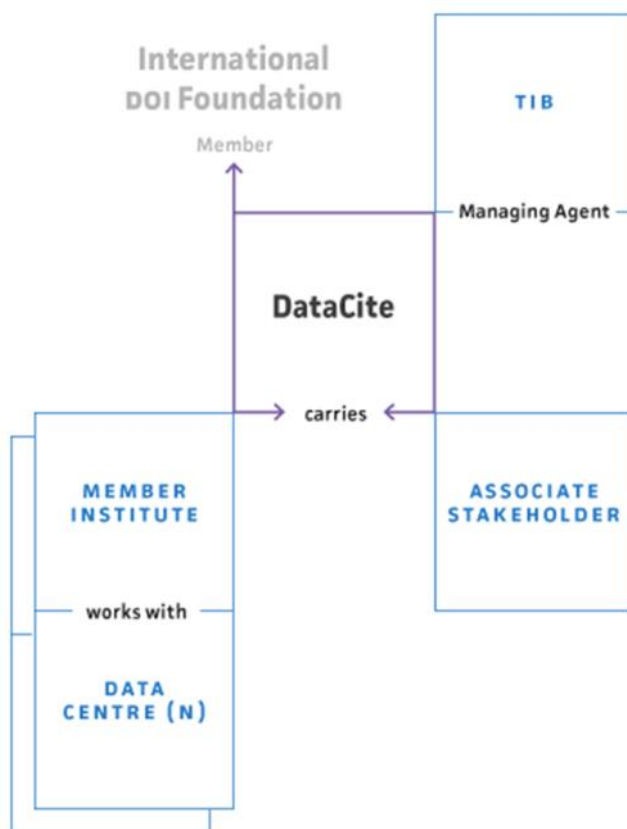


Figure 13. Structure of DataCite v3.1

Source: Retrieved from <http://www.datacite.org>

Technically DataCite provides a link between a research resource like a journal article and the data centre where the dataset resides in using Digital Object Identifiers (DOIs). These DOIs are permanent addresses which should be reachable for an indefinite time. If a DOI is clicked from a resource the dataset should be reachable or at least the researcher should end up at a landing page describing how the dataset can be obtained (e.g. by fulfilling the regulations of the data centre to obtain the dataset if it is not publically available).

The DataCite Metadata Scheme (current version: 3.1) is quite simple and reminds much of Dublin Core. This is the metadata information which would be stored for the landing page to give secondary researchers high-level information about the content. For storing questionnaires or cognitive instruments the scheme is much too basic. Table 3 shows the metadata scheme of DataCite (version 3.1).

Table 3. DataCite metadata scheme Source: DataCite, 2013 p. 6.

<b>Property</b>	<b>Obligation</b>
Identifier	Mandatory
Creator	Mandatory
Title	Mandatory
Publisher	Mandatory
PublicationYear	Mandatory
Subject	Recommended
Contributor	Recommended
Date	Recommended
Language	Optional
ResourceType	Recommended
AlternateIdentifier	Optional
RelatedIdentifier	Recommended
Size	Optional
Format	Optional
Version	Optional
Rights	Optional
Description	Recommended
GeoLocation	Recommended

Like Dublin Core the DataCite metadata schema is quite simple and straight-forward to make implementation easy due to lack of complexity. Only five of the 18 properties are mandatory which serves the main purpose to make finding an object from an external reference easier.

When discussing the right mix of metadata standards for representing cognitive items for computer-based assessment looking at DataCite makes sense for two purposes. On the one hand it could be part of the mix and if it is not sufficient it can be used on the other hand as an example to handle the linkage between journal articles and datasets by technical means.

### 1.2.4 Different metadata standards and their capabilities

The last sections gave some details of different metadata standards and their capabilities. To get a clearer picture in this section here is an overview. For the purpose of identifying which metadata standards are able to deliver which part of the survey process the Generic Longitudinal Business Process Model will serve as a guideline to the processes where the metadata standards will be categorized against.

Table 4. Comparison between GLBPM processes and different metadata standards

Metadata Standard	Evaluate / Specify Needs	Design / Redesign	Build / Rebuild	Collect	Process / Analyse	Archive / Preserve / Curate	Data Dissemination / Discovery	Research / Publish	Retrospective Evaluation
LOM v1.0	Partly	No	No	No	No	Partly	Partly	No	No
LTI v1.0	Partly	No	Partly	No	No	No	Partly	No	No
QTI v2.1	Partly	Partly	Yes	Partly	No	No	No	No	No
APIP v1.0	Partly	Partly	Yes	Partly	No	No	No	No	No
SCORM 2004 (4 <sup>th</sup> ed.)	Partly	No	Partly	No	No	No	Partly	No	No
DDI v3.2	Yes	Yes	Yes	Partly	Partly	Partly	Partly	Partly	Partly
SDMX v2.1	No	Partly	Partly	No	Yes	Yes	Yes	Partly	Partly
Dublin Core	No	No	No	No	No	Partly	Partly	Partly	No
DataCite Metadata Scheme	No	No	No	No	No	Partly	Partly	Partly	No

The first result which can be seen from the table with the high-level processes is none of the metadata standards is able to cover the demands of the business model in all details. The metadata standard filling most of the fields is Data Documentation Initiative v3.2 which fulfils at least all processes partly. Nevertheless it is lacking sub-processes like “Setup Collection” or “Run Collection” in “Collect” as it does not focus on paradata or disposition codes. In DDI 3.2 the answer to the question on how to archive proceedings from field work would be to store this information in a file and document it with the “Archive / Preserve / Curate” processes. This is not sufficient if this information should be available for researchers in searchable databases. As the DDI Alliance is open for internal and external input the chances of filling these gaps are quite high in the long run. DDI Lifecycle developed from a metadata standard which was primarily used in archiving into a standard which can also be used for survey design. An extension towards data collection seems to be a next logical step.

Also seen from the comparison can be the complimentary use of DDI and SDMX which is the basis for the DDI and SDMX dialogue in 2010 and 2011. DDI focuses on the instruments while SDMX strengthens statistical production and fulfils most requirements there. A combination of these two



standards would make sense as long as they do not cover each other's domains this well. Unfortunately as stated before there has not been much progress since 2011. The reason for this might lie in the development of GSIM as the combining information model and the work on DDI 4 which has similar goals.

Furthermore as the model shows the educational metadata standards are lacking more processes as similar models from social sciences. The main focus is either to connect e-learning platforms (like in LOM or LTI) or pure item banking (like in QTI or APIP). There is currently no standard which even remotely tries to fulfil any requirements of data management or data archiving. Research work on this topic is therefore very much on the beginning.

The Generic Longitudinal Business Process Model does not only cover singular processes but also offers continuous processes like project management or the ingest of external metadata.

Here a comparison of the metadata standards with the ongoing processes in GLBPM.

*Table 5. Comparison between GLBPM's ongoing processes and metadata standards*

Metadata Standard	Project Management / Quality Management	Metadata Management	Use of External Metadata Standards
LOM v1.0	No	No	No
LTI v1.0	No	No	No
QTI v2.1	No	No	No
APIP v1.0	No	No	No
SCORM 2004 (4 <sup>th</sup> ed.)	No	No	No
DDI v3.2	No	Partly	Partly
SDMX v2.1	Partly	Partly	Partly
Dublin Core	No	No	No
DataCite Metadata Scheme	No	No	Partly

The ongoing processes are the ones which are difficult to model in metadata standards and the question is if rather external sources should be used. It is therefore not surprising all metadata standards are lacking support here. For project management or quality management it makes sense to resort to known standards like PRINCE2 (project management) or ISO 9000 standards (quality management).

From the literature review it can be seen metadata standards for educational sciences and especially computer-based assessment do not fully fulfil the demands of data management and data archiving for long-term preservation. Also, the standards from social sciences are further developed in those fields, but cannot be used "as-is" for our demands. Therefore a requirements analysis is necessary to figure out how either educational metadata standards can be improved or if standards from social sciences can be modified accordingly. In the next section the differences between the survey processes in social sciences and the educational sciences will be discussed to get a clearer picture about those demands. This will be the foundation for the survey on those requirements where researchers will be asked for their opinion on different processes in metadata documentation.

### **1.3 Description of the current situation regarding metadata and data in educational sciences and social sciences**

As can be seen from the previous section, metadata standards in the educational sciences do not cover the needs of data management. Unfortunately the research process of educational surveys produces a lot of information and not all of them can be stored especially if post-processes are necessary. It is therefore worth analysing the research process in computer-based studies to see where metadata and data could be produced. Furthermore there will also be some examples of different data management portals from different large-scale studies. They might not comply with a common metadata standard but use a proprietary data model, but nevertheless the data management demands for educational science can be extracted from the user interfaces and functionalities. In the next sub-section there will be a comparison of different software packages for computer-based assessment where their capability or readiness for metadata management will be analysed. Finally, the information about metadata standards from the last section and the analysis of the current demands in educational sciences will lead to research questions which are the basis of modelling questionnaire items for a web survey.

#### **1.3.1 Overview of the survey and assessment cycle**

In the last years statistical agencies and social sciences institutions prefer to have lifecycle models to organize the flow of data within their organizations as can be seen from the GSBPM, GLBPM, DDI or SDMX and offer a data archive for secondary analyses. Similar notions can also be seen in the educational sciences if one looks at the domain of formative assessment (Black & William 2009) or the connection between questionnaires and cognitive items in large-scale studies like the Programme for the International Assessment of Adult Competencies (PIAAC) (Kirsch & Yamamoto, 2013) or the German National Educational Panel Study (NEPS) (Blossfeld, von Maurice & Schneider, 2011). It therefore makes a lot of sense to build the basis for further analytics in taking a look at those models. A representation of a lifecycle model in the educational sciences can be found in formative assessment.

In formative assessment the lifecycle model is used to link the results of every round of assessment to individual recommendations for improvement for students. The results of these processes look very similar to the lifecycle models discussed before in the field of social sciences. Figure 14 shows a typical example for an assessment cycle in formative assessment.

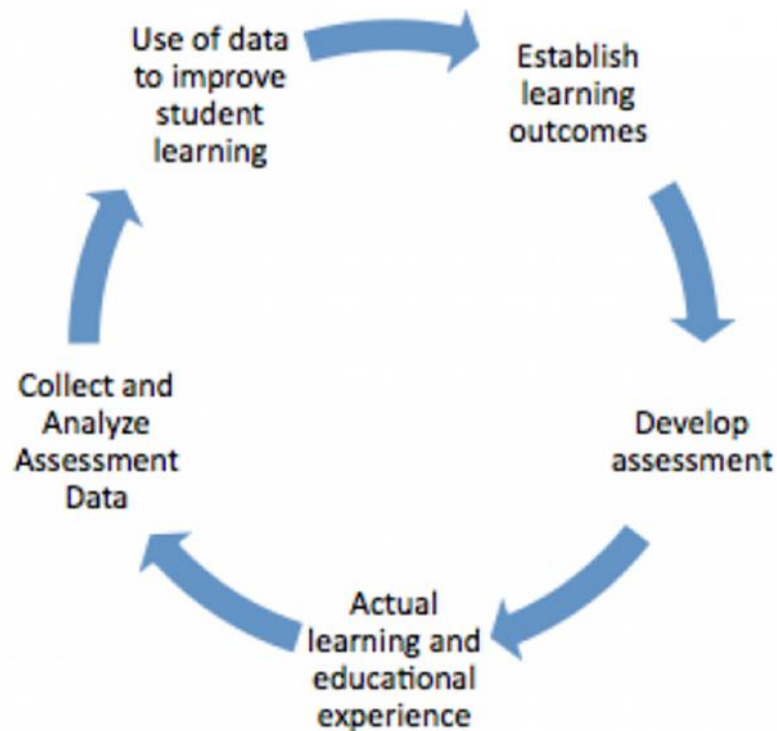


Figure 14. Assessment cycle in formative assessment

Source: Hagerstown Community College, 2014.

A computer system which is supposed to support this assessment cycle needs necessary processes of data management as the whole model can only work if the items and results from previous rounds can be processed in later rounds as well. If data formats are changed without transition processes or documentation gets lost the reporting on individual learner progresses cannot work anymore. This becomes even more crucial if the results from those assessments are also used for a summative assessment for reporting on district, state, national or international level. If a reporting is supposed to cover trends in education over longer periods the instruments and results have to be stored, documented, and versioned.

The understanding of the need of data management based on survey and assessment cycles can also be seen in international studies. PISA, PIAAC and NEPS offer data centres where the metadata and results of the studies can be downloaded for additional research (to be shown in detail in section 1.3.3). Furthermore, as those studies are either repeated cross-sectional studies with a trend (PISA, PIAAC) or longitudinal panel studies (NEPS), their internal logic bases on a lifecycle model where metadata and results from previous rounds are optimized in a retrospective phase for the next rounds. Another interesting design issue when looking at those large-scale studies is the overlap between social sciences and educational sciences. All of these studies use Background Questionnaires (e.g. student questionnaires, teacher questionnaires, parent questionnaires, participant questionnaires) linked with cognitive items from different domains (e.g. literacy, numeracy, complex problem solving, and problem solving in technology-rich environments). This means to store the computer-based instruments e.g. from PIAAC data management, processes are needed to cover both the information from the questionnaire and the cognitive items as they cannot be separated and even base on the same technology.

It is therefore worth also for the educational scientist to take a look at the processes from social sciences as they are not only similar and can be used as a blue print for further developing educational metadata standards but also as they are more and more needed in mixed mode research. Nevertheless, the standards from social sciences cannot be used “as-is” as they do not cover all the needs from educational assessment. Therefore the differences between these two modes have to be analysed.

**1.3.2 The demands of Data Management for computer-based assessment studies**

The basic structure of a computer-based assessment study does not really differ from a social sciences survey. As can be seen in studies like PIAAC, questionnaires and cognitive items can be combined into one bigger instrument and data collection can be performed with the same software and computer platform. The results of the questionnaire and the cognitive items end up in the same international and national databases. Figure 15 shows the structure of a PIAAC interview which will serve in this section as a use case.

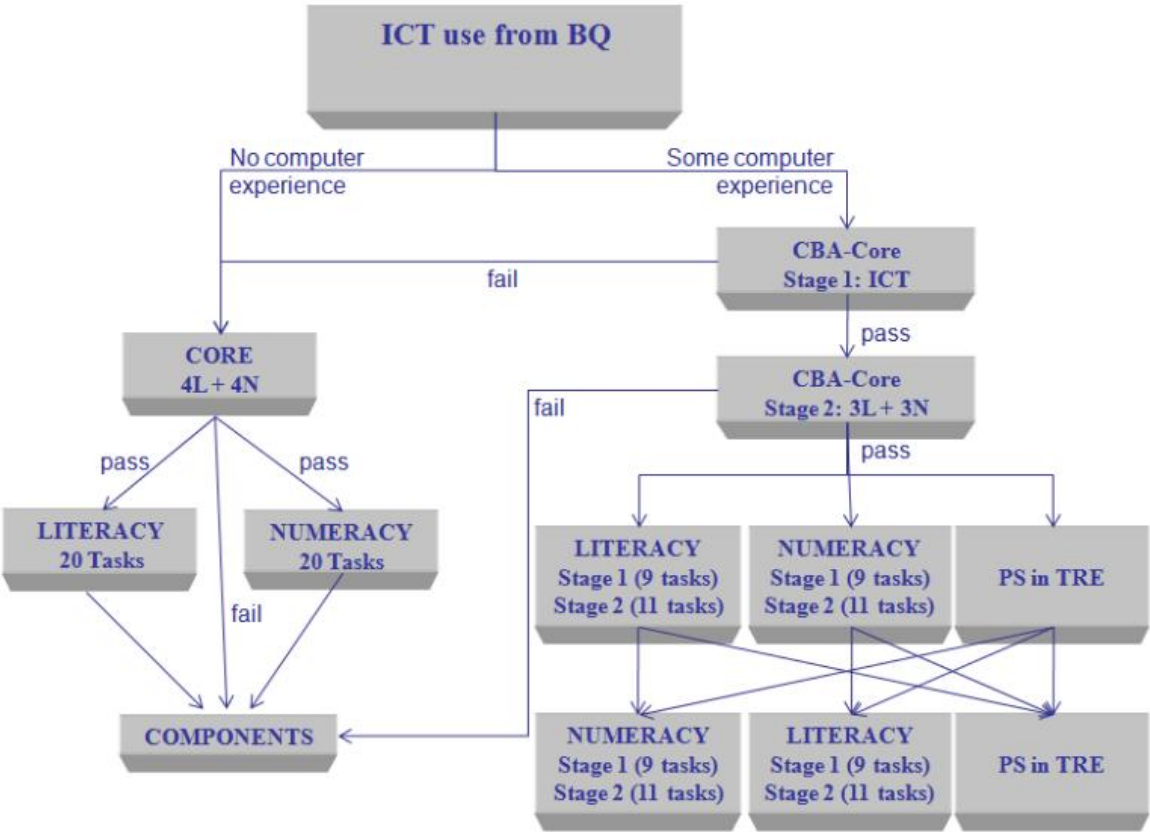


Figure 15. PIAAC Assessment Design

Source: Kirsch and Yamamoto, 2013.

PIAAC is a household study organized by the OECD which is currently in its first wave (2011) with two rounds of extension countries (Round 2: 2012 – 2016, Round 3: 2014 – 2018). In the original Round 1 24 countries participated which were extended by nine countries in Round 2. Currently Round 3 is in

its starting phase and another 10-12 countries are expected to participate. The data collection mode is a Computer-assisted personal interview (CAPI) where each country has to collect 1500 cases for the field test and 5000 cases for the main study per national language. As can be seen from the assessment design the study starts with a Background Questionnaire (40 min + 5 min national extensions) where the interviewer asks questions to the interviewee by using the laptop. If in the main test design the interviewees respond they do not have computer knowledge or fail one of the two basic computer tests (CBA Core Stage 1 and 2) they will be sent into a paper-based test. Otherwise they will perform a 60 minute cognitive test with booklets from literacy, numeracy and problem solving in technology-rich environments. Though PIAAC does not use computer adaptive testing (CAT) it uses an adaptive booklet design. The difficulty of the first booklet is chosen by educational variables specified in the Background Questionnaire (e.g. qualification) and the participants get, based on their performance on the booklet before, a more difficult booklet, an easier booklet or a booklet on the same level.

What makes PIAAC interesting as an example to show the differences in design between the data management for surveys and assessment is it contains a lot of these issues within the same study (OECD, 2013).

- The Background Questionnaire is processed by the interviewer, but the cognitive assessment by the interviewee.
- Both parts of the interview nevertheless happen on the same machine and software.
- The PIAAC design allows pauses and break-offs during the Background Questionnaires, but not while the cognitive part is running.
- In the Background Questionnaire it was allowed to change languages on runtime, the cognitive part had to stay within the same language
- Screen size issues do not play any role for the Background Questionnaire and the interviewer questionnaire at the end, but for the cognitive items the screen size and resolution were standardized (Original Design: 14" 1024x786, Round 1 Main Study: 15,4" 1280x800, Round 2 Main Study: 15,6" 1366x768).
- All international data management processes are handled by the same software (Data Management Expert by IEA/DPC).

PIAAC uses as for its overall instrument different software platforms which are bundled together into a Virtual Machine (VM) to make the IT processes for the national teams easier. Most parts of the platform were modules from the Testing Assisted par Ordinateur (TAO) system developed by the Centre de Recherche Public Henri Tudor (CRP) and the University of Luxembourg (Latour, 2007). In Round 1 the domains of literacy and numeracy were handled by the CBA Itembuilder which was developed by the German Institute for International Educational Research (DIPF) in cooperation with Softcon (Rölke, 2012). The CBA Itembuilder is constructed to integrate with the TAO delivery platform. In Round 2 also the literacy and numeracy items were rendered by the TAO engine as a preparation for the PISA 2012 and PISA 2015 platform. Figure 16 depicts the German delivery system for PIAAC.

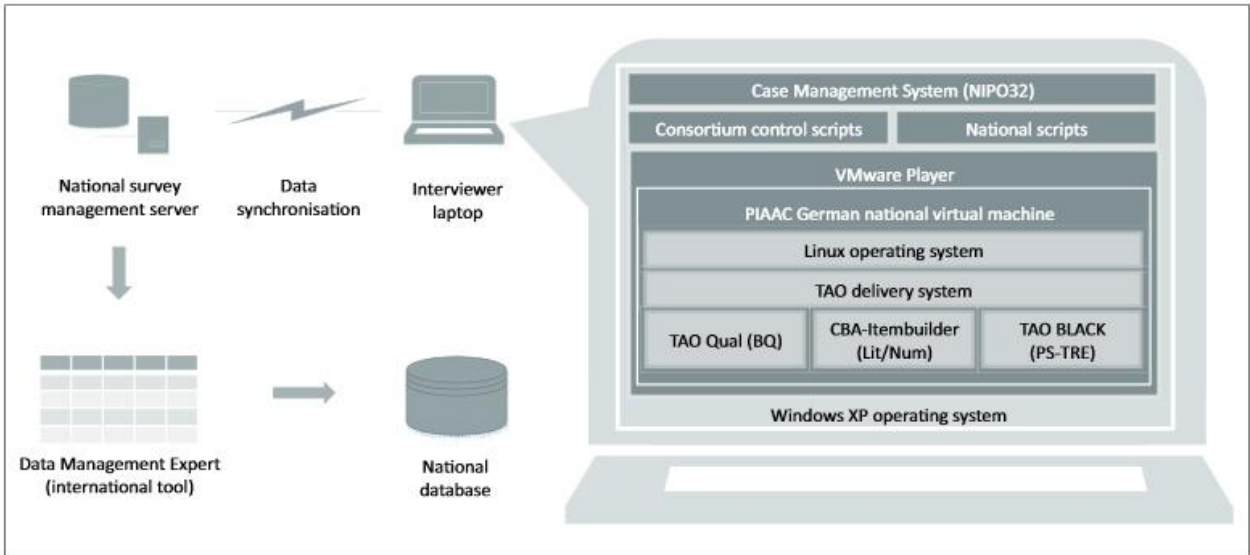


Figure 16. German PIAAC delivery system

Source: Zabal., Silke, Massing, Ackermann, Helmschrott, Barkow, & Rammstedt, 2014.

Though TAO is mainly an assessment system it was also used for rendering the Background Questionnaire. For this purpose the TAO Workflow Engine “TAO Qual” was modified to represent the logic of questionnaires (e.g. routing, branching, dynamic text, soft breaks and hard breaks). Nevertheless though the same core software is used the questionnaire and the cognitive items in PIAAC look quite differently. Figure 17 shows an example for a question item in PIAAC:

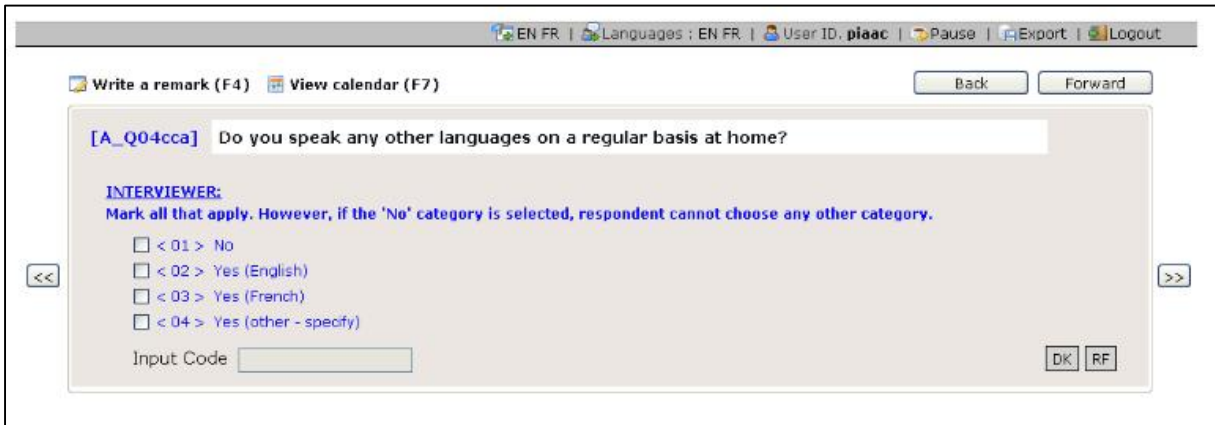


Figure 17. PIAAC questionnaire item

Source: Latour, Jadoul and Wagner, 2013.

Though running on the same platform the design and functionality of the cognitive part differed a lot. Figure 18 presents for comparison reasons a cognitive item from the domain of literacy.

Look at the list of preschool rules. Highlight information in the list to answer the question below.

**What are the two rules about taking medicine to the preschool?**

### Preschool Rules

Welcome to our Preschool! We are looking forward to a great year of fun, learning and getting to know each other. Please take a moment to review our preschool rules.

- Please have your child here by 9:00 am.
- Bring a small blanket or pillow and/or a small soft toy for naptime.
- Dress your child comfortably and bring a change of clothing.
- Please no jewelry or candy. If your child has a birthday please talk to your child's teacher about a special snack for the children.
- Please bring your child fully dressed, no pajamas.
- Please sign in with your full signature. This is a licensing regulation. Thank you.
- Breakfast will be served until 7:30 am.
- Medications have to be in original, labeled containers and must be signed into the medication sheet located in each classroom.
- If you have any questions, please talk to your classroom teacher or to Ms. Marlene or Ms. Tree.

Figure 18. PIAAC literacy item

Source: Upsing., Gissler, Goldhammer, Rölke, & Ferrari, 2013.

The screen design and also the answer mode are completely different. While the questionnaire part uses multiple choice answer mode Figure 18 shows a simple example of the ‘highlighting’ answer mode which can be compared to a digital text marker. Highlighting does not exist as an answer mode in typical survey systems, but also not as an interaction in educational metadata standards like QTI (IMS 2006). The challenge becomes even bigger as PIAAC contains items with a much higher complexity especially in the field of problem solving in technology-rich environments where simulated web browsers, mail clients and business software are used.

PIAAC as a study can be described with the processes defined by the Generic Longitudinal Business Process Model (GLBPM) if the high level categories are used. The survey design and also preservation processes fit to the demands of a social sciences repeatable cross-section study. The problem lies in the details of the sub-processes where new input has to be considered if the GLBPM is to be extended to also fulfil the demands of educational sciences or the model has to be transformed into a Generic Educational Business Process Model if the gaps are too big to be covered in one model. In the next sections the GLBPM and DDI are used as a guideline for the survey processes and the necessary extensions needed for cognitive items from computer-based assessment in education will be added.

### 1.3.2.1 Study Concept in Educational Assessment

“Study Concept” is the first phase of planning a new study or the next wave of a longitudinal or multiple cross-sectional survey. The Generic Longitudinal Business Process Model defines in the same phase “Evaluate / Specify Needs” the following sub-processes (Barkow et al. 2012):

- Define research question, universe and high-level concepts
- Evaluate existing data and publications
- Establish output and needed infrastructure
- Define specific concepts to be measured
- Plan, create time-table and identify needed infrastructure
- Prepare proposal and get funding

These processes are very similar to the study design used by PIAAC in the Technical Report provided by the OECD (Kirsch & Yamamoto, 2013) or the German PIAAC Technical Report (Zabal et al., 2014). From a first look setting up an educational assessment has similar processes to the social sciences studies. If this is the case the process Study Concept from DDI which is much related to the “Evaluate / Specify Needs” can be used to describe also the content of educational assessment.

This is true for computer-based assessment as long as the study is a summative assessment of a larger population in a controlled environment like in PIAAC or PISA. The situation changes if the focus is directed towards formative assessment or high stakes testing for certification processes. “Study Concept” in DDI and GLBPM can describe the structure of a study which is organized by researchers and where there is the preparation of a scientific dataset at the end. In educational assessment items might be prepared by teachers to test the individual further development of their students. As can be seen in the assessment cycle from formative assessment in Figure 14 these processes do not necessarily have to result in scientific publications, but might be the basis of a reporting on individual student, class, school, state or national level.

For describing metadata on this level a model from e-learning like LOM might make more sense as it describes the context in the curriculum for specific learning domains. This is also the reason why LOM has to be adapted to the school system of individual countries as e.g. the school system in the UK differs completely from the three-structured school system in Germany. Figure 19 shows the UK LOM Core Model as a country representation.



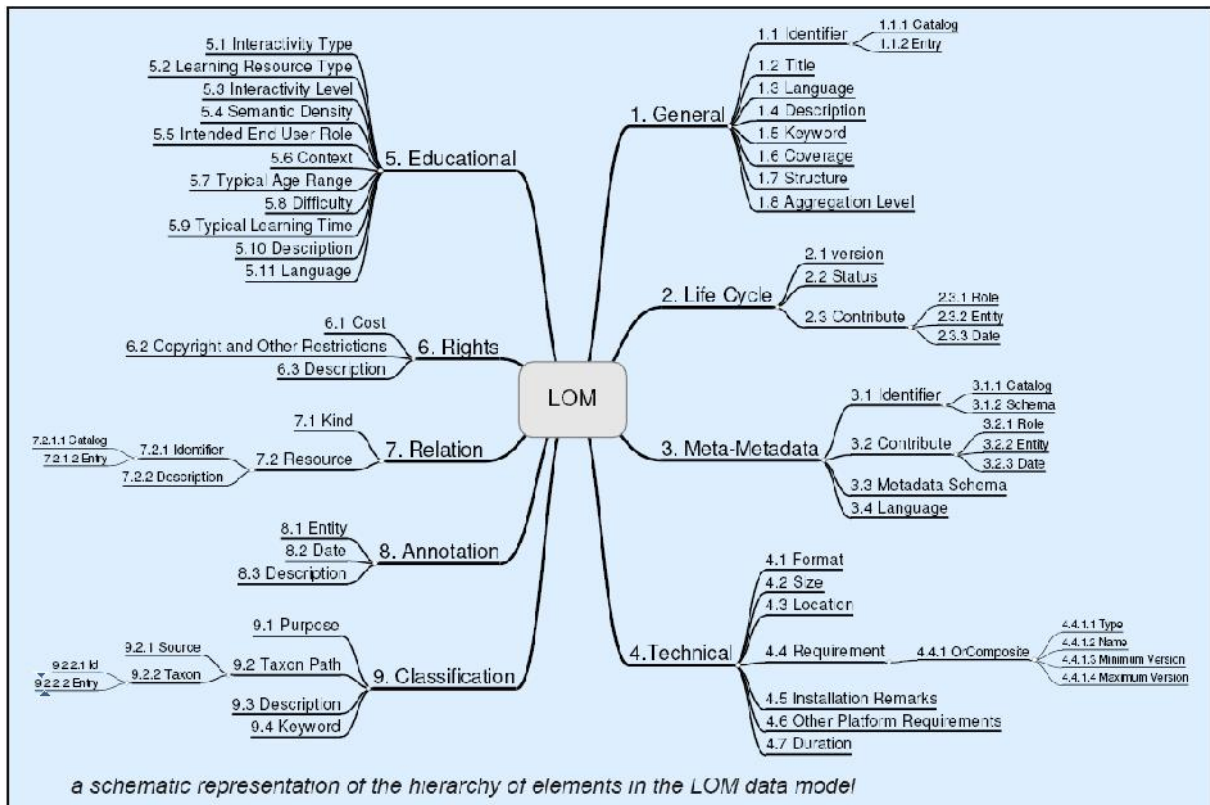


Figure 19. UK LOM Core

Source: Barker, 2005.

As Figure 19 shows LOM can be used to describe the school context, but none of the nine main elements contains concepts for the scientific domain (e.g. survey information, prime investigator, research team, information on data collection). Items for the scientific context need another metadata representation than items produced for school measurement. A new metadata model for education therefore needs both components. The situation becomes even more complex if commercial high-stakes testing or educational tests for Human Resources have to be modelled. It can be doubted one metadata standard can model all those cases like a “Jack of all trades, master of none” approach. For educational sciences the answer might be to work with different metadata standards for different purposes. Educational studies could use a slightly adapted or modified DDI while national school monitoring might benefit from a model based on LOM. This would have also the advantage of connectivity to E-Learning and Blended Learning platforms.

### 1.3.2.2 Item Design in Educational Assessment

Another difference between computer-based studies from social sciences and educational assessment is the higher complexity of the items. The problem fields are answer modes of complex item types, screen size or layout problems, mode effects and scoring. Cognitive items in computer-based assessment can resemble much complete simulations of work environments or contain scenarios. Figure 20 shows such a more complex example from the PIAAC use case.

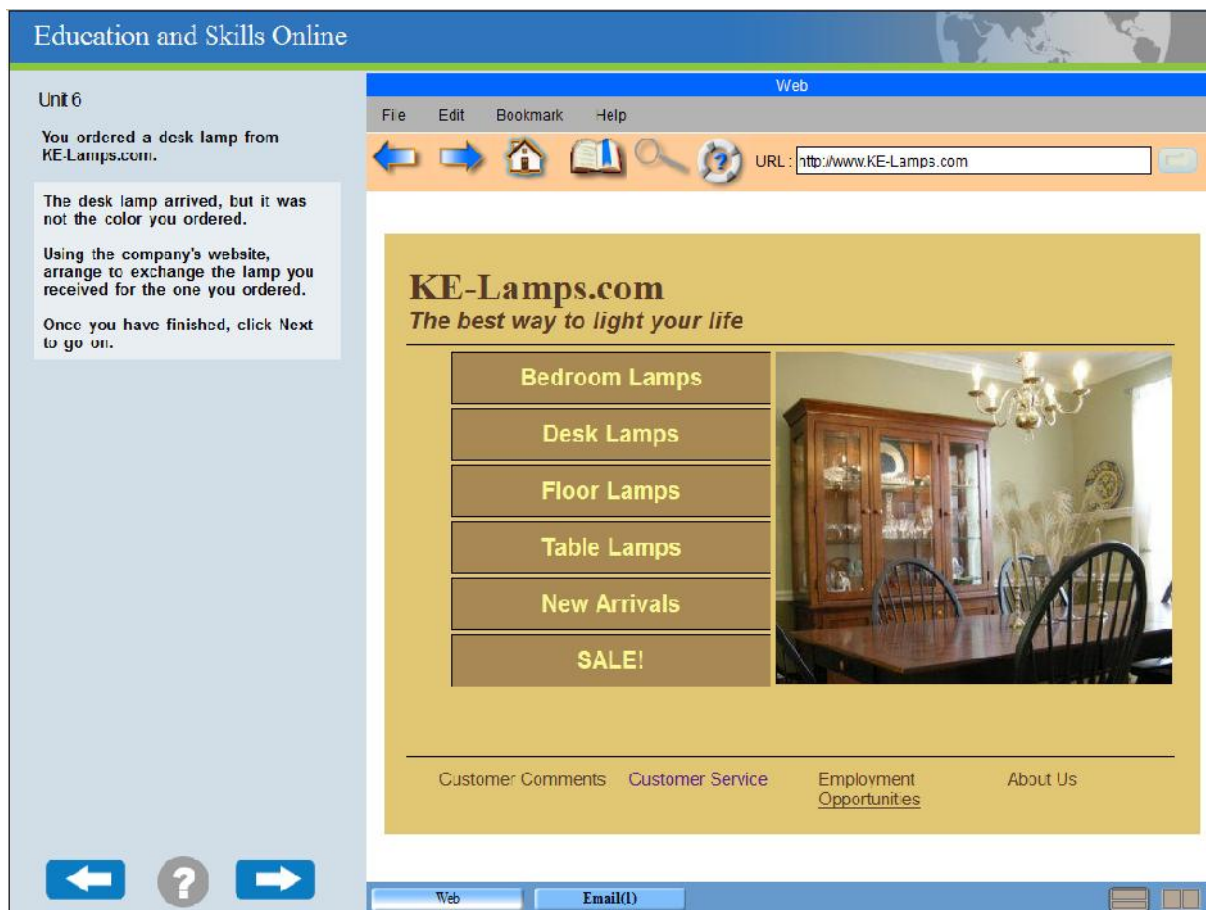


Figure 20. PIAAC Problem Solving Item

Source: OECD, 2013.

The participant of the study has to use a simulated website to exchange a desk lamp in a web store. The item has several web pages, a shopping basket and a mail application. It uses a complex stimulus to come near to the reality of similar applications in the internet. This is a completely different level of complexity for long term preservation than a simple item using e.g. a picture as stimulus and a battery of four multiple choice answers as answer categories. Neither DDI nor QTI are currently able to model an item like described above. Assessment software like CBA Itembuilder and TAO can produce such items, but for this purpose they do not use any metadata standard.

To add to this complexity there is also a second dimension in computer-based testing and this is fixed layout and screen size. Other than questionnaires in social sciences the information to be measured can change heavily if in an exercise the key word of solving the task is the last word in the second line of the first paragraph or the second word of the third line of the first paragraph. Complex items like the one above change their difficulty if the fonts change their size. Unfortunately web standards like Hypertext Markup Language (HTML) have the problem they do not support a fixed layout, but the elements on a web page arrange themselves according to the screen size (W3C, 2009). Additionally the layout and fonts of web pages are interpreted differently by different browsers (e.g. Microsoft Internet Explorer, Firefox, Chrome or Opera). This also holds for different versions in one browser family (e.g. Internet Explorer 7 vs. Internet Explorer 8).

In a worst case scenario the item was designed for a bigger screen and is now displayed on a smaller size. This can lead to much too small fonts which cause problems in readability for the elderly especially

if long texts are used or the text does not fit on the screen anymore which leads in modern operating systems and browser to the display of only parts of the text with adding scroll bars. Using scroll bars is from an assessment perspective a different task than answering a literacy item and therefore the difficulty of the item is completely distorted. In the past computer-based assessment systems have therefore been implemented in technologies like Adobe Flash to circumvent this problem, but these carry the problem of compatibility with several devices or operating systems.

Another additional metadata information which does not exist in social sciences driven standards like DDI is scoring. In a computer-based assessment one of the advantages is the computer will evaluate for most answer modes on run-time if the item has been answered correctly or incorrectly. This automatic scoring offers plenty of opportunity. It is possible to define partially correct answers. Furthermore, changes of the answers during the assessment can also be evaluated. These proceeding are the basis for Computer Adaptive Testing (CAT) (e.g. Weiss & Kingsbury, 1984) where after each cognitive item another item with the same, a higher or a lower difficulty can be chosen. This enables much shorter testing times (candidates in a high-stakes test may fail after only 10-12 questions). For data management of computer-based assessment items nevertheless this means the scoring rules as well as statistical measures like item difficulty or Cronbach's Alpha (Cronbach, 1951) have to be stored in a reproducible manner for later re-use or later analysis by other researchers who were not involved in the original study. Otherwise paradigms like Item Response Theory (IRT) (e.g. Lord, 1980) do not have their input parameters to work properly. QTI is therefore able to store scoring rules including adaptive algorithms to a certain extent.

A completely new metadata standard for computer-based assessment therefore needs more answer modes than QTI so also more complex items could be stored. This is quite complex especially in the light of the layout problems mentioned above. Currently no metadata standard from social sciences or educational sciences stores layout information in detail or can provide a fixed screen layout. For simulations also the flow logic of the individual sub-processes of an item have to be stored in a way they are still understandable in coming years. This tasks can be considered quite challenging.

### **1.3.2.3 Data Collection in Educational Assessment**

A further speciality of educational assessment is an upcoming stronger focus on log files and paradata. Paradata are administrative metadata about the data collection process on a micro-level (e.g. how many days the interview took, contact attempts by the interviewer, timing of answers by the interviewee) (Couper, 1998). Analysing paradata in the social sciences is mainly done for checking interviewer behaviour (e.g. to look for cheating) or project management of field work. Normally the paradata about happenings during data collection comes back from the data collection in form of a field report.

In educational assessment paradata like the timing of interviewee answers can provide highly interesting research insights as the complete answer path can be traced by replaying the log file. For the PIAAC use case DIPF is currently analysing log files from the German PIAAC field test with the following research questions (Goldhammer, Naumann, Stelter, Tóth, Rölke, & Klieme, 2014).

- Processing time: Integrated measurement models for responses and response times (IRT-RT models) serve to empirically determine the relation between domain specific skills and speed constructs.

- Processing steps (theory-based): For the domains of literacy and problem solving, theory-based process measures are constructed describing the quality of a navigation path and serving as predictors of successful outcomes as well as competence indicators.
- Processing steps (exploratory): The recorded sequences of events are also to be analysed automatically by means of process mining techniques, to explore and determine from empirical processing steps an underlying process model of actual processing behaviour.

This research is fully based on receiving the unprocessed log files from the study so every process step can be analysed. Unfortunately there are several challenges to this as different assessment platforms use completely different log file formats with different granularity.

Figure 21 shows an example a comparison between the log file output of TAO BLACK (PIAAC Assessment platform) and CBA Itembuilder EE.

```

475 HISTORY_BACK      stimulus
    pageid=unit10apage3|*$title=JobService|*$url=http://www.jobserv
ice.de
603 DOACTION          stimulus
    action=as://historyBack(toolbar_back_btn)
110 MENU              stimulus    id=wb-bookmarks-menu
645 MENUITEM          stimulus    key=bookmark-add
425 BOOKMARK_ADD      stimulus
    pageid=unit10apage3|*$title=JobService|*$url=http://www.jobserv
ice.de
475 BUTTON            stimulus    id=add_bookmark_validation
476 DOACTION          stimulus    action=as://hide(addbookmark_window)
780 TOOLBAR            stimulus    id=toolbar_bkm_btn
781 DOACTION          stimulus    action=as://hide(addbookmark_window)
783 DOACTION          stimulus    action=as//addBookmark(tabs-
toolbar.managebookmark_list,add_bookmark_book_title_input,add_bookma
rk_bookmark_url_input,add_bookmark_validation,add_bookmark_cancel)
840 BUTTON            stimulus          id=add_bookmark_cancel
842 DOACTION          stimulus
    action=as://hide(addbookmark_window)
034 MENU              stimulus          id=wb-bookmarks-menu

```

Figure 21. TAO BLACK log file from PIAAC

Source: adapted from DIPF (<http://tba.dipf.de/en/Projects%20a-z/dataanalyzer>)

DIPF developed a tool called DataAnalyzer to convert the output from the log files into useful tables which can be analysed by researchers into input for the common statistical packages (SPSS, Stata, M-Plus). In comparison Figure 22 shows a log file from the Execution Environment of the CBA Itembuilder.

```

899914      15      {"sender": "15", "type": "unloading", "data": ""}
900006      15      {"sender": "15", "type": "variable_change", "data":
      {"name": "item_score", "value": "<?xml version=\"1.0\"
encoding=\"ASCII\" ?>      \u000a<cbascoreingresultmm:ItemScore
xmi:version=\"2.0\" xmlns:xmi=
      \"http://www.omg.org/XMI\" xmlns:cbascoreingresultmm=
      http://cbascoreingresultmm/1.0\" name=\"Item15_Algen_Task1\",
      result=\"true\" nbHits=\"1\" hitWeight=\"1\" nbMisses=\"0\"
      missWeight=\"0\" creditClass=\"1\" creditWeight=\"1\"
execTime=\"67\" execTimeTotal=\"67\" reactionTime=\"15\"
reactionTimeTotal=\"15\" nbInteractions=\"7\"
      nbInteractionsTotal=\"7\">\u000a
      <hitList name=\"Item15_Algen_Hit1\", weight=\"1\" />\u000a
      </cbascoreingresultmm:ItemScore>\u000a"}}}
900009      15      {"sender": "15", "type": "variable_change", "data": {
      "name": "_15_NumberOfActions_state", "value": {"score": "7"}}}

```

Figure 22. Log File from CBA Itembuilder Execution Environment (CBA EE)

Source: retrieved from DIPF (<http://iba.dipf.de/en/Projects%20a-z/dataanalyzer>)

Both log files measure interactions with the assessment system (e.g. clicked buttons, mouse movements) with according timestamps. Unfortunately they are not compatible to each other. The PIAAC version of the CBA Itembuilder stored logfile information within TAO BLACK so data analysis can be performed with the same methods (this was reached by conversion and injection - the actual internal logfile had a different format), but newer versions of both tools lost this feature. In fact neither TAO (current version: 2.6) nor CBA Itembuilder (current version: 6.0.1) create the same log files like the PIAAC versions from the year 2010. Tools like the DataAnalyzer which have been developed to transform the log file output from the PIAAC Field Test into SPSS or Stata format have to be modified every time as newer version of the assessment tools are released. The same is valid also for other assessment systems as currently there is no common metadata standard to describe log file formats for computer-based assessment. A new or extended metadata standard should therefore take defining paradata from log files into consideration. Otherwise there will be a lot of additional transformation work needed to bring the individual log files from different studies and systems into a format which is sustainable for long-term preservation.

### 1.3.2.4 Data Analysis in Educational Assessment

The processes of data analysis in the social sciences and educational sciences are quite individual and also depend on the statistical tools used for data production. The process in the GLBPM describes the transformation process from raw data from data collection into files the researchers can use for publication later. The raw data cannot be used for this as several processes might have to be performed for quality assurance or legal reasons (Barkow et al., 2012).

- Data cleaning (e.g. remove broken off cases)
- Classification or coding (e.g. for open answers)
- Impute missing data

- Construct new variables
- Calculate weights
- Calculate aggregates
- Anonymize data

These processes are quite common and happen both in social sciences and educational studies as the problems of data collection and legal obligations are similar. This process step does not describe the analysis work of a secondary analysis of the data using dissemination products. This step defines the work of the agency which does the data collection in liaison with the primary research team to clean the output from field work into usable statistics.

DDI and especially SDMX describe the statistical processes quite well so they can be used in educational sciences as well or serve as a blue print for metadata standards in the educational sciences.

### **1.3.2.5 Data Archiving in Educational Assessment**

Data Archiving is next to the actual data collection the most complex and time-consuming of the processes in the GLBPM. The sub-processes are:

- Ingest data and metadata
- Enhance metadata
- Preserve data and metadata
- Undertake ongoing curation

The first step is very typical for data archives or research data centres. Normally the metadata and data are not stored at the place where the research originally happened, but at a third party facility. In the United Kingdom research normally has to be handed in centrally at the UK Data Archive (UKDA) in Essex. Next to the legal considerations of depositing data at another facility (e.g. access control, confidentiality, costs) ingest of metadata and data which were not produced in-house are a challenge especially if the data producer did not use any metadata standards so the information could be easier harmonized. In the data archive researchers have to analyse the data structures of the incoming metadata and data to see how it fits to the storage structure of the data archive. As these researchers are not the original investigators of the study and therefore have to rebuild the information of the study from different sources the process within the data archive can be considered as data forensics as the reconstruction happens from small information bits and is very often difficult and time-consuming. This is also part of the next step.

Very often a study does not have all the metadata information which is needed for the data archive to present it for their data users or to describe it sufficiently for their internal storage structures. The reason for this is the original research did not consider the necessities of long-term preservation. Policy makers like the European Union try to create an awareness in researchers by asking them for a data management plan when they hand in research proposals (Barsky, Abel, Deardorff, Henderson, Mitchell, Soper, & Taylor, 2014 or Digital Curation Center, n.d.), but current studies still lack an understanding of these processes. In this case the data archive has to use a process called “enhanced metadata” which means the metadata which was delivered with the study is combined with additional new metadata e.g. from the data forensics process or other sources. As the enhancement is time-consuming and leads to additional costs it normally cannot be performed to all incoming studies and raises the question who has

to pay. The UKDA estimates with their 60 persons staff they can process about 5% of the 3000 studies handed in yearly to the level of enhanced metadata (CURCAT 1 - curation category 1, former category A\*). All other studies have to be stored in different levels of less processing (see Corti, 2006). Here an overview of the curation categories:

*CURCAT1: Data collections selected for long-term curation. These data collections are made available for download, or accessible via online access tools.*

*CURCAT2: Data collections selected for “short-term” management. These data collections will not (initially) be retained for long-term preservation, rather they will be backed-up (i.e., bit-level preservation only), made accessible and discoverable through online access tools (including Nesstar, InFuse, UKDS.Stat, etc.) or via in-house repository software (ReShare).*

*CURCAT3: Data collections selected for ‘delivery’ only, e.g., where data from third parties are accessed via APIs/web services and delivered to end users via a UK Data Service interface. Issues such as level of trust in owner, what documentation/metadata are required, and how rights/registration are handled need to be agreed.*

*CURCAT4: Data collections selected for “discovery” only. These collections will not be brought formally into the holdings of the UK Data Service, they will exist only in other (institutional) repositories, but the UK Data Service will harvest (or in exceptional circumstances, create) metadata records to allow these data collections to be found more easily.*

*A fifth category (CURCAT5) relates to preservation-only which falls outside the scope of the UK Data Service Appraisal, and which is handled by the UK Data Archive. Data collections may be moved into higher or lower categories over time if the need arises.*

A long-term goal of data centres and data archives therefore should be to raise awareness for using metadata standards already in study design and field work to let ingest and enhance processes start at a much higher level.

The processes of “preserve data and metadata” and “undertake ongoing curation” serve a long-term basis. To make computer-based sources available for long-term preservation due to the ongoing technical changes all metadata and data have to be evaluated constantly if they are still usable in modern environments. If this cannot be guaranteed they either have to be skipped by which they become unavailable or have to be moved to newer formats or standards. These processes normally differentiate as described at the beginning research data centres from data archives. Research data centres like the DIPF RDC for Education have a preservation time of ten years which means they offer the processes “ingest data and metadata” and “enhance metadata”, but not “preserve data and metadata” and “undertake ongoing curation”. This would be the task of a data archive which offers long-term preservation. This leads again to questions of costs as the number of studies outweigh the capacities of data archives, e.g. the DIPF RDC for Education is planning to store its data after the ten years period at the data archive at GESIS the Leibniz Institute for Social Sciences, but it is currently not clear if the funding will be sufficient to preserve the whole content.

In the field of computer-based assessment for the educational sciences this situation even becomes more complex. As described in the sub-section “Item Design in the Educational Sciences” we are currently struggling to harmonize the layout, scoring rules and mode effects within the same period with similar software tools. The challenge becomes even greater if multimedia content has to be stored in a long-term perspective. Currently metadata standards like QTI have no answer to this challenge as they do not offer the storage of layout information or complex content. Metadata standards in the long run need to

have a solution for this challenge or it will not be possible to preserve computer-based items on a long term.

### **1.3.2.6 Data Dissemination in Educational Assessment**

Data dissemination describes the processes of data analysis by data users which were not necessarily members of the original research team but the data is provided by a research data centre or data archive as a data product. These data products can have multiple forms to be used by the researcher:

- Public Use File (data which has been heavily anonymized and aggregated so it can be put onto a website without access control)
- Scientific Use File (data files with a moderate level of anonymization which are available from data centres or data archive if a scientist registers with a research proposal)
- On-site facilities (the data centre or archive offers guest scientist seats in a secure environment where data analysis can be performed)
- Data enclaves, secure data services or Virtual Research Environments (VREs) (the data archive offers secure remote access to the data to be analysed) - see Brandburn et. al. (2006), Lane et. al. (2008) or Barkow et. al. (2012)
- Remote calculation or job submission services (the scientist is not allowed to see the data directly, but can by using mock data create scripts for analyses which can be handed in to the data centre to be executed on the real data)

As providing metadata and datasets are the main tasks for data centres and data archives where they create their demand from an important sub-process of the data dissemination process is the promotion of the dissemination products so researchers find and use the data for secondary analyses. From a metadata standard perspective data dissemination can supported by organizational standards like Digital Objects Identifiers (DOIs) in the DataCite Metadata Scheme as it links data to publications and might increase the interest of researchers to perform additional research on the datasets as reading the journal article might trigger research ideas.

Nevertheless a challenge to research data centres and data archives is the question how to preserve the secondary analyses of the following researchers. While the data analysis process between the primary researchers and the data collection agency in large surveys is at least standardized and quality controlled the calculations which are performed by the secondary researchers often state a mystery. Here an interesting research question could be if mechanisms can be developed to encourage secondary researchers to ingest their findings as well like the original research team (thus becoming a secondary data producer) or if they are at least willing to hand in the statistical scripts which produced the new datasets and variables.

The current metadata standards except for DataCite in case of linkage from data to publications do not support secondary dissemination process neither in the social sciences nor the educational sciences.



### 1.3.2.7 Publication of data in Educational Assessment

The final process in the GLBPM (except for the Retrospective Evaluation which is only necessary for longitudinal studies) and also the pinnacle of research work is the publication process. As described in the section before, this process can be supported by Digital Object Identifier (DOIs) like in DataCite or Universal Resource Locators (URN) like in DDI and SMDX.

In a best case research scenario, publications and data are linked and openly available under an Open Access License as they reach the widest spread within the research community. DIPF is currently trying to link its Open Access Library pedocs with the data stored within the Research Data Center for Education datasets. Figure 23 shows an overview of pedocs within Fachportal Pädagogik (the German Education Portal).

The screenshot shows the 'pedocs' interface on the 'FACHPORTALpaedagogik.DE' website. The page features a search bar with a 'Search' button and a 'Publish texts ...' link. Below the search bar, there is a section for 'Top-Downloads of the last month (October 2014)' with a table of results:

Rank	Author	Title	Downloads
1.)	Looser, Dölf	Soziale Beziehungen und Leistungsmotivation. Die Bedeutung von Bezugspersonen für die längerfristige Aufrechterhaltung der Lern- und Leistungsmotivation Book (monograph)	2020
2.)	E-Learning 2009	Lernen im digitalen Zeitalter Book (monograph)	1340
3.)	Schwarzer, Ralf; Jerusalem, Matthias	Das Konzept der Selbstwirksamkeit Article (from a serial)	566
4.)	Dozio, Edo	Evoluzione della politica scolastica sull'insuccesso e il disadattamento: il caso del Canton Ticino Article (journal)	432
5.)	Wilbers, Karl	Wirtschaftsunterricht gestalten. Lehrbuch Book (monograph)	393

On the right side, there is a 'Latest publications' section with a list of recent works, including 'Qualifizierungsangebote in der Pflegehilfe für Personen mit geringen Grundbildungskennntnissen' by Stuckatz, Diana; Wagner, Comelia (2014).

Figure 23. pedocs at Fachportal Pädagogik

Source: retrieved from DIPF - <http://www.pedocs.de/index.php?la=en>

This is an actual example how different metadata standards can work together. The datasets in the repository of the Research Data Centre for Education are documented in DDI, which also builds the basis for the landing pages used for the DOIs for DataCite. In the last stage these DOIs will be connected to the Open Access texts available in pedocs. Currently in the case of educational assessment this process is only available for instruments in paper and pencil format. Extending this to the field of computer-based assessment is actually the motivation for this dissertation.

### 1.3.3 Examples for Data Management in large-scale studies

The last two sections looked at different metadata standards and the principal demand of metadata management for different processes in the educational sciences. To elaborate more details it is worth to look at examples how data management is performed in large-scale studies as there is an automatic demand for secondary research as the data are interesting to many researchers which did not participate in the original data collection. For this reason at least the large OECD studies like PISA and PIAAC developed processes how at least the international data can be accessed.

#### 1.3.3.1 OECD Programme for International Student Assessment (PISA)

Since the year 2000 the Programme for the International Student Assessment (PISA) developed by the Organisation for Economic Co-operation and Development (OECD) has become one of the most prominent studies for the comparison of national students' achievements. Though it was not developed as a longitudinal study, but as a cross-sectional study using a three years' rhythm (currently PISA 2015 is reaching the main study phase while PISA 2018 is in the planning phase) it gives countries the opportunity to regularly benchmark the development of their school system in comparison to other countries in a national ranking. In some countries PISA therefore directly triggers decisions from policy makers.

The participation of Germany in PISA 2000 led to something which all major German newspapers afterwards called "the PISA shock". German students overall reached only a position in the lower average between all countries (overall score Germany in PISA 2000 – position 22 of 32 tested nations), in the domain of mathematic competence even only a position in the lower end. Even the favoured students from a Gymnasium did not score well in the three domains literacy, mathematics and science. Therefore heavy criticism from the public came up questioning the whole school system, especially the division into three different school types as the Scandinavian countries with a unified school form scored at the top positions (especially Finland is always quoted to be an example). Barz (2011) compares the "PISA shock" for Germany to be on the same level like the Sputnik shock in 1959 for the United States. A fundamental common general believe of superiority in one field (in the US in rocket science, in Germany education) has been ultimately destroyed.

Especially in Germany there is a lot of criticism towards PISA like Jahnke and Meyerhöfer (2006), Wuttke (2009) or Krautz (2012). Their main points can be categorized in these main points.

- *Statistical failures*
- *Intercultural bias*
- *Quality of the items*
- *Quality of the overall framework*

The interesting point from the perspective of data management is not the criticism per se as large studies like PISA will produce supporters and adversaries due to their size and impact. Much more relevant is the availability of all international and very often also national data for all PISA rounds from 2000 to 2012 so researchers interested in own research on the study can perform it. The international datasets from PISA 2012 as well as the US national datasets are available in the PISA Data Explorer at the

National Center for Educational Statistics. The website also offers information and summary results from PISA 2003, PISA 2006 and PISA 2009. Furthermore the Data Explorer is available for International Association for the Evaluation of Educational Achievement (IEA) studies like the Progress in International Reading Literacy Study (PIRLS) or the Trends in International Mathematics and Science Study (TIMSS). Figure 24 shows a screenshot of the PISA Data Explorer:

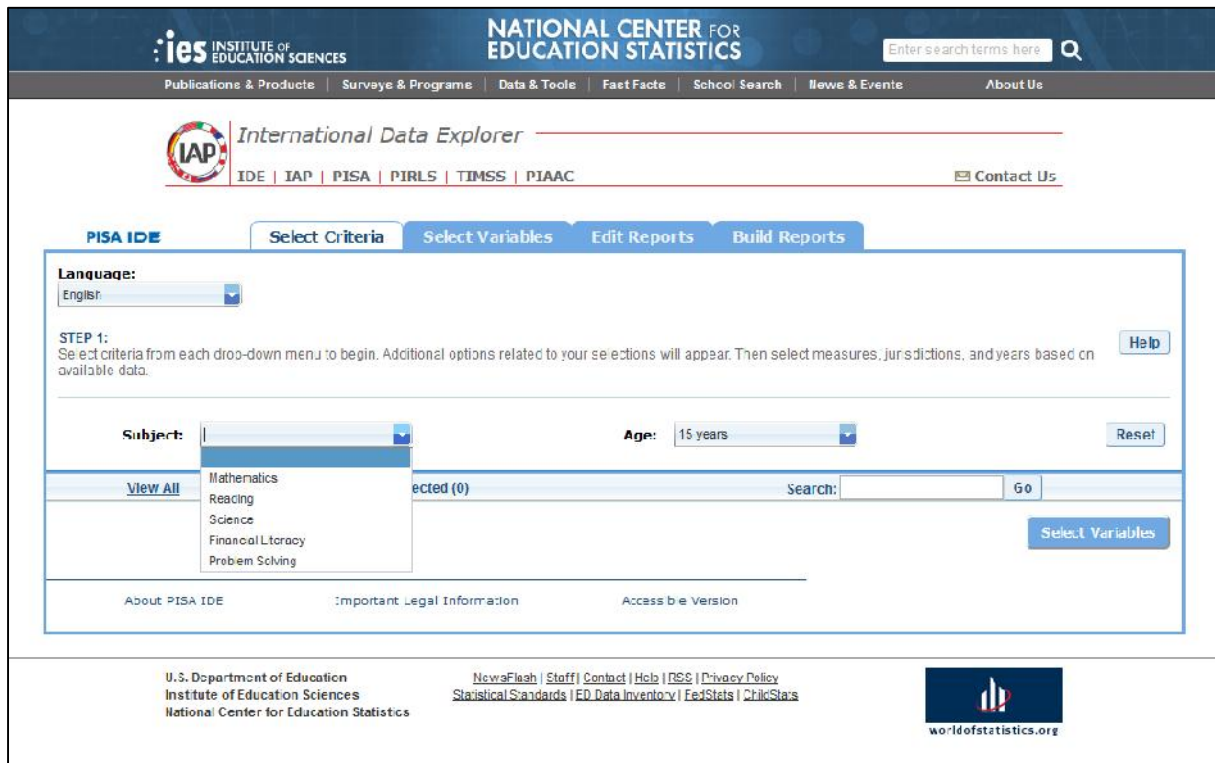


Figure 24. PISA Data Explorer

Source: retrieved from the National Center for Education Statistics (NCES) - <http://nces.ed.gov/surveys/pisa/idepisa/>

Users can use the International Data Explorer to create reports on aggregate variables from different domains in different countries and match them (e.g. with the OECD average). The tool is easy to use and users have only to agree to the terms of the NCES to enter. Within the same tool all different studies (PISA, PIRLS, TIMSS, PIAAC) can be accessed for creating different kinds of reports.

Though the tools can be considered a step in the right direction from convenience for the user there are some features lacking from the perspective of a researcher who wants to do deeper data analysis.

- The Data Explorer does not use metadata standards
- The items cannot be viewed as they are confidential
- The data is very high-level and aggregated like in a Public Use File (which is not surprising as everybody can enter after accepting the terms)
- Data cannot be exported for further use in a statistical packages (which would not make sense either way as the data is too aggregated)

PISA Data Explorer can be considered a first step, but should be extended by an access to Scientific Use Files, a data enclave or remote job submission. The Australian Council for Educational Research

(ACER) e.g. offers job submission for the PISA databases from 2000 to 2009 (<http://pisa2000.acer.edu.au/index.php>).

In Germany the national PISA databases are handled by the research data centre of the Institute for Educational Progress (IQB) in Berlin. Here access is only possible by travelling to the guest scientist seats at their facility after a request for accessing them has been submitted. On the IQB homepage the request for proposal, the PISA scale handbook, the variable handbooks and empty SPSS datasets to inform about the structure can be downloaded to prepare the analysis before travelling (<https://www.iqb.hu-berlin.de/fdz/studies>). This handling is usual for data with a high degree of confidentiality.

To sum up PISA data is available from multiple sources but the data management is currently in comparison to other domains not up to date.

### 1.3.3.2 OECD Programme for the International Assessment of Adult Competencies (PIAAC)

The structure of PIAAC has been described in a previous section and it can be considered to be the OECD's PISA version for adults though structure and content are quite different. Nevertheless both studies are successfully marketed as the OECD's flagship products for educational measurement and create a lot of recognition in this field. The data from both studies are therefore in high demand and create the need for data management. In the case of PIAAC the structures for data retrieval are therefore very similar to PISA. PIAAC uses the same technology of the international data explorer with the difference it is directly available at the OECD website. Likewise the U.S. PIAAC datasets have been incorporated into the Data Explorer on the NCES website. Figure 25 presents the International Data Explorer on the OECD website.

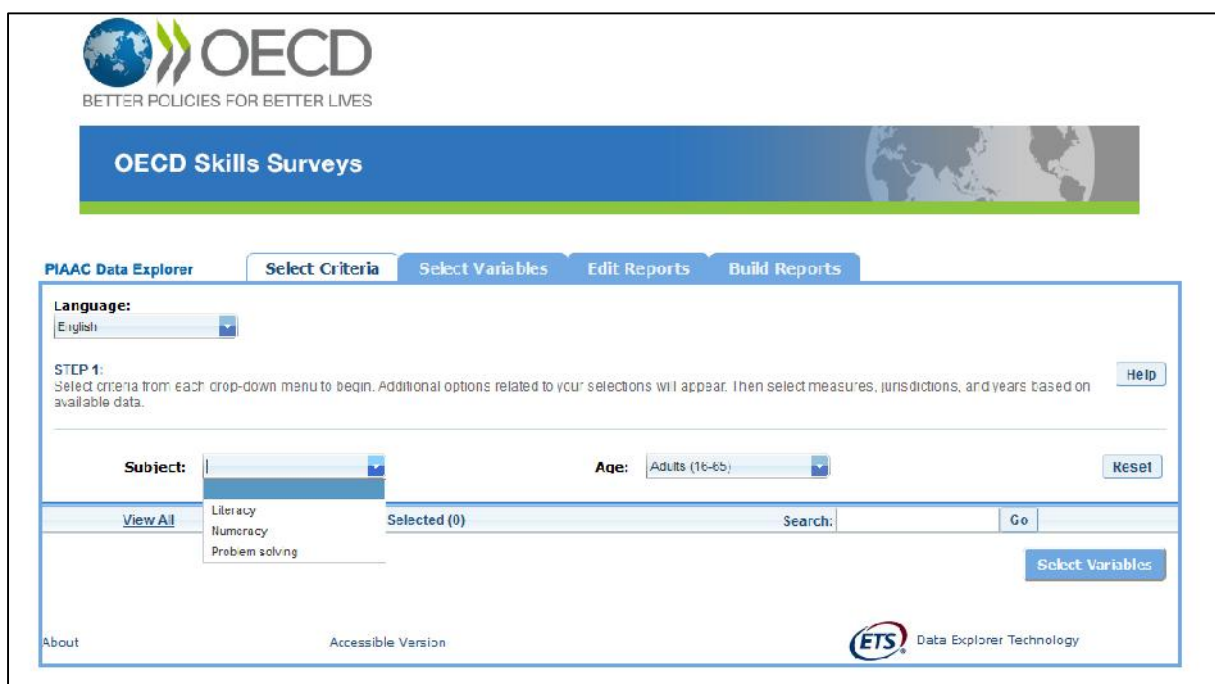


Figure 25. PIAAC International Data Explorer

Source: retrieved from the OECD - <http://piaacdataexplorer.oecd.org/ide/idepiaac/>

Additionally to the International Data Explorer the website PIAAC Gateway (<http://piaacgateway.com/datasets/>) offers the international Private Use File as well as a link to the GESIS data archive which contains a reduced version of the German dataset as well as the dataset from Cyprus. Especially interesting about the storing of PIAAC in the GESIS data archive is the study information has been coded in DDI 2.5 and DDI 3.1 as these are currently the institute standards for data documentation. This means it has to be processed by metadata enhancement as PIAAC originally does not use any metadata standard. Figure 26 shows a snippet of the DDI instance.

```
<ddi:DDIInstance id="gesis_ZA5845" agency="de.gesis"
version="1.0.0" versionDate="2014-11-30"
xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1"
xmlns:pd="ddi:physicaldataprod:3_1"
xmlns:pi="ddi:physicalinstance:3_1"
xmlns:c="ddi:conceptualcomponent:3_1"
xmlns:l="ddi:logicalproduct:3_1" xmlns:r="ddi:reusable:3_1"
xmlns:dc="ddi:datacollection:3_1" xmlns:a="ddi:archive:3_1"
xmlns:xhtml="http://www.w3.org/1999/xhtml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dce="ddi:dcelements:3_1"
xmlns:dc2="http://purl.org/dc/elements/1.1/"
xmlns:n1="ddi:physicaldataprod_ncube_tabular:3_1"
xsi:schemaLocation="ddi:instance:3_1
http://www.ddialliance.org/Specification/DDI-
Lifecycle/3.1/XMLSchema/instance.xsd">
```

```
<r:Citation>
```

```
<r:Title>DDI3.1 study level documentation for study ZA5845 Programme
for the International Assessment of Adult Competencies (PIAAC),
Germany - Reduced Version</r:Title>
```

```
<r:Creator>GESIS - Leibniz Institute for the Social
Sciences</r:Creator>
```

```
<r:Publisher>GESIS - Leibniz Institute for the Social
Sciences</r:Publisher>
```

```
<r:PublicationDate><r:SimpleDate>2014-11
30</r:SimpleDate></r:PublicationDate>
```

```
<r:Language>en</r:Language>
```

```
<r:InternationalIdentifier
type="DOI">doi:10.4232/1.11865</r:InternationalIdentifier>
```

```
<r:Copyright>All metadata from GESIS DBK are available free of
restriction under the Creative Commons CC0 1.0 Universal Public
Domain Dedication. However, GESIS requests that you actively
acknowledge and give attribution to all metadata sources, such as
```

the data providers and any data aggregators, including GESIS. For further information see <https://dbk.gesis.org/dbksearch/guidelines.asp> </r:Copyright>"

*Figure 26. DDI Lifecycle v3.1 Instance*

*Source:* GESIS, 2014 – retrieved from <https://dbk.gesis.org/dbksearch/>

This means the study information can be imported, parsed and processed by data management systems which are able to work with DDI. This enables GESIS to share these metadata information with other repositories like IAB, DIPF or CentERdata who are using the same metadata standard.

Unfortunately the cognitive items are not available due to confidentiality and the background questionnaire is only a PDF document and not included in the DDI instance, but the progress can be seen. GESIS is taking steps into a more process-oriented data management also in education which can be explained by their history in the social sciences and therefore familiarity in those standards. Available from the website are only the Public Use Files, for a more specific analysis GESIS has to be visited in Cologne by using a proposal procedure similar to the one of IQB for PISA.

### **1.3.3.3 The National Educational Panel Study (NEPS)**

The German National Educational Panel Study (NEPS) runs an own data management team which consisted from 2008 to 2013 from representatives out of four work packages (AP 12 – Methods and Statistics, AP 13a - Data Warehouse Requirements, AP 13b - Data Warehouse Programming and AP 14 - User Support). Figure 27 shows the overall structure of NEPS including the work packages

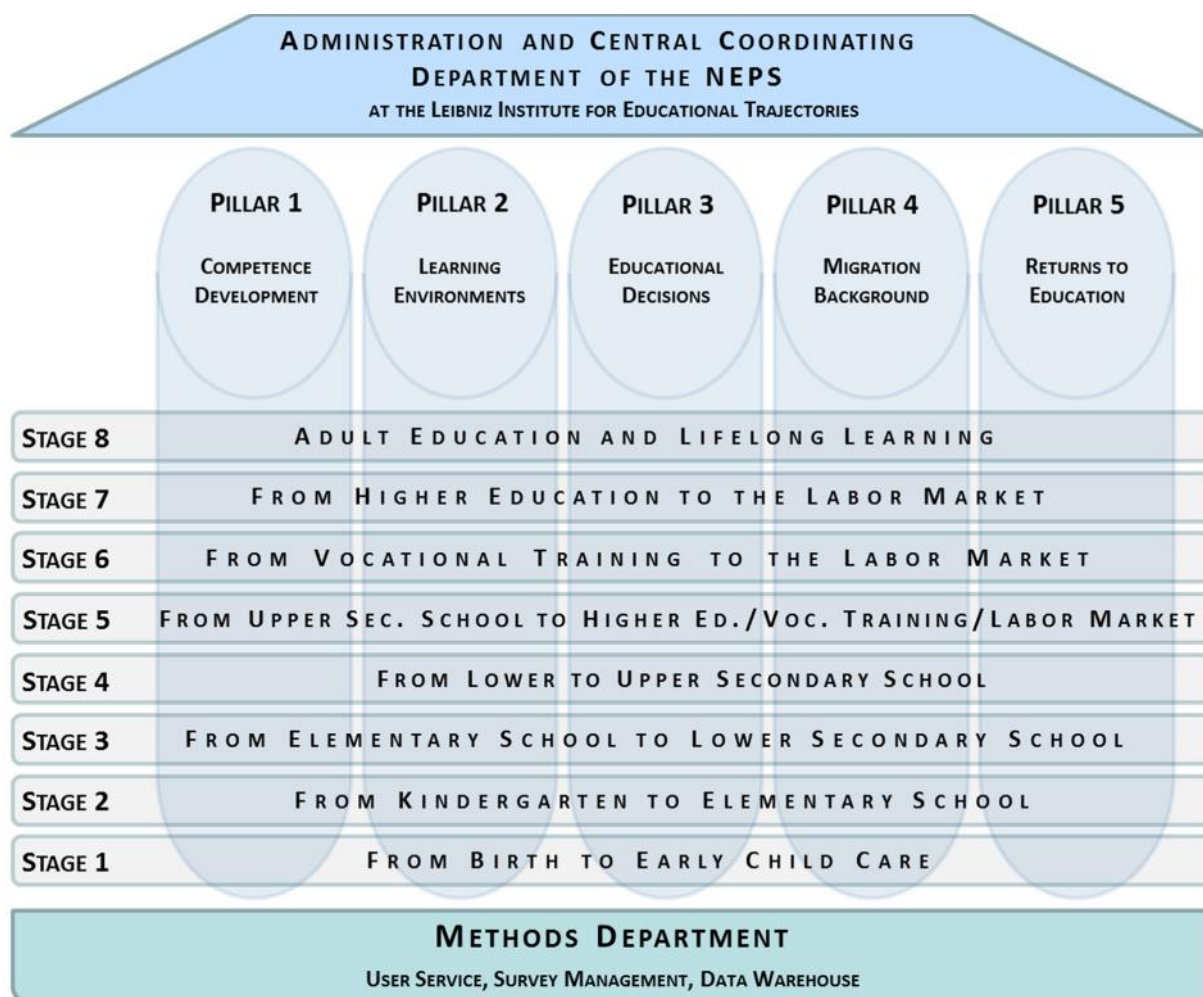


Figure 27. NEPS Structure.

Source: Blossfeld et al., 2011.

NEPS reaches due to its design as a longitudinal panel study with eight parallel cohorts a level of complexity which makes data handling on a normal file basis not feasible anymore. The idea was therefore to collect all data into a big data warehouse structure and generate the files for researchers based on their demands. In NEPS all study information is stored in a relational database where it can be extracted into different formats, e.g. it is possible to create a codebook out of this metadata by using reporting functionalities. The whole database contains about 100 tables with further information like instruments, universes, items, variables, schemes, attachments. Access to the data is possible via Scientific Use Files, guest scientist seats in Bamberg or a secure data service called RemoteNEPS (Barkow et al., 2012). Figure 28 depicts the multicohort sequence design of NEPS.

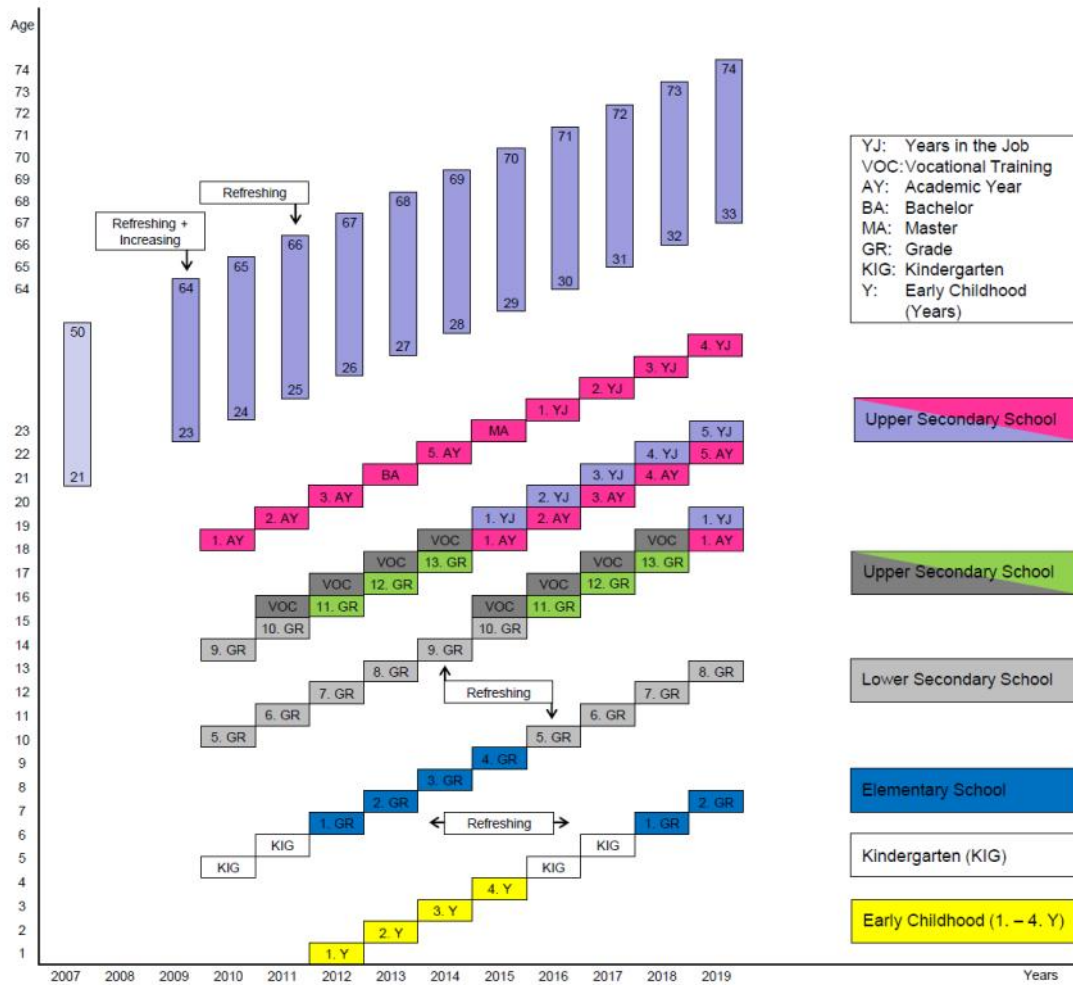


Figure 28. Multicohort Sequence Design

Source: Blossfeld et al., 2011.

Figure 28 shows the data collections from early childhood to adults in lifelong learning between 2007 and 2019. NEPS is supposed to run like the Socio Economic Panel for 20-30 years with 65.000 participants. On the other hand the NEPS is split between 25 different organisations which reside over many states in Germany which means also the design tasks for the instruments are quite complex. The sheer amount of metadata and data has therefore to be handled with structures which are capable of supporting such processes.

For this reason a large database structure has been created including support tools for data documentation and portal solutions for scientists to make the metadata more accessible. The multicohort sequence design has been modelled into a tool called the NEPSplorer where users can drill down through the cohorts to specific metadata they are interested in. Figure 29 depicts a screenshot from the NEPSplorer.



The screenshot shows the NEPSplorer tool interface. At the top, there are navigation tabs for 'Home', 'Project Overview', and 'Data Center'. Below this is a search bar and a 'Help' button. The main content area is divided into two tabs: 'Study overview' and 'Constructs'. The 'Study overview' tab is active, displaying a table with the following columns: 'Starting cohorts / Year', 'StudyDesign', 'SUF', and ten columns representing waves from 2007/08 to 2013/14. The table lists various study cohorts such as 'Early Childhood', 'Kindergarten', 'Regular School', 'Grade 5', 'Grade 9', 'First-Year Students', 'Adults', 'Organizational Reform Study in Thuringia', and 'GB Reform in Baden-Wuerttemberg'. Each row shows the study design (e.g., 'panel' or 'cross section'), the number of subjects (SUF), and the specific waves in which the cohort was observed.

Starting cohorts / Year	StudyDesign	SUF	2007/08	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14
Early Childhood	panel								
Kindergarten	panel	1788				4 - 580	4 - 818		
Regular School	panel	1788				4 - 580	4 - 818		
Grade 5	panel	2578				7 - 666	15 - 906		
Regular School	panel	2578				7 - 666	8 - 906		
Special School	panel						8 - 800		
Grade 9	panel	4007				7 - 846	14 - 1588		
Regular School	panel	4007				7 - 846	10 - 1584		
Special School	panel						6 - 150		
First-Year Students	panel	2417				4 - 917	2 - 1007		
Adults	panel	1724	1 - 43		2 - 924	1 - 623			
Organizational Reform Study in Thuringia	cross section	1810			3 - 171	3 - 170			
GB Reform in Baden-Wuerttemberg	cross section	1430				5 - 100	4 - 5		

Figure 29. Snippet from the NEPSplorer tool

Source: Retrieved from the Leibniz Institute for Educational Trajectories (LifBi)

<https://www.neps-data.de/en-us/datacenter/studydocumentation/nepsplorer.aspx>

Figure 29 shows the main window of the NEPSplorer tool which allows user to drill through multiple waves of the study to retrieve metadata and basic statistics about e.g. the data collection or the variables. It does not contain the datasets but shows the conditions under which they are available. Unfortunately NEPS currently does not support any metadata standards though the use of DDI for the questionnaires has been discussed for some time during development. From 2008 to 2013 NEPS also only used paper and pencil instruments for cognitive testing which is currently changing due to the work of AP 13c – Mode Effects under the lead of Ulf Kröhne. Sooner or later NEPS will run into the same challenges which this dissertation tries to describe.

### 1.3.3.4 The Database for the Quality of Schools (DaQS)

The Database for the Quality of Schools (DaQS) differs significantly from the large-scale studies discussed before. Essentially DaQS is a service offered by the German Institute for International Educational Research (DIPF) to make scales or answer schemes from studies like Pythagoras, Deutsch Englisch Schülerleistungen International (DESI) (see Klieme & Beck, 2007) or Internationale Grundschul Leseuntersuchung (IGLU) (see Bos, Lankes, Prenzel, Schwippert, Valtin, Voss, & Walther, 2005) available for the use in own research. DaQS stores complete item batteries from paper & pencil data collections including metadata and statistical values from the studies like standard deviation, variance and Cronbach’s Alpha. Figure 30 depicts a screenshot of the item battery “aggression” from IGLU 2001 included in DaQS.

**DaQS**      Datenbank zur Qualität von Schule  
Instrumente zur Erfassung von Schul- und Unterrichtsqualität

FACHPORTALpaedagogik.de

Startseite      **Details der Skala: Schulklima: Aggression**

Suche

    Einfache Suche

    Erweiterte Suche

Konstrukte

    Schema

    Liste

Studien

Fragebogenarchiv

Wie über uns

Hilfe / FAQ

Feedback / Kontakt

Rechtliche Hinweise

    Nutzungsbedingungen

    Datenschutz

Impressum

Skala: Schulklima: Aggression

Konstruktzuordnung in DaQS: [Klassenklima, Deviantes Verhalten](#)    [andere Zuordnung vorschlagen](#)

Theoretische Zuordnung in der Ausgangsstudie: Schulklima

Studie: [IGLU 2001 - Internationale Grundschul-Lese-Untersuchung](#)

Ursprung: Eigenentwicklung

Quelle: Bos, Wilfried / Lankes, Eva-Maria / Prenzler, Manfred / Schwippert, Knut / Valtin, Herate / Voss, Andreas / Walther, Gerd  
IGLU - Skalenhandbuch zur Dokumentation der Erhebungsinstrumente, Münster: Waxmann 2005, ISBN: 978 3 8300 1681 2  
[Titel in FIS-Bildung](#)

Theoretischer Hintergrund:

    Zielgruppe(n): Schülerinnen der Primarstufe

    Erhebungszeitraum: 2001

    Anzahl der Items: 6

    Anmerkung: Die Skala ist rekodiert. Ursprünglich war die Kodierung 'ja' (1); 'nein' (2)

    Einleitender Text: Sind irgendwelche dieser Dinge im letzten Monat an deiner Schule passiert (soweit du es mitbekommen hast)?

    Items:

Item-Formulierung	Mittelwert	Standardabweichung	Trennscharfe
In der Schule ...			
ist mir etwas gestohlen worden.	0,22	0,42	0,32
ist jemand anderem aus meiner Klasse etwas gestohlen worden.	0,44	0,50	0,38
bin ich von einem anderen Schüler immer wieder geärgert worden.	0,46	0,50	0,38
ist jemand anderes aus meiner Klasse von einem Schüler immer wieder geärgert worden.	0,60	0,46	0,36
bin ich von einem anderen Schüler geschlagen oder verletzt worden.	0,29	0,45	0,40
ist jemand anderes aus meiner Klasse von einem anderen Schüler geschlagen oder verletzt worden.	0,52	0,50	0,47

Figure 30. Snippet from Database on School Quality (DaQS)

Source: Retrieved from [http://daqs.fachportal-paedagogik.de/search/show/instrument/1580\\_37](http://daqs.fachportal-paedagogik.de/search/show/instrument/1580_37)

Figure 30 shows a search result about an item in German about the aggression at schools from the IGLU 2001 study. As can be seen here DaQS stores the item battery including study information, target population, content and statistical aggregated values from data collection. The battery can be exported into PDF and ODF format. Furthermore users can give feedback about the battery.

Interesting about DaQS is it uses next to the concepts which were originally used in the study an own concept tree to sort item batteries and scales into common categories. Figure 31 shows the concept model of DaQS.



Figure 31. DaQS Construct Model

Source: Translated from DaQS - <http://daqs.fachportal-paedagogik.de/constructlist/graphic>

Though DaQS is an interesting approach for storing items and currently serves as an inspiration to build the DIPF Repository within the Research Data Centre for Education it is lacking some features to make it usable also for computer-based items.

- DaQS does not use any metadata standards
- The structure was built to reflect paper and pencil instruments, not for computer-based assessment
- DaQS only contains metadata, it cannot handle datasets

DaQS was originally designed for handling metadata for educational sciences on a high level perspective to forward researchers who are interested in the datasets and items to the respective organizations. In the terminology of the German Council for Social and Economic Data (RatSWD) this would fit to the organizational structure of a data service centre (Solga and Wagner 2007). In the meantime the DIPF RDC decided to integrate DaQS into a new service structure for quantitative and qualitative metadata called Rogatus Repository which supports DDI and SDMX as metadata standards and also references to datasets within a database structure. DaQS can therefore be considered as a first step into a modern repository structure for the educational sciences.

#### **1.3.3.5 EDK HarmoS School Monitoring**

HarmoS is an initiative by the Swiss Directorate for the Cantonal Directors of Education (EDK) to build up a national school monitoring system and harmonize items and learning material between the different cantons in Switzerland. Similar to the German states in Switzerland every canton has its own educational minister and also education system. In the last years some cantons decided to implement school monitoring using different assessment tools. Some of them were produced by commercial vendors (e.g. Arcadix Testdrive) see: <http://www.arcadix.com/>). The national government decided these approaches should be centralized so items and other data are available for a bigger research audience. In a second stage of the project it is planned to make the items available for educators on a structures for school servers where they can re-use the items for teaching purposes.

For these two stages an item bank called TIPO (TAO-based item portal) was developed where different assessment platforms (currently TAO, CBA Itembuilder and paper&pencil material) can be attached to the central database so computer-based and paper-based cognitive items become available in the first stage to item developers, researchers or policy makers. TIPO is not able to migrate items between different platforms like TAO and CBA Itembuilder automatically, but users are able to see which items are available for which platform and then are able to open the authoring tools to rebuild them in other platforms as well. In the long run the goal is also to include other assessment platforms used in Switzerland like Arcadix Testdrive, Trifork or Questify. Figure 32 displays the concept of the TIPO system.

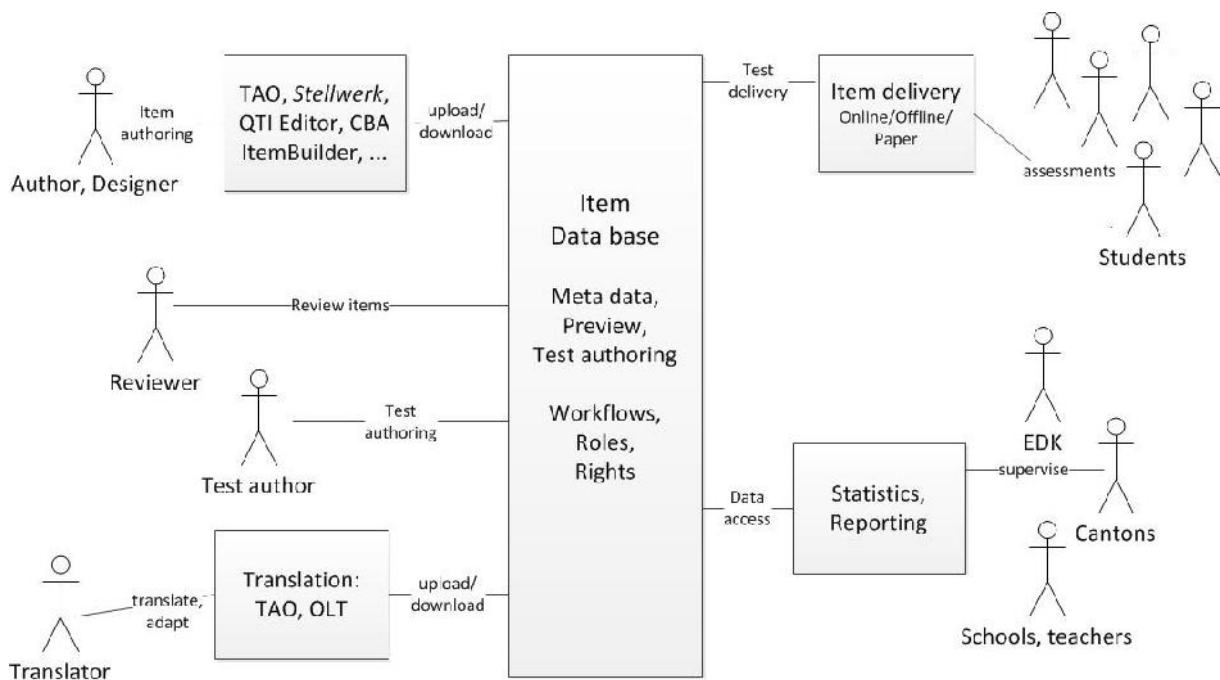


Figure 32. Conceptual structure of TIPO

Source: Rölke, 2013.

In a certain perspective TIPO can be considered to be a workaround regarding the lack of metadata standards in computer-based assessments in the educational sciences. As items cannot be easily documented and re-used between different tools TIPO acts as central database for different assessment tools which are otherwise not able to communicate with each other. TIPO also supports two metadata standards. For documenting the concepts of items for different target groups it uses the Learning Object Model in the country adaptation for Switzerland (LOM-CH). On the other hand an instance of TAO is embedded which allows the use of items basing on QTI. Nevertheless TIPO only makes the management of computer-based and paper&pencil items easier. It does not solve the overall problems of re-usability. Furthermore it also does not store any information about the data collection or resulting datasets as the test delivery is still performed on the original assessment systems. Therefore it also does not act as a repository for the resulting datasets. These functionality could be implemented in the long run but is currently not available.

Again like with other systems TIPO can be considered as a first step and serve as a basis for further developments, but it is currently not able to model the whole assessment cycle.

### 1.3.4 Different software packages in computer-based assessment and their capabilities for metadata management

In the last section different large-scale studies and the software they are using have been discussed. As some of the names are appearing in multiple context it is necessary to look at these software tools in more detail to discuss if they can be adapted to use metadata standards more broadly.

### 1.3.4.1 Testing Assistee par Ordinateur (TAO)

Testing Assistee par Ordinateur (TAO) is an open source software initiative started by the University of Luxembourg and the Centre de Recherche Public Henri Tudor (CRP) in 2002. The idea behind TAO is to develop a computer-based assessment system which bases on an abstract item model so items can be reused in a world of changing technologies. Before TAO computer-based items were mostly developed by programming them directly in a specific programming language for a distinctive computer platform. From a data management perspective this is difficult to handle as with the next technology step the computer platform might become obsolete and transformation costs to reprogram the item on another system are simply too high. By these proceedings many items got lost from the 1980s until the beginnings of the 2000s. TAO uses a RDF-based item model which abstract item descriptions and generates the items using a compilation and rendering process. If new technology enters the scene the renderer can be replaced and the item can be regenerated on the new platform. In the case of TAO this happened with the change from Flash technology to Rich Internet technology during the TAO Transfer project which was done collaboratively with DIPF as the new third project partner. Figure 33 shows the structure of TAO as an overall system.



Figure 33. Structure of TAO

Source: Latour and Martin, 2007.

As can be seen from this graphic TAO uses a life cycle model to describe its internal processes. Furthermore it is also interesting TAO considers itself to be a complete tool suite with supports test

development from candidate organization to test development and test delivery up to reporting. The proceeding from previous testing rounds can be used for the next cycles of testing.

Though TAO bases on an interesting concept and is quite successfully used in different prominent large scale studies like PISA and PIAAC there are some drawbacks. The process of rendering items from an abstract model makes the overall system in overall performance slower than a system which uses a native programming of computer-based items. Especially this became a challenge in testing larger groups as the problem increases with the number of users. In PIAAC R1 countries complained about the slow performance of the items in testing (Zabal et al., 2014). The Hungarian school monitoring project „Developing Diagnostic Assessments” at the Institute for Educational Sciences at the University of Szeged tried to do the data collection using TAO but it was not suitable for the numbers of students so a decision was taken to develop an own system (Csapó, 2011).

The performance problems were reduced in the meantime. On the one hand the advancements in hardware performance counterbalanced the issues and on the other hand TAO has been performance optimized. As the rendering of items from the abstract model was costing too much performance TAO 2.X uses pre-compilation processes and checking out of the RDFs into a relational database model to save CPU and hard drive processing time. The newer versions therefore are considerably faster than older versions.

From version 2.X TAO supports QTI as a metadata standard and can therefore re-use items from other compliant QTI platforms like Questify. As TAO uses RDF for metadata it would also be possible to implement other metadata standards into the platform. Unfortunately this has not happened yet so the metadata model used in TAO for Study Concept, Delivery and Publishing is proprietary and very basic. It does not fulfil demands from data management like they are specified by the lifecycle model of DDI or GLBPM. Nevertheless it would be possible to integrate those into TAO from the RDF model. Another advantage is as TAO is Open Source it can be extended by other interested parties. If an agency is interested in DDI it would be possible to integrate the standard into TAO by extending the RDF model and also adapt the source code of the workflow engine to make it possible to import DDI instances. The same proceedings can be performed for other metadata standards as well.

#### **1.3.4.2 CBA Itembuilder**

CBA Itembuilder was originally developed by DIPF and Softcon in 2007 as the Hypertext Builder in the context of an extension of PISA 2009 called Electronic Reading Assessment (ERA) as a plug-in for TAO mentioned in the section before. The concept was to deliver an authoring tool which is able to design simulated web browsers by using pre-defined templates. At this time TAO only contained a very basic authoring tool for simple item types. The success of the Hypertext Builder in PISA led to several further developments of the tool which was rebranded as the CBA Itembuilder. The name change was also based on a change in technology happening in the background. Like TAO CBA Itembuilder also uses an abstract item model to author items and can therefore use different renderers to create the graphical content. From Hypertext Builder to CBA Itembuilder this technology was changed from Adobe Flash to Rich Internet Technologies based on a JAVA Eclipse framework as Flash caused to many problems on Linux (used e.g. for delivery in PIAAC) and the slowly upcoming mobile platforms (e.g. Apple devices do not support Flash).

Some more features were extended into CBA Itembuilder.

- Support for translation by using the XLIFF format
- Support for audio and multimedia content
- Design of Dynamic Automaton (MicroDYN) items for dynamic problem solving scenarios (e.g. for PISA 2012)
- Design of Finite Automaton (MicroFIN) items for problem solving scenarios using finite state machines for complex workflows
- Templates for web pages, mail clients and business software (e.g. word processor, spreadsheets)
- Recording of audio and video as an answer mode

The items designed by CBA Itembuilder can still be plugged into TAO, but in the meantime the editor comes with an own delivery platform called the Execution Environment (EE) so it can also be used stand-alone. In total the CBA Itembuilder takes its main focus on complex problem solving items. Figure 35 displays the screenshot a MicroFIN item.

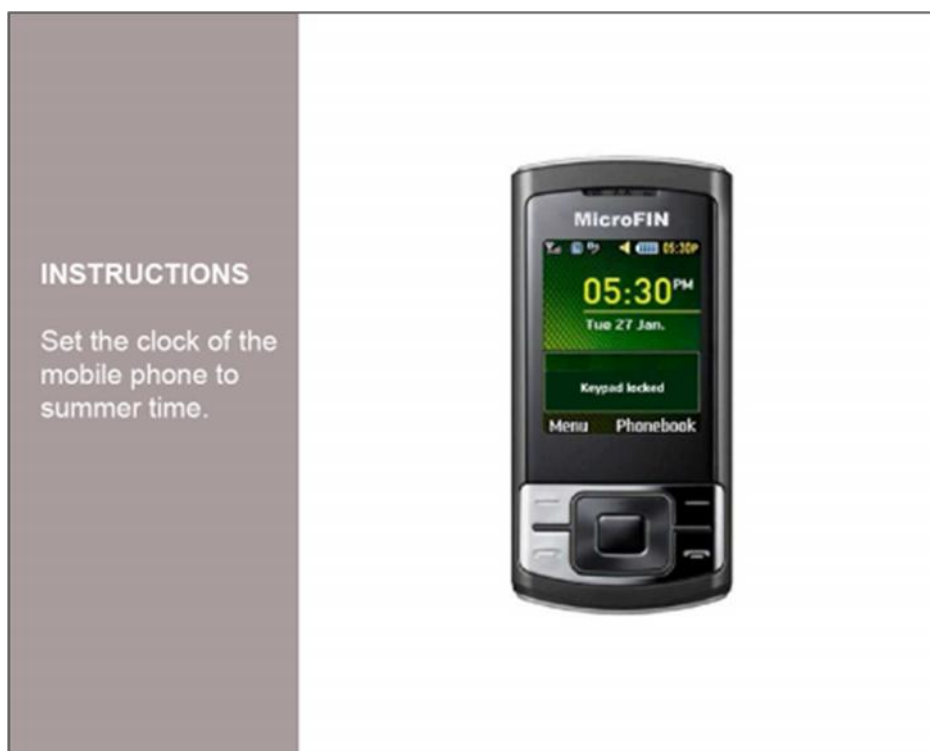


Figure 34. MicroFIN item in CBA Itembuilder

Source: Rölke, 2012.

The task in this item is to set the clock of the mobile to summer time. Therefore the user has to solve a foreign scenario using steps pre-defined by the finite state machine (e.g. unlocking the mobile, change to the menu, search for the settings, change the time format) which can only be done in a certain sequence. If the simulated mobile is not used for a certain time it will lock itself again. CBA Itembuilder uses templates and an editing functionality for the rules of the finite state machine and scoring to produce such items.

Though CBA Itembuilder is an impressive tool from its functionalities it was not designed to perform any kind of data management:



- It does not support any kind of metadata standards in regards to survey design or item development
- It does not contain any means to document background information of surveys at all
- A functionality like in TAO to attach own metadata to an entity is missing
- The items are rendered from a propriety object model which is very complex and not designed to be used with other tools
- The scoring rules and finite machine states are basing on an own development

Introducing a metadata model will be very difficult for CBA Itembuilder simply because of the complexity of the problem solving items it was built for. A workaround to keep at least the survey documentation is to use TIPO mentioned in a previous section as a front end item bank. Nevertheless there is currently no tool on the market which can author these kinds of complex items. As metadata management is currently struggling to preserve much less complex item types for computer-based assessment building a model for these complex items might be a very long-term endeavour.

### 1.3.4.3 Questify

Questify is a commercial assessment platform which was developed by Cito – a research institute in the Netherlands - since 1999 but only got this brand name in 2012 (see <http://www.questify.eu>). From its long history it supports the authoring of a lot of different item types and also has all the functionalities for candidate and delivery management. It therefore fulfils the needs according to industry standards like platforms from competitors like Questionmark or MyItemWriter from Prometric. Originally the dissertation was not supposed to include commercial platforms as they very often promote indirectly the idea of a “vendor lock-in” (i.e. by default the items cannot be exported to another platform so the customer cannot change the software) which completely negates the idea of data management and were normally designed for high-stakes testing for certification processes in test centre scenarios (e.g. Microsoft or Oracle certification exams). Therefore they address a completely different use case than computer-based testing for scientific research.

Nevertheless, Questify is an interesting exception. Though it is a commercial platform it supports QTI and LTI to enable the import and export of items to other platforms. Furthermore Cito became one of the owners of Open Assessment Technologies (OAT) who are the producers of the open source platform TAO in 2013 and since then promote a dialogue to bundle both products actively together. At the European Association of Test Publishers conference 2014 (E-ATP 2014) in Budapest Cito and OAT showed a technical demo how both systems can interact by using the same standards. Due to LTI they also took Moodle as an E-Learning platform into the approach. Figure 36 displays a screenshot from the technical setup.

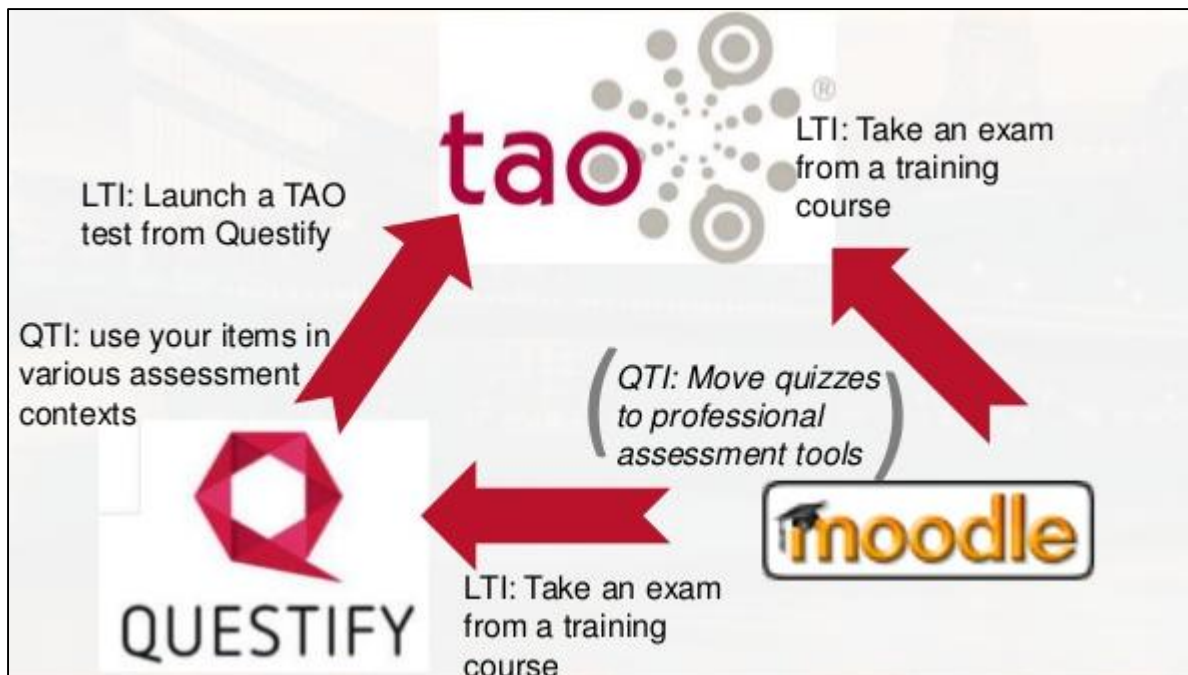


Figure 35. Connecting Questify to TAO and Moodle

Source: Molenaar and Garrard, 2014.

From a data management perspective this use case is very interesting as it uses two metadata standards to connect three different platforms and shows the possibilities as well as the problems such an approach conveys. Molenaar and Garrard (2014) report they got this prototypical approach running but faced a lot of problems as the standards do not integrate this seamlessly.

- Only three simple interactions of QTI were used (out of 21 possible)
- Rendering of interactions in QTI is vendor-based therefore the graphics look different in the systems
- These effects have to be psychometrically validated which has not happened yet

Molenaar and Garrard (2014) therefore showed a combination of QTI and LTI could “do the job”, but they recommend further developments as the integration is still very basic. This matches the considerations about QTI and LTI defined in the previous sections. The practical application shows QTI and LTI can make the integration process between different platforms easier for the application in the same time period, but the neither solve the integration problems fully nor offer a perspective for the long-term preservation of items (if the item storing in the present is already difficult it is not feasible for storing items from the past). The integration via QTI and LTI between Questify and TAO is also interesting from a technical standpoint as both platforms are not compatible from their code base. Questify is a .NET application written in C# while TAO uses Apache, MySQL and PHP as its base. Items can therefore not be ported directly by programming adaptations.

Nevertheless the processes of harmonizing two platforms show the need of more interaction between vendors as the customers obviously are not willing to pay for the re-development of items in migration scenarios. This is at the same time a possibility for data management to extend the demand also with long-term preservation scenarios.

### 1.3.4.4 Electronic Diagnostic Assessment System (eDia)

The Electronic Diagnostic Assessment System (eDia) was developed in the second phase of the Hungarian school monitoring project „Developing Diagnostic Assessments” at the Institute for Educational Sciences at the University of Szeged (Csapó, 2011). It is still constantly extended in other projects, e.g. as an online diagnostic system for primary school (e.g. Csapó, L. Rincz, & Molnár, 2012). In the first phases of the “Developing Diagnostic Assessments” project first TAO and later CBA Itembuilder including the Execution Environment (EE) were used. Obviously the two systems could not cope with the number of students a national school monitoring needs and therefore the development of a leaner platform was necessary. The eDia platform therefore has very similar item types like TAO or CBA Itembuilder including the MicroDYN dynamic problem solving items described in a previous section. Figure 37 shows a screenshot of eDia.

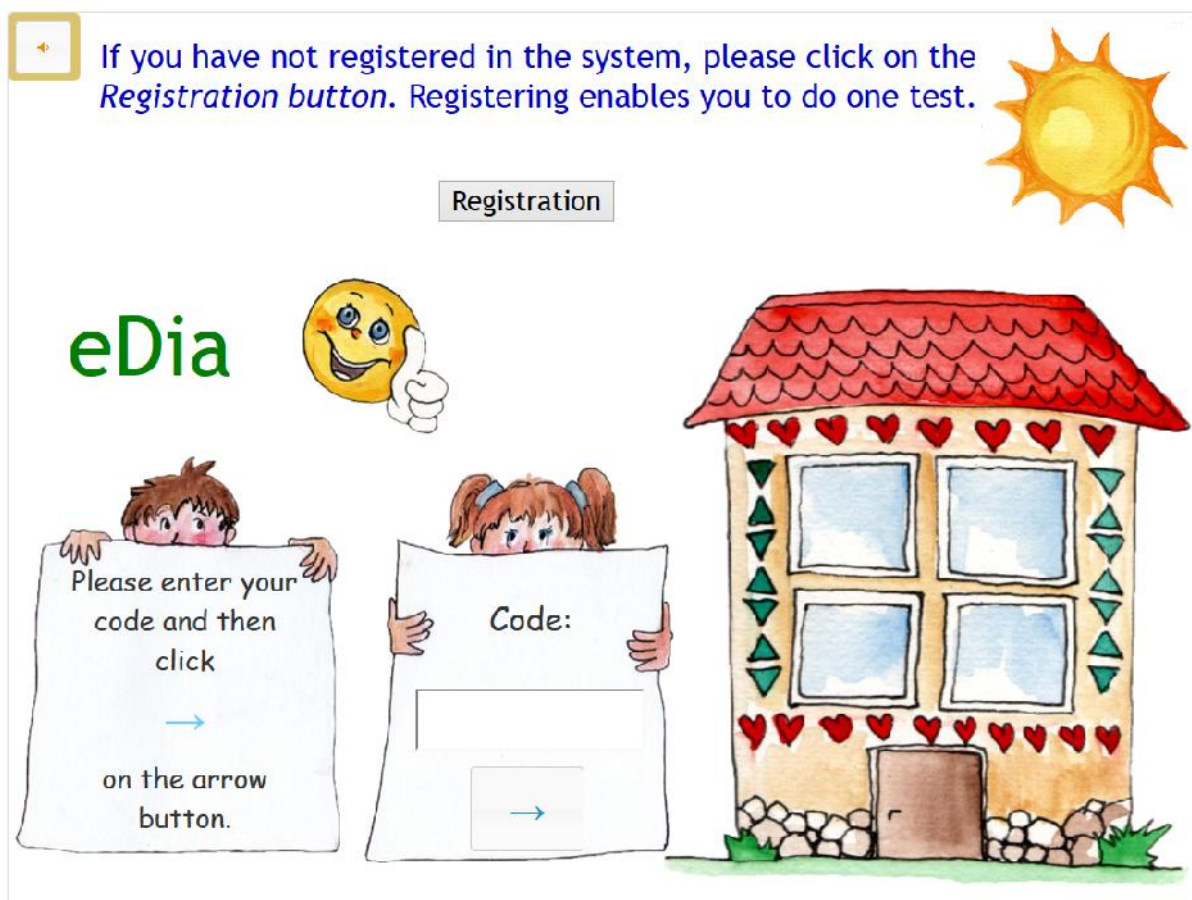


Figure 36. Item from eDia

Source: Csapó, Molnár and Nagy, 2014

eDia solves the performance problem caused by the object generation in TAO and CBA Itembuilder by being closer to programmed items though it uses item authoring capabilities as well. Nevertheless the item authoring leads to direct code and not to a representation in an object model. This trade off enables eDia to be leaner and much faster but also limits the possibilities of migrating content to other platforms or newer technologies. Currently eDia does not support any kind of metadata standards or data management and also is not based on an object model for items. The risk of not being able to move to a newer platform is therefore even higher for eDia than the other platforms described before as the

rendering capabilities cannot be changed this quickly due to the close connection between items and delivery without abstraction layer. Extending eDia with such capabilities would therefore be a recommendation if the items were to be preserved on a longer basis.

### 1.3.5 Conclusion

There are metadata standards for computer-based assessment and they are also implemented into different platforms. Nevertheless the standards and software products are not sufficient for the needs of data management if the content is supposed to be stored for research on a middle-term or long-term perspective. Larger studies like PISA, PIAAC and NEPS have data management processes in place but they rather archive the results than store the items and instruments in a computer-based format. It can therefore happen the paper&pencil material like in NEPS can be reconstructed also in 10-20 years but the computer-based items might get lost as the processes to preserve them over time simply do not exist and make a full re-programming necessary if they are supposed to be re-used and understood at a later point of time.

Decision makers therefore should be aware of this problem especially as the development of calibrated computer-based items is a costly endeavour. The question is if scientist are willing to take part in these processes which leads to the following research questions.

### 1.3.6 Research questions

Currently it is not known how much scientists know about data management processes for computer-based assessment and their willingness to participate in such tasks. Data management is time intensive and puts a burden on both scientists as well as the developers of assessment software as the software has to be extended in various respects from survey documentation to item development and finally to data analysis.

This leads to the following research questions:

- Which categories of different metadata standards do scientist know?
- How can those metadata standards be integrated into their daily work?
- What are the requirements of educational scientists towards metadata standards?
- What is missing to provide a high degree of re-usability (e.g. items, instruments)?
- What is missing to describe as much as possible for researchers who want to do a secondary analysis of the data?
- How could a model for storing a computer-based study including all metadata and paradata in the long term look like?
- Do researchers in the educational sciences need a work-around in the meantime?
- How can especially complex item types like simulations in complex problem solving (e.g. MicroFIN, see *Greif & Funke, 2010*) be modelled in a meaningful way?
- How can the results be stored for long time preservation or archiving?

These questions can be answered by analysing input given by educational and social sciences researchers. For this reason a survey was started to get opinions on those topics. The next major section will therefore describe the survey and its design.

### 1.3.7 Hypotheses

Before the start of the survey between the research questions and the field work also some hypotheses with underlying can be formulated on basis of the literature review.

- We expect more familiarity with data management and metadata standards from social scientist or researchers working in both fields than from pure educational scientist.

This is based on the higher qualitative level of the social sciences metadata standards and the higher number of institutions available for this purpose. Educational scientist have less opportunities to use a proper standard or to find and consult an expert.

- We expect a general positive attitude towards metadata management as researchers understand the advantages, but a mediocre or even negative attitude towards current standards and their implementation.

Based on Punch (2009) secondary analysis of datasets generates interest of scientist and furthermore these techniques are used by large scale studies. Therefore the message of usefulness should have reached most scientists. Nevertheless the previous sections show what metadata standards especially in the educational sciences are lacking. This should create a certain unhappiness only to be counterbalanced if the scientist do not know what they are missing.

- We expect most metadata standards will only be known by name or not at all by most researchers. There will not be many experts.

Normally data management (though it is demanded from stakeholders) is a process most scientists are not familiar with. They might only know those standards if the software they are using for their research supports it in the background or as a 'sales argument' of research data centres offering their services. It cannot be expected researchers are familiar with the standards as they normally do not fall into their domain.

- We expect significant interest of educational researchers in secondary analysis of datasets.

As surveys and field work become more and more expensive researchers get more interested in re-using previous datasets or instruments.

- We expect younger researchers to have a more positive attitude towards sharing of their research data than more senior researchers.

Re-using of datasets needs an attitude of sharing data which has only be promoted in the recent years. Before a high linkage between data and primary researcher existed. Therefore we expect some traditional views ("this data is mine") from scientist which started their careers with a different paradigm.

- We expect the average time for data management and documentation in studies and surveys not to be sufficient for generating high level metadata.

This hypotheses bases on the professional experience of the other and the missing elements of most metadata standards. A normal ingest of research data into a data archive consists of a codebook describing the core variables and the dataset in the format of a statistical package (e.g. SPSS, Stata) in different levels of quality. There are normally no further information (e.g. instruments, reports, analytical scripts).

*“Without data you’re just another person with an opinion.”*

(W. Edwards Deming)

## **2. Survey Design**

The previous section showed different metadata standards, different computer-based assessment products and data management procedures in large-scale assessments. It was also shown none of the current processes are fully sufficient to build a complete workflow for storing computer-based assessment instruments. To answer the research questions and to lay the foundation for further research and upcoming better software tools for data management it seems necessary to get more input from the involved party. Therefore as part of this dissertation a survey has been designed which addresses the needs of data producers as well as researchers who do secondary analysis. The following sections will show methodological considerations about the format and design of the survey. Afterwards the content of the questionnaire and its construction will be shown.

### **2.1 Methodological considerations in questionnaire design**

The first question about a survey about metadata management for educational sciences is the mode in which it should be performed. Basically the following modes are available (Bethlehem & Biffignandi, 2012).

- Paper and Pencil Interview (PAPI) – paper questionnaire conducted in house by an interviewer
- Computer-Assisted Personal Interview (CAPI) – computer-based questionnaire conducted in house by an interviewer
- Computer-Assisted Web Interview (CAWI) – web survey filled out by the participants themselves
- Computer-Assisted Self Interview (CASI) – computer-based questionnaire filled out by participants in a facility, sometimes observed by audio or video
- Computer-Assisted Telephone Interview (CATI) – computer-based questionnaire conducted by an interviewer via phone

As this dissertation is not directly connected to a research project the data collection has to be organized in a feasible way, e.g. the use of trained interviewers in an international context is not possible. This eliminates the modes PAPI, CAPI and CASI. Though the costs for CATI are considerably lower than sending out interviewers the costs for a phone survey are still too high. The only affordable solution is therefore conducting a web survey or CAWI mode though web surveys have clear advantages and disadvantages.

Wright (2005) lists the following advantages.

- Web survey allow access to unique populations
- Web surveys save time for researchers
- Web survey save costs for researchers

The drawbacks on the other hand are:

- Sampling issues (e.g. pre-selection bias, no representative sample possible)
- Self-reported demographic information is not reliable
- Access issues (e.g. invitations to mailing lists are considered rude or spam)

Web surveys are the cheapest survey mode, but also the one with the most methodological problems and lowest response rates. The next section therefore references best practices in web survey design to make the questionnaire more applicable and attractive to the participants. Invitations for web surveys via mailing lists are challenging and therefore the graphical aspects play an important role.

### **2.1.1 Web Survey Design**

This section will discuss the design considerations which were taken for the web survey to make it attractive for the participants. The importance of a good web survey design cannot be underestimated.

Couper (2000) identifies the following reasons for observing design standards.

- Respondents extract meaning and seek guidance on how to respond not only from the question wording itself, but how the question and response options are presented
- Other design and layout features may distract from – or even interfere with – the question-answering process
- In this way, design may directly or indirectly affect measurement error or the quality of answer we obtain
- Poor design may also increase nonresponse or breakoffs

Therefore a decision regarding survey software, survey settings and answer modes has to be planned before designing the questionnaire itself.

#### **2.1.1.1 Survey software**

There is a big market for different web survey systems with different costs. The author conducted the research for his master thesis on the statistical analysis on the value of IT certifications in businesses in 2007 (Barkow, 2007) with a software package called Unipark/Globalpark. The software provided very good survey capabilities including an editor for branching and routing rules and some analytical functionalities. Nevertheless, Unipark/Globalpark is not a free product and the subject of data management and focus on reusability makes an open source product more attractive. The most prominent open source web survey system as of 2014 is called Limesurvey and is available under a GNU General Public License.

Limesurvey started in 2003 as PHPSurveyor and is currently available in version V2.05 as a free product which can be installed on any Linux-based web server and offers the following features (Schmitz, 2012).

- Unlimited number of surveys at the same time
- Unlimited number of questions in a survey (only limited by the database)
- Unlimited number of participants in a survey
- Multi-lingual surveys
- User-management



- More than 20 different question types
- WYSIWYG HTML editor
- Quotas management
- Integration of pictures and movies into a survey
- Creation of a printable survey version

Next to those features it also offers an expression manager for branching and routing rules plus basic statistical analysis. The biggest advantage nevertheless is Limesurvey as an open source product can be adapted by programmers or the community with own functionalities. The author therefore tested Limesurvey in August 2014 for its functionalities and the decision was taken it was sufficient for the purposes of the web survey. Figure 38 shows a screenshot of the main user interface of Limesurvey V2.05.

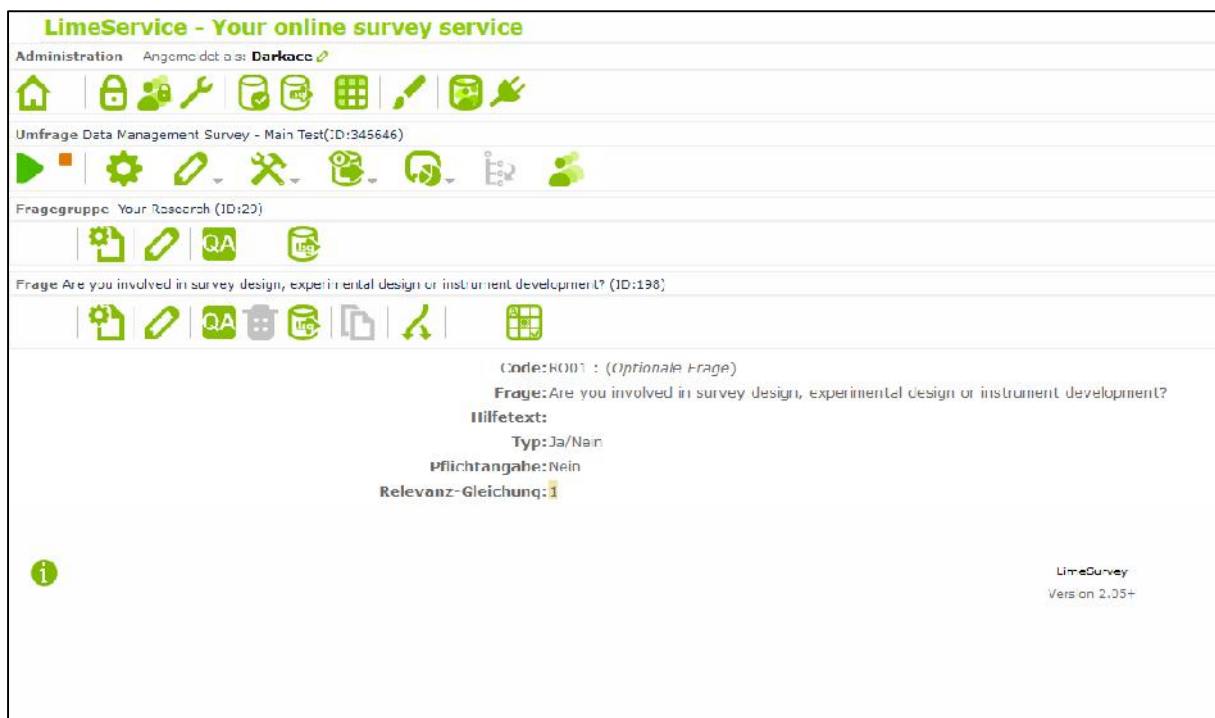


Figure 37. Limesurvey V2.05 Main Administration Window

Source: retrieved from <http://www.limeservice.org>

Limesurvey can be installed for free on own web servers in the internet. The producers of the tool nevertheless also offer managed servers for surveys at low costs. The payment is connected to packages of completed interviews. The author bought a package of 1000 participants for different surveys for 50€ which is a price of 5ct per completed interview. A hosting on an own server at the home organizations German Institute for International Educational Research (DIPF) or the University of Szeged would have led to higher costs.

The system was configured not to use the default Limesurvey template, but a community-based template called SkeletonQuest (<http://extensions.sondages.pro/skeletonquest/>). SkeletonQuest uses a responsive design for questionnaires which means it is also optimized for mobile devices. Wijnant and De Bruijne (2013) state 7% of users do web surveys by tablets and 2% by mobiles even if they are told the survey is not supposed for mobile devices. The numbers are growing from year to year. Using a responsive design therefore will accommodate users who prefer to use a mobile device and furthermore might make

the survey overall more attractive. Limesurvey offers the functionality to check the layout for different screen sizes. Figure 39 displays an example for mobile phones.

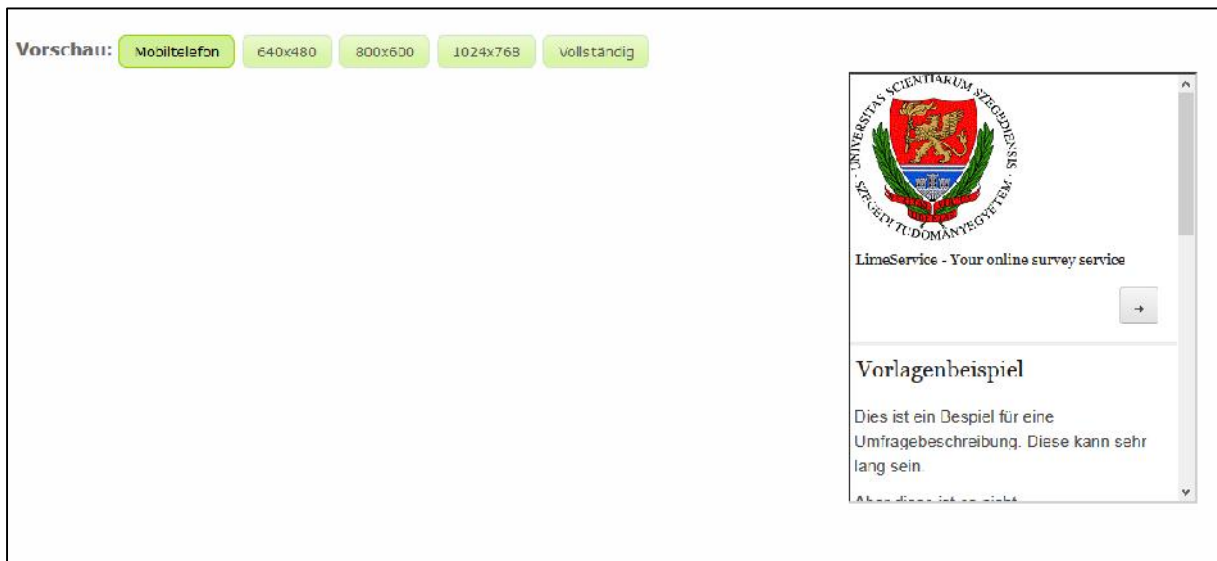


Figure 38. Snippet from Limesurvey Template Editor

As can be seen SkeletonQuest adds own buttons for mobile devices but the screen resolution for a questionnaire is still not ideal. Nevertheless the resolution is fitting for tablets. As the survey will be performed fully anonymous due to data protection laws there will be no results regarding browser usage especially concerning mobile browsers.

For the metadata management survey the SkeletonQuest template and Limesurvey were modified in two regards.

- Logos from the University of Szeged to provide a scientific background
- Normally Limesurvey offers a “Clear Survey and Exit” button which was removed from the software code (as it deletes also partial responses)

With these changes the template was configured for data collection.

### 2.1.1.2 Survey settings

Next to the general installation of the software and the template Limesurvey offers a lot of individual settings how the survey engine behaves while the survey is running. Here are some decisions which were taken due to survey design principles:

- Language: English

Though Limesurvey supports multiple languages the survey will only be performed in English due to the planned international audience.

- Format: Question by Question

Instead of showing the whole questionnaire in one page or each group of the questionnaire separately the metadata management survey will show each question separately. Though Peytchev et. al. (2006) did not find any difference between scrolling versus paging the study contains a lot of matrix questions and they might scare off participants due to their combined size. Therefore a question by question format was used.

- Progress Bar: off

As the study heavily uses branching a progress bar does not have any benefit as the jumps might be too big. Recent studies also show little benefit of progress indicators (e.g. Villar, Callegaro, & Yang, 2013).

- “Survey consists of XX questions”: off

This is caused by the same rationale like the progress bar. The number of questions heavily depends on the branching.

- Anonymized Answers: yes
- Cookies: off
- Store IP Addresses: off
- Save Referrer-URL: off

All of these settings regard the internal handling of users. As the survey is planned to be anonymous it would violate German data protection laws to claim to have an anonymous study and leave these settings as a back door.

- Save Timings: yes

This setting leads to the creation of a log file containing time stamps for every answer. It does not violate data protection laws as no user information is stored, but it can be used for additional research using log file analyses. Without this setting only the final answers and the overall interview time would be stored in the database.

- Mandatory questions: no

The participants are not forced to answer any of the questions as the participation is fully voluntarily. This means in a worst case scenario a participant can answer a questionnaire by skipping all questions. Bandilla et. al. (2000) define these participants as “lurkers” and they would simply break off the interview if they are forced to answer. Couper (2000) recommends not to force participants for an answer but rather encourage the user to finish the survey by keeping it short, interesting and easy to do.

### **2.1.1.3 Answer modes**

Couper (2000) recommends a mix of different answer modes like lists and grids to keep the survey more interesting. Furthermore the whole answer categories should be visible on one screen without scrolling thus long lists and matrixes have to be avoided. The field test of the survey tried therefore different answer modes for similar questions, e.g. a question about the delivery of different material by the data collection agency was asked one time to data producers as a matrix and to data users as a ranking question. The ranking question proved to be much more time intensive and was sometimes skipped. Therefore, this kind of answer mode was removed before the main data collection. Nevertheless this had

also a drawback as the main study version now contained a lot of matrix questions which are according to Couper (2000) the answer type with the highest break-off rate. Splitting the matrixes into separate questions nevertheless would have prolonged the questionnaire to unacceptable length. Figure 40 shows an example of one of the ranking questions from the field test.

Please rank the **five most important** metadata for an item in a questionnaire or assessment which are most relevant for preservation or re-use in later studies?

*Double-click or drag-and-drop items in the left list to move them to the right - your highest ranking item should be on the top right, moving through to your lowest ranking item*

*Please select at most 5 answers*

Your choices	Your ranking
Analytic scripts	
Answer categories	
Concept	
Flow logic (e.g. filtering, routing, dynamic text)	
Fragments of software (e.g. code which makes the item reusable)	
Interviewer instructions	
Layout definitions	
References to research literature	
Scoring rules	
Statistical values from previous studies	
Stimuli in a reusable format	
Variable definitions	

Figure 39. Item from the Data Management Survey Field Test

This question was replaced in the main study data collection with a matrix about importance. Furthermore some of the options were extended with examples based on the feedback from participants in the field test. Figure 41 shows in comparison the main study version of the same question.

How important to you are the following metadata of an item or a question to be re-used in your own studies?					
	Very important	Important	Neither important nor unimportant	Not that important	Not at all important
Analytic scripts (e.g. SPSS or Stata code that generates variables)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Answer categories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information about the concepts to be measured	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Flow logic (e.g. routing, filtering, dynamic text)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fragments of software (e.g. programming code for a specific item)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Interviewer instructions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Layout definitions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
References to research literature	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scoring rules (e.g. information in a cognitive test which answers were correct or incorrect)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Statistical values about the items from previous studies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stimuli in reusable form (e.g. pictures, videos, audio files)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Variable definitions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 40. Item from the Data Management Survey Main Study

Both answer modes are optimized by SkeletonQuest to show their full properties at a 1280x800 screen resolution which fits to modern browsers and modern tablets so they will most likely fit onto one page for the majority of the users. Nevertheless both are too big to be handled on a smart phone. Currently the number of mobile users taking part in surveys according to Wijnard and De Bruijne (2013) is 2% and they are most likely able to finish the survey also with scrolling. If the numbers are rising as predicted later studies most likely need to have an own additional template for mobile users which breaks down matrixes into a better format. Here it is of no concern.

### 2.1.2 Target group

An important step for the survey is to consider who the target group is. The survey is not constructed for the general public as the topics are outside of the frame of knowledge of a person not involved in research or data collections. Furthermore to get a better insight into the proceedings of data producers

and secondary users of data there should be researchers from both parties. As the survey covers the demand from educational sciences and social sciences mainly participants from those groups should be addressed. Nevertheless knowledge about metadata management for the educational sciences and social sciences is also common for personnel working in data collection agencies, market research agencies and vendors of software products around data management.

Therefore the following groups have been identified as participants:

- researchers in the field of social sciences and educational sciences
- personnel at data collection or market research organizations
- software developers of computer-based questionnaire or assessment software
- researchers working at research data centres or data archives

The survey was planned to contain filters to send those groups to relevant questions to their daily work. Furthermore the survey is planned to contain results from international sources.

### **2.1.3 Sampling**

#### **2.1.3.1 Setting a frame for the data collection**

As there is no world-wide central register of the groups described in the last section and the participation is completely voluntarily the sampling has to be quite random. It is only possible to identify the candidates by demographic information and their job biography they give at the very end of the questionnaire. As discussed at the beginning of the section this information is not reliable. On the other hand as the target group are researchers the expectation is they will be more interested in the results themselves. To get a sufficient number of respondents an invitation for participation was sent out to big research organizations, mailing lists for researchers and individual contacts. In total the invitation was sent to 26 mailing lists with some of them containing several thousands of recipients. A high five-figure number (50.000 – 70.000) can therefore be estimated. Unfortunately most of the list do not publish exact numbers. The different mailing lists and organizations can be found in the appendix.

This process already shows a typical problem of web surveys. Though the author is very well connected due to a long work history in several studies and therefore has contacts who promised to forward it is difficult to know how many persons will be reached in the different organizations. Bigger mailing list also do not publish their numbers of recipients. It will be therefore not possible to calculate a response rate though it can be expected as to be very low as invitations through mailing list are considered to be spam or junk mail. The invitation tries to invoke interest in the topic and especially for the data protection concerns which are quite prevalent to Germans the line about data protection has been put onto special emphasis by printing it in bold letters. Also the link to the study has been optimized so it does not look a marketing plot. Also, the researchers were asked to forward the mail to their colleagues. The invitation can be found in the appendix.

### 2.1.3.2 Checking the sample sizes in regards to effects

Section 2.1.3.1 shows a typical problem of web surveys a real sampling with sampling plan is not possible as there is no world-wide register of social scientists or educational scientist. Also, web surveys usually have very low response rates. The identification of researchers is therefore only possible by questions about highest educational degree, scientific background, academic titles and years spent in research. As the data analysis will separate between different groups (e.g. social scientists vs. educational scientists) the results have to be checked if the number of collected interviews is sufficient for data analytics. To enable these analysis the Cohen table (Cohen, 1988 or Cohen, 1992) on effect sizes will be used. Table 6 shows an extract from the Cohen table

*Table 6. Cohen table on effect strength*

	Effect size Index	Small	Medium	Large
t-test on Means	d	0.20	0.50	0.80
t-test on Correlations	r	0.10	0.30	0.50
F-test ANOVA	f	0.10	0.25	0.40
F-test regression	f <sup>2</sup>	0.02	0.15	0.35
Chi-Square Test	w	0.10	0.30	0.50

The results from Table 6 are nowadays embedded in tools for statistical power analyses. A very prominent tool for this purpose is G\*Power (Faul, Erdfelder, Buchner, & Lang, 2009). This tool will be used in the later data analytics sections to determine if the group sizes are sufficient for the effects to be considered significant as the response rate can be expected to be small.

### 2.1.4 Time frame

The field test of the survey with 25 researchers from the areas of survey methodology, data management, data collection, educational sciences and social sciences was performed between the 21<sup>st</sup> of October and 31<sup>st</sup> of October. In November the results from the field test were analysed and the questionnaire was optimized especially with two colleagues from the survey methodology department of GESIS the Leibniz Institute for Social Sciences in Mannheim. The changes between field test and main study can be found at the end of the next section.

The main study data collection started on November 27<sup>th</sup> and ended on December 19<sup>th</sup>, 2014. A longer data collection was not feasible due to the start of the holiday season around middle of December which in Europe traditionally lasts until the second week of January. The hope was to reach a certain degree of participation especially between in the second week of December as the workload of personnel starts going down before Christmas so they might find the time to participate in the study.

## 2.2 Content of the questionnaire

This section shows the content of the questionnaire for the main study data collection and the rationale why the questions were asked. Originally in the field test the chapters of the questionnaire were labelled

like the according phases of the Generic Longitudinal Business Process Model (GLBPM) from “Study Design” to “Publication” but researchers were unfamiliar with the terms. For the main study the chapter names therefore have been changed to more common names.

In total the questionnaire contained 43 questions in 13 groups. The next sections contain an overview of the questions, flow logic, background and answer categories.

The next sections will show the content of the questionnaire in an abbreviated form (only a summary of questions per chapter). The full questionnaires of the field test, main study and second data collection can be found in the appendix. Furthermore routing questions have also been omitted.

### **2.2.1 Questions on “Your Research”**

The chapter “Your Research” first tries to identify if the person is involved as a researcher in the design of questionnaires or assessment instruments. If they are not the whole chapter is skipped. Nevertheless if they are the most important aspects are to find out if they have to use data management plans (i.e. they have to plan their data management efforts due to the involvement of stakeholders) and how much time they are using for it in the whole survey process.

Here the questions from the chapter.

- How often do you have to provide a data management plan (i.e. a plan that describes how to document or store the data for later usage)?
- How do you store your research data, e.g. survey or experimental data?
- Approximately how many months did the last study, e.g. survey or experiment you were involved in take? This overall time may include survey design, data collection, analysis and publication.
- Approximately how many months did the last study, e.g. survey or experiment you were involved in take? This overall time may include survey design, data collection, analysis and publication.
- How many persons were involved in your last study?
- Approximately how much time, as in percent of the total project time, was used for documenting data in your last study?

In the field test the chapter “Your Research” was called “Research Organization” thus the code list was abbreviated with “RO” for naming codes and variables.

### **2.2.2 Questions on “Metadata Standards”**

As metadata standards play an important role in the dissertation and especially in Chapter 1.1.2 they also need their own chapter in the survey. Originally in the field test there was a filter question which asked researchers if they know metadata standards at all. This filter question was removed after feedback from the survey methodologists as most researchers could not follow the term “metadata standards” but were only able to answer when they saw the list with examples.



Here are the questions on metadata standards in the survey.

- Metadata standards play an important role in data archives. We are interested in how often researchers use them in their surveys and projects. Please specify to what extent do you use these metadata standards?
- How satisfied are you with the metadata standard Data Documentation Initiative (DDI)?
- Which versions of DDI are you using?
- How satisfied are you with the metadata standard Question and Text Interoperability (QTI)?
- How much do you agree on the following statements regarding metadata standards?

In this section a limitation of Limesurvey was discovered. In the original planning of the research the idea was to ask all respondents about their satisfaction about metadata standards by first using a matrix question to identify which metadata standards they use and afterwards use a second matrix question about satisfaction only with the sub questions from the previous matrix question where they answered “I use this metadata standard often”. Unfortunately Limesurvey is in the default setting not able to filter matrix to matrix (it is possible to filter e.g. between matrix and multiple choice). Therefore the decision was taken to ask about the two most prominent standards from the previous chapters in the dissertation – DDI and QTI.

### **2.2.3 Question on “Data Sources”**

Metadata standards is followed by a short chapter only containing one question about data sources. Here the important question is what kind of survey or experiments the respondents are involved in. The main question is if they are using questionnaires or cognitive tests. The other options are there for completeness, but actually they are out of the scope of the dissertation (e.g. it is not concerned with metadata standards for experiments or qualitative research though these might be interesting topics for further papers or research).

The only question in “Data Sources” is the following – “How often do you use the following data sources for your research?”. In the field test the chapter “Data Sources” was just called “Data” thus the code list was abbreviated with “DA” for naming codes and variables.

### **2.2.4 Question on “Research Documentation”**

Similar to the previous chapter also this chapter consists of one questions. Originally in the field test it was called “Study Unit” after terminology from the GLBPM or DDI. Nevertheless this was unknown to the experienced researchers giving feedback after the field test. The name therefore has been changed to a more general term. Still the internal codes therefore use the abbreviation “SU” for the variables, options and sub questions.

The question about the “Research documentation” is as follows – “Which of the following categories do you think are important if a research project, e.g. a survey or study has to be documented?”. This question was originally a ranking question in the field test but has been changed to a matrix due to complaints from the respondents.

### **2.2.5 Questions on Data Collection Instruments**

This chapter represents the next step in the lifecycle models described in the previous chapters. To shorten the questionnaire time the first section is only asked to respondents who are involved in the creation of items while the second section only concerns persons who are re-using items or are planning to re-use them for their own research. Originally in the field test this chapter was just called “Items” leading to the code designation “IT” for variables and sub questions.

Here are the questions about data collection instruments.

- How often do you provide these different types of metadata with a questionnaire or cognitive item?
- Have you been re-using or are planning to re-use items or questions developed by other persons or institutions?
- How often do you provide these different types of metadata with a questionnaire or cognitive item?

### **2.2.6 Questions about Data Analysis**

Similar to the last chapter which tried to identify the metadata which is delivered with an item this chapter tries to identify which metadata is delivered with data files for data analysis. Originally these questions were part of a chapter about Data Collection in the field test thus the designation “DC” for variables and codes. In DDI and GLBPM these questions represent the “Process / Analyze” step of the survey cycle therefore the term “Data Analysis” was more fitting to the subject.

Here the questions about data analysis processes.

- Do you analyze data which is provided by a data collection agency, research data center or data archive?
- How often are the following metadata or paradata information included with the data sets delivered to you?
- In general, how satisfied are you with the metadata and paradata information you receive from data collection agencies, research data centres or data archives?
- How often do you use the following data sources for your research?
- Do you provide datasets to others?
- Are the users of your data mostly internal or external?
- Which metadata or paradata (i.e. metadata derived from data collection processes) information do you usually provide to the users of your data?

The data producers do not get a similar mirrored question as such services are normally only provided by research data centres or data archives. Data collection agencies normally delivered datasets in formats of statistical packages (e.g. SPSS, Stata) which have been cleaned by internal quality control processes, but do not offer services like remote calculations or data enclaves.

### 2.2.7 Questions about Paradata

The chapter on “Paradata” originally belonged to the “Dissemination” process as the term describes e.g. the generation of log files during the interview process and the analysis afterwards. Thus the questions, variables and codes have a “DI” designation at the beginning. Unfortunately the term “Data Dissemination” was not known to most of the researchers taking part in the field test. As the chapter was shortened to include only the dissemination of paradata the title of the chapter was changed. Nevertheless as paradata is also not a very well-known term for technical metadata collected during the interview process a lot of explanations and examples were needed.

Here the questions on the dissemination of paradata.

- Do you analyze paradata from log files collected during the survey (e.g. timestamps)? Note: paradata are administrative metadata about the data collection process on a micro-level (e.g. how many days the interview took, contact attempts by the interviewer, timing of answers by the interviewee).
- When you analyze paradata from log files, which formats do you prefer?
- In general, how satisfied are you with the documentation of paradata from log files you use for your research?

This chapter does not contain too many questions about data dissemination processes from external researchers as they are out of the scope of research data centres. In the long run it might be possible to offer certain analytic possibilities within data documentation portals but currently the process ends for research data centres when the secondary researchers obtained the datasets plus additional information.

### 2.2.8 Questions about Publication

The “Publication” process is the last process of the GLBPM and DDI if the study has only one wave (otherwise the retrospective phase is the last). From a data centre perspective the important part is to connect publications done with the data with the datasets itself into a product called “enhanced publication”. For research data centres this process is needed to show decision makers the necessity to provide research data infrastructures. Basis for this process is not only the need of secondary analysts to obtain datasets but the willingness of the original data producers to share their data, invest time into documentation and connect them to their research papers.

Here are the questions about the publication process.

- Are you publishing scientific research in journals?
- How important to you are the following factors in choosing a journal for publication?
- Do you share the metadata, paradata or data from your research with others?
- With whom do you share your data?
- Please indicate the reasons why you do not share your data.
- Which of the following could be important motivating factors for you to share your data with others?

With the questions about metadata standards and data models the scope of research data centres is left. The rest of the questionnaire is concerned with additional and demographic questions.

In comparison to the GLBPM the supplementary processes “project management” and “quality management” (“use of external metadata standards” is covered in one of the chapters) have been omitted. This is due to the length of the questionnaire which did not feature enough time to cover all the main processes. This might be part of further research.

### **2.2.9 Question about Software**

The next chapter contains an additional question which is very similar to MS01 (“Please specify to what extent do you use these metadata standards?”) in an indirect manner. As an earlier chapter of the dissertation described which software supports which standards researchers might use metadata standard through software though they are not aware of this.

The question in contrast is the following – “Some software products support the use of metadata standards. We are interested to know which of these products are known to researchers. To what extent do you know the following software products for use in the area of questionnaires or computer-based assessment?”

### **2.2.10 Question on Organisation**

The questions on “Organisation” were previously called “Research Organisation” and are the start of the demographic information. To construct groups for descriptive statistics for the later chapters the filter questions which identify data producers and data users are not sufficient. It is also important in which discipline the researchers is working, the educational level and the organisation which is represented.

Therefore here the question about the organisation the respondent is working for – “Who is your employer? If you have more than one employer please specify the one where you spend the most of your time.”

### **2.2.11 Questions on Information about the Researcher’s background**

The chapter “Organisation” and “Information about the Researcher’s background” used to be in one chapter. Nevertheless they were separated for the main study as the organisational question can also be asked to non-research personnel and therefore the title “Information about the Researcher’s background” (which also can be seen in Limesurvey during data collection) would not be appropriate. OR01 serves as a filter question for the whole chapter as these questions will only be asked to persons who said they work in a university, public research institute or private research institute.

Here are the questions about the researcher’s background:

- What is your current academic status? Please choose the option which reflects best your daily work even if more than one role fits to you.
- How many years have you been working in academia?

- Which areas of research have you been working in?

### **2.2.12 Questions on Region**

As the study is supposed to provide an international result it is necessary to inquire from which region the respondents are coming from or rather working in. The invitation list has a strong focus on the European Union and North America (to be precise Germany, Hungary and the United States) therefore these regions will be separated more clearly. For Asia, South America or Oceania a participation in very low numbers is expected therefore a specific country list has not been implemented.

Here are the questions on region:

- Which region do you work in primarily?
- In which country of the European Union (EU) do you work in primarily?
- In which country in North America do you work in primarily?

### **2.2.13 Questions on Personal Information**

Finally the survey contains very basic personal information. This is limited to gender and age as more information might cause problems with German data protection laws and requires respondents to sign off the survey. For descriptive statistics to identify a trend in data management deeper information was not considered to be necessary thus this high level information is sufficient.

Here are the questions on “Personal Information”:

- In which year were you born?
- What is your gender?

The question about gender ended the questionnaire. It was followed by a screen which thanked the respondent for participating in the interview and informed where the results can be found after the analysis.

The dissertation will be available as an Open Access resource in the Pedocs archive of DIPF while the datasets will be stored in the Research Data Centre for Education (FDZ Bildung). As the survey was constructed around the topic of data management, enhanced publications and availability of datasets the handling of the results should serve as an example how these processes can be put into practice.

### **2.2.14 Questionnaire Flow Logic**

The data management survey was built not to be sequential so respondents only get the questions fitting to areas where they have experiences. Though the filter logic was described in the individual questions in the previous chapter flow charts describing the filter logic of the whole questionnaire for the main study and the field test questionnaire can be found in the appendix.

## 2.3 Data Collection

As planned the data collection took part between November 27th and December 19th, 2014. During this phase 221 interviews were collected by Limesurvey. Out of these 221 interviews 120 were fully completed while 101 remained unfinished. The raw data was saved in SPSS format to be cleaned and analysed in Chapter 3. Furthermore a log file with timestamps was saved to be the basis for a potential log file analysis.

During the data collection there were no reports about technical incidents and the server logs show a 100% uptime of all systems. The whole study was daily backup up as an archive (Limesurvey LSA file) and the data files extracted in SPSS format. Nevertheless a use of the backups was not necessary.

### 2.3.1 Feedback from users during Main Study data collection

The invitation to the survey asked participants to send in questions and feedback during the main study. In total this happened three times:

- A survey methodologist from the United States reported he found the questionnaire to be difficult especially at the beginning. A special problem was question RO05 where the participants were asked to give the number of co-workers in their last survey. The researcher is working several studies in parallel therefore “last survey” does not apply to him. The interesting fact is the reduction to “last survey” was introduced between field test and main study on request of the survey methodologists of GESIS in Germany who had exactly the opposite opinion. As the data collection was in the middle stage there were no changes to the questionnaire.
- An IT specialist from Germany remarked the link to Pedocs in the end screen is not the official address. This has been corrected during the data collection as this is only a cosmetic bug (the other link also worked).
- A researcher from Germany remarked Learning Object Model (LOM), Learning Tool Interoperability (LTI) and Question and Test Interoperability (QTI) are not metadata standards per se though they use a lot of metadata but rather technical definitions. The author and he had an interesting conversation by mail where the metadata standard definition of the dissertation was exchanged. The researcher then agreed it could be defined in the manner of the survey.

The amount of feedback was therefore relatively small and the author thanks all participants for the input and responses. After finishing the field work the data analysis started as to be described in the next chapter.

*„Statistics is the most important science in the whole world: for upon it depends the practical application of every other science and of every art: the one science essential to all political and social administration, all education, all organization based on experience, for it only gives results of our experience.”*

(Florence Nightingale)

### 3. Results from the first data collection

The previous chapter showed the survey design of the questionnaire which went into the field from November 27th, 2014 until December 19th, 2014. This chapter will now show the analysis of the data using different paradigms. As the subject is a questionnaire the analysis might not provide as much possibilities as the cognitive assessments normally used in educational sciences as there is no scoring involved (e.g. there is no correct or incorrect answers to the questions).

#### 3.1. Data management practices of the researchers

In the section scientist were asked about background information for their studies in regards to data management. The chapter starts with a filter question on the involvement in survey design. The majority (72.5% of the respondents) is involved in this process which is one of identifying parameters of being a data producer. 24.2% of the participants are not involved and 3.3% of the respondents did not answer the question.

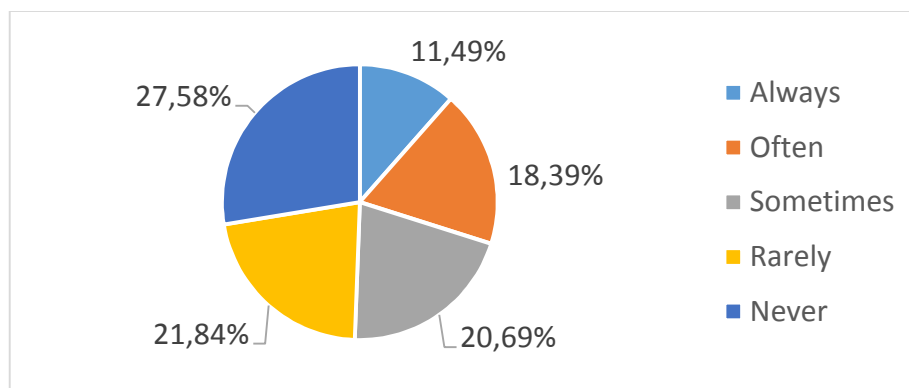


Figure 41. Frequency of providing data management plans (%) (N=87)

Though data management plans are in the meantime a requirement from funding bodies (quote) they do not seem to be a standard for researchers. The numbers of “always” and “often” almost equal the “never” category. “Sometimes”, “Rarely” and “Never” are clearly dominating (see Figure 39). This is no surprise as most software packages have no means in supporting researchers in those processes. Here clearly a demand for support can be identified.

Table 7. Frequency of the storing the research data (%)

Storing research data	Frequency (%)
Research data is stored in a research data facility or archive which belongs to the institution I work for.	36.57
Research data is stored in a third party research data facility or archive.	9.70
Research data stays with Prime Investigator (PI) or project team.	26.86
Research data stays on the IT infrastructure of the project.	25.37
Do not know	1.50

Note: N=134

Research data is clearly an internal process in research facilities. Distribution of the answers is in Table 7. Most researchers store their data within own archives or with the internal project team. Nevertheless the answers “Research data stays with the PI or project team” or “Research data stays on the IT infrastructure of the project” convey another problem as they do not solve the issue what happens after the project is finished, e.g. IT infrastructure is normally shut down after a certain time (in normal IT cycles between 4-5 years). Surprisingly getting external help does not seem to be popular as third party facilities are not widely used.

In median values the average research project lasts roughly 18 months. The RO06 question (“How many persons were involved in your last study?”) did not have an ideal wording in the main study what can be seen in the maximum and average values. In the original question from the field test the addition “in your organization” was used which was removed by advice of the external survey experts. Some persons now obviously counted the participants in which leads to the high values. The median value of 7 persons therefore seems to be much more plausible. In another round of the survey this question has to be changed.

In median values the researchers spent 10% of the project time in data documentation. In combination with the project length of 18 months this means all documentation is done within six weeks. This time does not seem to be sufficient to prepare data onto a level which can be used for archival purposes.

### 3.2. Usage and satisfaction of metadata standards

In this section researchers had to describe how familiar they are with different metadata standards and evaluate how satisfied they are with those standards. It can be stated metadata standards are not widely used or the researchers do not know them. From respondents the most known standard is DDI with 16.9% of researchers who used it “often” while especially the educational standards are lacking behind. QTI, LTI, LOM and SCORM are hardly known which can also be proven from Table 8.



Table 8. Frequency (%) of Metadata Standards

Standard	Usage “often”	Usage “rarely”	Usage “never”	Usage “Do not know”
Data Documentation Initiative (DDI)	16.9	8.9	33.1	41.1
Dublin Core	7.2	10.4	20.8	61.6
Questionnaire and Text Interoperability (QTI)	4.2	2.5	16.7	76.6
Machine Readable Cataloguing (MARC)	2.5	5.7	14.8	77.0
Statistical Data and Metadata Exchange (SDMX)	1.6	3.3	22.8	72.4
Learning and Test Interoperability (LTI)	1.6	0.8	16.4	81.1
Learning Object Model (LOM)	0.8	3.3	19.9	76.0
Shareable Content Object Reference Model (SCORM)	–	6.6	14.0	79.3
Metadata Encoding and Transmission Standard (METS)	–	2.5	19.7	77.9
Text Encoding Initiative (TEI)	–	2.5	17.4	80.2
QueDex	–	–	12.3	87.7

Note: Hyphen means that there was no answers in that category. N=130.

The least known standard is QueDex which is used for qualitative data, but also for all other standards can be said researchers rather do not know them. On the first glance it looks as if the topic of metadata standards is not important to researchers. This contrasts nevertheless with another result from the survey.

Table 9. Distribution of the statements regarding metadata standards

Statement	1	2	3	4	5	6	7
Importance of metadata standards (MS0603)	16.66	46.66	10	0.83	2.50	12.50	10.83
Identification of metadata standards (MS0601)	4.16	16.66	11.66	17.50	9.16	30.83	10.00
Metadata standards and domains (MS0602)	2.50	7.50	10.83	21.66	26.66	20.83	10.00
Complexity of studies in regards to metadata standards (MS0604)	2.50	5.00	12.50	33.33	20.00	16.66	10.00

Note: Hyphen means that there was no answers in that category. 1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree; 6 = Do not know; 7 = No answer. N = 130.

Researchers are aware of the advantages of using metadata standards (see Table 9). The question is therefore why does this gap between the acceptance of metadata standards and the knowledge about them exist. Clearly the researchers have problems identifying the metadata standards which are important for their projects. They know about the importance of metadata standards and also that they exist in their respective domains which can be seen from the next graph. Nevertheless they do not have the detailed knowledge to use metadata standards on their own. The problem is also not the complexity of the data to be represented by standardized metadata as shown in Table 8.

Researchers are willing to use metadata standards and are generally positive about the features, but obviously have difficulties in identifying the correct standard or lack the knowledge how to use them. The question is if this knowledge is really necessary for researchers or if rather the external processes should be improved like automatic generation of metadata within software products or closer connections between research data centres and research facilities while studies or projects are running.

When metadata standards are used in social sciences and educational sciences the satisfaction with the two main standards DDI and QTI differs. Unfortunately in both cases the number of researchers using the standards were low so the results can only be seen as a trend. Table 10 shows the satisfaction of researchers with DDI.

*Table 10. Satisfaction with DDI and QTI (%)*

Metadata	Very satisfied	Satisfied	Neither satisfied nor dissatisfied	Dissatisfied	Very dissatisfied	No answer/ do not know
DDI	16.66	55.55	16.66	11.11	–	–
QTI	–	20.00	80.00	–	–	–

*Note: Hyphen means that there was no answers in that category. n=18.*

Researchers who use DDI are generally satisfied with the standard. As there are several versions of DDI which are not compatible to each other (e.g. DDI Codebook V2.5 vs. DDI Lifecycle V3.2) there was another question about usage (see Figure 40). The newer versions of DDI Lifecycle V3.0 to DDI Lifecycle V3.2 are clearly trending though there are still users also of the older versions before DDI Codebook V2.5.

The situation for satisfaction about QTI nevertheless looks different (see Table 10). Though the numbers of researchers using QTI are quite low in the survey (n=5) there seems to be a trend of a certain degree of dissatisfaction in comparison to the DDI users. Maybe a standard which can represent a variety of simple computer-based assessment items without any other metadata is simply not good enough.

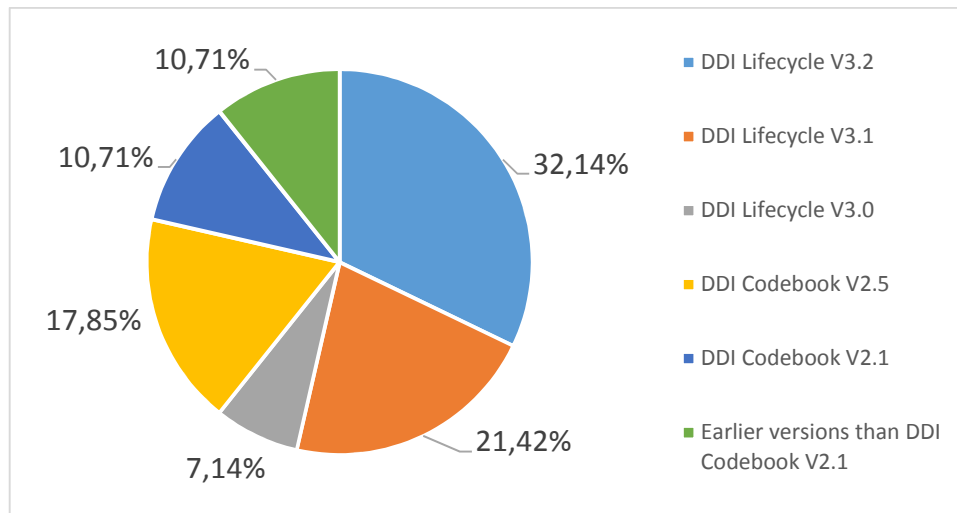


Figure 42. Frequency of DDI versions usage (%) (n=28)

DDI was in the descriptive statistics the most popular metadata standard and it is based on the social sciences. Therefore the expectation is DDI is used by social scientists and not by educational scientists. Table 11 shows the results. As expected DDI does not play the same role for educational scientists like for participants which identify themselves as social scientists or both ( $\chi^2(2)=11.58, p<.05$ )

Table 11. Cross tabulation on usage of the DDI standard between social scientists and educational scientists

Research area	Not used	Used
Social science (n=23)	17.33	13.33
Educational science (n=28)	36.00	1.33
Both (n=24)	24.00	8.00

Another hypotheses stated educational scientists and social scientists will show the same positive attitude towards the use of metadata standards in general. As expected there is no significant difference in the attitude towards metadata standards (MS06\_c). In all three groups it can be considered positive as already described in the descriptive statistics. Nevertheless there was a significant difference in the answer “My studies are too complex to use metadata standards for data documentation.” (MS06\_d) ( $\chi^2(2)=9.59, p<.05$ ). Educational scientist researchers who identify themselves as belonging to both groups agree with this statement much more than social scientist ( $U=93.5, p<.05$ ).

### 3.3. Usage and differences of data sources among scientists

To continue with the identification what a metadata standards for the educational sciences should be able to cover the participants were asked which data sources they are normally using for their research. Rational behind this is how much effort should be put into requirements for data management processes to accommodate those needs. Table 12 shows the ranking of the different sources for all participants.

Table 12. Ranking on data sources (%)

Rank	Data Source	Usage “always”	Usage “often”
1	Structured questionnaire (e.g. multiple choice)	40.0	38.5
2	Structure cognitive test (e.g. computer-based assessment)	11.5	23.1
3	Qualitative text data (e.g. narrative interviews, oral history)	5.4	17.7
4	Experimental designs	5.4	11.5
5	Observational data (e.g. recordings, transcripts)	4.6	10.8

Note:  $n=130$

Interestingly the qualitative data sources which need the most effort to document and to be transformed into research data for secondary usage are the ones which are less used than quantitative sources. Reasons for this can be the complexity of data management for these domains which can also be seen by the fact the metadata standards like QueDeX are hardly known. The data management for qualitative data needs more research, but as the focus of this dissertation is not within this scope this task will fall to other researchers.

Within the scope of the dissertation it can be seen questionnaires are very common while structured cognitive tests are lacking a bit behind. Nevertheless there is enough demand for them to construct better metadata standards.

The biggest difference between social scientists and educational scientists can be expected in the data sources they are using. From the typical descriptions of their respective fields social scientists are expected to have a preference for questionnaires while educational scientist are more likely to use cognitive tests for their data collection.

As expected the differences started in the first answer category which was in regards to cognitive instruments ( $\chi^2(2)=23.45, p<.05$ ). Nevertheless the results were a bit surprising. According to the analyses social scientists are more likely to use cognitive tests than expected ( $U=89.5, p<.01$ ). Furthermore there is also no difference between all three groups in using questionnaires. The expectation was the usage would be lower for educational scientists. This strengthens the impression which was found in the literature analysis built in the first chapter both disciplines are so close to each other data management methods can be re-used or modified.

### 3.4. Practices of the research documentation

Research documentation describes which information about studies, projects and surveys researchers find important. As documenting the background of studies is time extensive the survey contained a question where the researchers were asked how important they consider aspects of research documentation. Table 13 shows the results of these questions.

Table 13. Ranking and frequency of research documentation (%)

Rank	Aspect	1	2	3	4	5	6
1	Concepts of the study	60	27.5	3.33	–	–	9.16
2	Target group / Universe	56.66	28.33	4.16	1.66	–	9.16
3	Structure of the study	56.66	25.83	5.00	0.83	0.83	10.83
4	Sampling methods	55.83	30.00	4.16	–	–	10.00
5	Analysis plan	39.16	37.5	8.33	3.33	2.5	9.16
6	Data cleaning methods	38.33	41.66	7.50	1.66	–	10.83
7	Quality control mechanisms	28.33	49.16	10.83	1.66	–	10.00
8	Research team	27.50	36.66	17.5	9.16	–	9.16
9	Funding sources	20.83	38.33	16.66	13.33	1.66	9.16
10	Previous publications	20.00	29.16	22.5	16.66	2.5	9.16
11	Organizations involved	15.83	42.50	21.66	8.33	1.66	10.00
12	Technical infrastructure	14.16	40.00	18.33	13.33	5.00	9.16

Note: Hyphen means that there was no answers in that category. 1 = Very important; 2 = Important; 3 = Neither important nor unimportant; 4 = Not that important; 5 = Not at all important; 6 =No answer. n=120.

These results can be used to form requirements for metadata standards to see where more effort in documentation should be put. Interestingly the focus is on concepts, target group and structure while funding sources, previous publications, organizations and infrastructure are less important.

### 3.5. Data documentation usage and differences of data producers

The chapter on data collection instruments contains the first comparison between data producers and users of secondary data. Here the main question is what data producers document about items and what do secondary users expect. The distinction was done by the filter questions which already provided interesting results.

It is not surprising most of the researchers (67.5%) are involved in item creation given the answers in the data sources part where questionnaires and cognitive tests were clearly dominating. 27.5% of the respondents are not involved in the process and 5% of the participants did not answer the question. Interesting was the comparison which the next filter question about secondary usage of data. The result was almost identical, most of the researchers (68.33%) have been re-using or are planning to re-use the items or questions developed by other persons or institutions. 25.83% of the participants haven't re-used others items or questions and are not planning to do that. 5.83% of the respondents didn't answer.

There is certainly a demand for items which have been used in previous studies or come in general from third parties. This means repositories for questionnaires or item banks for cognitive items are certainly needed. The advances section will clarify how big the overlap between the researchers answering "Yes" in both cases is. Nevertheless, here the answers between data producers and data users about items in direct comparison. Table 14 shows the metadata provided by data producers in regard to items.

Table 14. Metadata provided with items from data producers

Rank	Category	Always	Often	Sometimes	Rarely	Never	Do not know	No answer
1	Answer categories	49.38	19.75	19.75	1.23	2.47	3.70	3.70
2	Scoring rules	27.16	17.28	18.52	4.94	16.05	11.11	4.94
3	Variable definitions	25.93	28.40	16.05	4.94	9.88	11.11	3.70
4	Interviewer instructions	25.93	23.46	17.28	14.81	6.17	8.64	3.70
5	References to research literature	24.69	17.28	27.16	8.64	16.05	2.47	3.70
6	Concepts to be measured	23.46	32.10	25.93	7.41	6.17	1.23	3.70
7	Analytic scripts	22.22	19.75	18.52	11.11	23.46	2.47	2.47
8	Flow logic	19.75	16.05	25.93	7.41	6.17	1.23	3.70
9	Layout definitions	16.05	17.28	23.46	14.81	19.75	3.70	4.94
10	Statistical values from previous studies	7.41	23.46	25.93	14.81	20.99	3.70	3.70
11	Stimuli in reusable formats	6.17	11.11	19.75	18.52	30.86	9.88	3.70
12	Fragments of software	2.47	9.88	20.99	18.52	38.27	6.17	3.70

Note:  $n=81$

First, there were slight differences in the handling of different kinds of metadata by the phd student, no phd, phd and professor group. They could be found in IT0201 – “Analytic scripts (e.g. SPSS or Stata code that generates variables)” ( $\chi^2(3)=11.26, p<.05$ ) and IT0212 – “Variable definitions” ( $\chi^2(2)=6.17, p<.05$ ). Nevertheless, both lost their significance when further analysing the data.

The previous results showed about the importance of metadata elements for data users from the groups of social scientists and educational scientists, we will look at the same issue from a data producer’s perspective meaning which metadata elements do researchers from these groups offer. The only significant difference was in the delivery of statistical values (IT0210) ( $\chi^2(2)=6.96, p<.05$ ). Here the further analysis in regards to the groups. Educational scientists are more likely to provide statistical values as further metadata elements than their counterparts from the social science or researchers who identify themselves as both ( $U=71.5, p<.05$ ).

The results indicate some problems regarding the viewpoint of storing cognitive assessment items for computer-based assessment as layout definitions, stimuli in reusable form and fragments of software are ranked very low. This table for all complete answers from researchers will be broken down for groups of social sciences, educational sciences or both, but it shows a first trend. The same categories were asked to data user but with the question how important they are to their research. Table 15 represent the results.

Table 15. Importance of metadata on items from the perspective of data users (%)

Rank	Category	1	2	3	4	5	6
1	Answer categories	67.07	25.61	2.44	2.44	–	2.44
2	Concepts to be measured	60.98	28.05	3.66	4.88	–	2.44
3	Scoring rules	54.88	26.83	9.76	3.66	2.44	2.44
4	Variable definitions	42.68	37.80	6.10	3.66	6.10	3.66
5	Flow logic	36.59	30.49	12.20	9.76	6.10	4.88
6	Interviewer instructions	31.71	42.68	12.20	6.10	3.66	3.66
7	Statistical values from previous studies	29.27	39.02	17.07	8.54	1.22	4.88
8	References to research literature	25.61	39.02	14.63	13.41	2.44	4.88
9	Analytic scripts	24.39	36.59	19.51	10.98	6.10	2.44
10	Stimuli in reusable formats	21.95	30.49	17.07	13.41	9.76	7.32
11	Layout definitions	15.85	34.15	18.29	20.73	6.10	4.88
12	Fragments of software	4.88	26.83	28.05	15.85	19.51	4.88

Note: Hyphen means that there was no answers in that category. 1 = Very important; 2 = Important; 3 = Neither important nor unimportant; 4 = Not that important; 5 = Not at all important; 6 =No answer. n=82.

The expectations of data users differ in some respects from the metadata information delivered by the data producers. A prominent category is “concepts” which is on the second rank from the importance of data users but only on the sixth rank of delivery by data producers. With comparing those two tables new standards and software can be harmonized or features prioritized.

Another expectation was educational scientists will put emphasis on scoring and layout as this is more relevant in cognitive instruments. Scoring does not exist in questionnaires and layout is important as e.g. font sizes have a direct impact on the difficulty of an item. In this case the expectations were partly met. The deviating elements were “fragments of software” (IT1204) ( $\chi^2(2)=16.63, p<.01$ ) and “layout” (IT1207) ( $\chi^2(2)=7.53, p<.05$ ). As expected educational scientist have more demand for layout definitions ( $U=101.5, p<.05$ ). On the other hand social scientist have a much greater wish for fragments of software. Surprising is the not existing deviation between social scientists and educational scientists in regards to scoring definitions.

### 3.6. Differences in dataset delivery between data users and producers

The chapter on data analysis is again a comparison between the expectations of the data users versus the metadata delivery from the data producers. In this case the focus is on preference and delivery in regards to paradata and datasets. For identification to which groups the researchers belong to filter questions are used.

About half of the researchers (50.84%) do not get their data from in-house production but from external sources. This means they are depended on the preparation of datasets and log files from those third parties. Table 16 shows the metadata information they usually receive from those sources. The combination of codebooks and cleaned datasets is quite common while additional material like field reports, data cleaning reports and quality control reports are quite rare. The bad news for researchers involved in process mining is the huge lack of preserving log file data which is essential for their research. Nevertheless the lack of additional material does not affect the satisfaction with the data delivery too strongly as the general attitude is rather a satisfied one though not a very satisfied one (see Figure 44).

Table 16. Metadata and paradata information included in dataset delivery (data users' perspective)

Rank	Category	Always	Often	Sometimes	Rarely	Never	Do not know	No answer
1	Codebook with variable definitions	38.33	31.67	21.67	1.67	1.67	–	5.00
2	Cleaned datasets	33.33	35.00	15.00	6.67	5.00	–	5.00
3	Raw datasets	20.00	13.33	18.33	21.67	21.67	–	5.00
4	Field report	11.67	13.33	20.00	30.00	18.33	–	6.67
5	Description of data cleaning processes	8.33	15.00	35.00	15.00	20.00	1.67	5.00
6	Quality control reports	8.33	11.67	30.00	31.67	8.33	3.33	6.67
7	Log files from collected cases	6.67	11.67	16.67	33.33	26.67	–	5.00

Note: Hyphen means that there was no answers in that category. n=60.

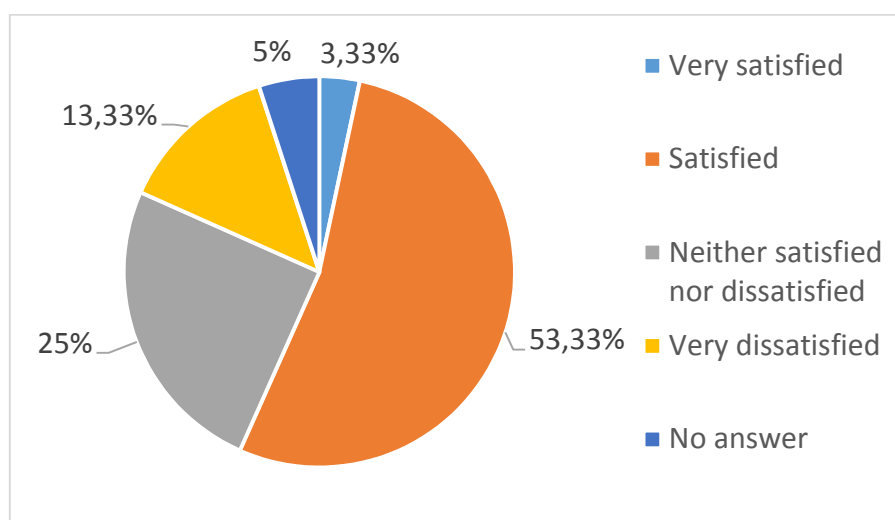


Figure 43. Satisfaction with metadata information from third parties (n=60)

Another important question for data analysis is the format or way in which the datasets are delivered. Table 17 represents the ranking the most common categories. There does not seem to be the “one way” to acquire datasets from third parties. Surprising are nevertheless the high rankings for raw data (Rank 2) and Public Use Files (Rank 3). Raw data should normally not be accessible to researchers from third parties as it normally contains sensitive information with falls under data protection regulations. Public Use Files on the other hand are usually very highly aggregated data to avoid problems with data protection (so they can be put for download on websites for all interested parties) so their scientific value for data analysis is very limited. Also, the modern ways of accessing datasets like Virtual Research Environments or personal extracts have not really changed the classical ways of data access.



Table 17. Frequency of the access to datasets (%)

Rank	Category	Always	Often	Sometimes	Rarely	Never	Do not know	No answer
1	Scientific Use Files	10.94	28.13	21.88	12.5	18.75	–	7.81
2	Raw datasets	8.33	23.33	18.33	20.00	20.00	1.67	8.33
3	Public Use Files	5.00	30.00	23.33	13.33	20.00	–	8.33
4	Data enclaves, secure data services or Virtual Research Environments	3.33	5.00	11.67	16.67	43.33	13.33	6.67
5	Relational database	–	13.33	11.67	18.33	43.33	6.67	6.67
6	Personal extracts created by portal websites	–	3.33	3.33	10.00	55.00	21.67	6.67
7	Data Warehouse	–	1.67	11.67	13.33	50.00	16.67	6.67
8	Remote calculation or job submission	–	–	6.67	10.00	46.67	28.33	8.33

Note: Hyphen means that there was no answers in that category.  $n=60$ .

Nevertheless here are the same questions to the data producers starting with the identification of this group. The group of data producers has a similar size to the data users (50% of the respondents). The first question for the data producers was to whom they normally deliver their data – external customers or internal users or both. In general the deliveries go to both groups with shifts to one or the other side. No data producer works only externally or only internally (see Figure 45).

To compare the impression of the data users to the data producers furthermore Table 18 shows the formats they usually provide to their customers. The perspective between data producers and data users does not differ much though the data producers have slightly higher values in the perspective of their deliveries.

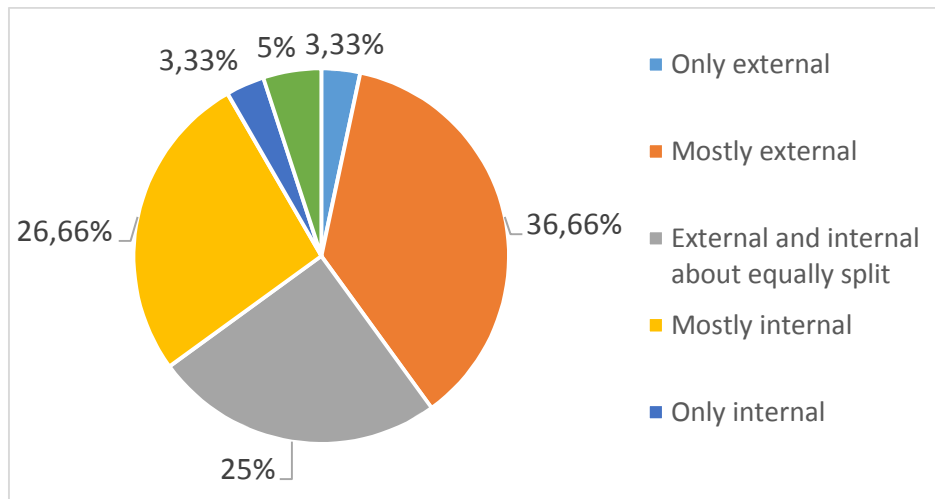


Figure 44. Delivery to external or internal users (n=60)

Table 18. Metadata and paradata information included in dataset delivery (data producers' perspective)

Rank	Category	Always	Often	Sometimes	Rarely	Never	Do not know	No answer
1	Codebook with variable definitions	51.67	28.33	10.00	1.67	5.00	–	3.33
2	Cleaned datasets	46.67	30.00	8.33	3.33	5.00	3.33	3.33
3	Raw datasets	15.00	11.67	18.33	15.00	33.33	3.33	3.33
4	Description of data cleaning processes	11.67	23.33	25.00	21.67	10.00	5.00	3.33
5	Field report	10.00	11.67	16.67	26.67	30.00	1.67	3.33
6	Quality control reports	8.33	10.00	18.33	30.00	26.67	3.33	3.33
7	Log files from collected cases	6.67	8.33	23.33	25.00	31.67	1.67	3.33

Note: Hyphen means that there was no answers in that category. n=60.

### 3.7. Habits and differences of analysing paradata

Paradata and specifically the analysis of log file paradata are a new and upcoming topic at several faculties around educational sciences (e.g. the ProKom project at DIPF). The first question is therefore again a filter question to see whom of the researchers uses these methods at all. 27.5% of the participants analyze paradata from log files, 64.16% of the respondents do not use this technique and 8.33% did not answer the questions.

Obviously this topic is not as popular with researchers as classical data analysis of datasets, but it is at least concerning a third of them. One challenge of log file analysis is there is no predefined format for them so there is a huge effort in converting them to useful applications. The next question therefore tried to analyse which formats researchers prefer for log file data analysis. Table 19 contains the results.

Table 19. Preferred formats for log file analysis

Formats of log files	Frequency (%)
CSV file	45.45
Excel file	27.27
Mplus file	12.12
R file	15.15
Relational Database	18.18
SAS file	3.03
SPSS file	39.39
STATA file	36.36
TXT file	27.27
Word Document	12.12
XML file	21.21

Note: The participants could mark multiple responses.  $N=130$

CSV files, SPSS and STATA are clearly dominating, but in general the preference in regards to formats seems to be a mixed bag. Peculiar is the choice of Word Documents by four participants as this format does not offer any proper handling for statistical analysis. The classical formats for log files which are TXT and XML files are named, but in less demand than tabular formats. Obviously the focus is on computability in statistical packages which means in most cases a conversion from a text format into a tabular format.

Based on descriptive statistics the question on the usage of log file formats caused a problem as there was no clear preference. To get a clearer picture which also represents experience in research the results were analyzed in regards to the “phd student”, “no phd”, “phd” and “professor” group. Table 20 as an example the cross-tabulation of the first result which can be seen as the prototype for all other computations of this question.

Table 20. Cross tabulation on the preferred log file format – example of CSV files (%)

	Academic status			
	PhD students	No phd	Phd	Professor
Not chosen	29.11	13.92	25.31	17.72
Yes	2.53	3.8	5.06	2.53

Note:  $n=79$ .

Unfortunately also the inferential statistics do not give any further insight into the preferred format to be used for log file data. There is no clear preference for a format regardless of group affiliation. It seems there is no agreement within the research community on a clear process to handle or document this kind of data.

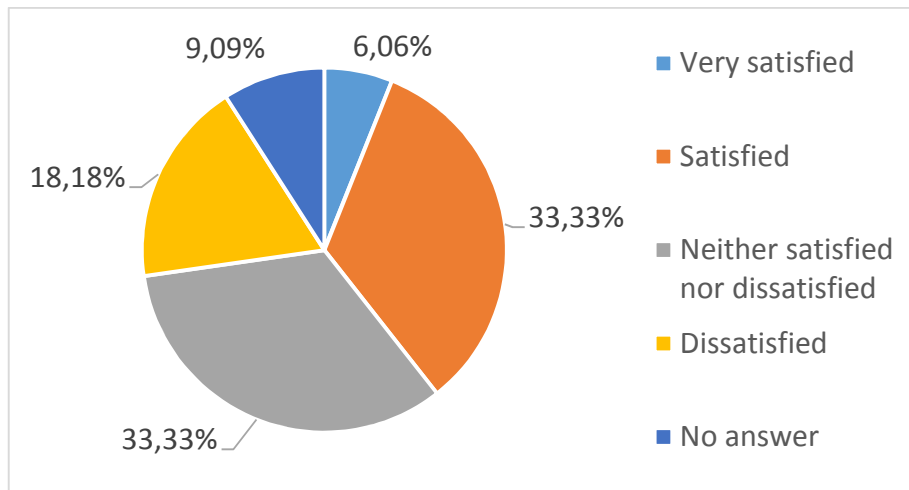


Figure 45. Satisfaction on paradata documentation (n=32)

The last question concerned the overall satisfaction with the log file paradata (see Figure 46). The researchers do not seem to be fully dissatisfied, but there is a small overall tendency to dissatisfaction or at least a mediocre satisfaction. Obviously there is room for improvement especially in comparison to other metadata described in the earlier chapters.

### 3.8. Background factors in publication processes

Support for the publication process is the last step in the GLBPM before the retroactive phases and therefore it is important to know how to link publications to research data and metadata. For this reason a chapter within the questionnaire was built to get more insight on those issues. First the researchers were separated from non-publishing staff by a filter question. As expected most participants of the survey are researchers and therefore involved in publication processes (67.5% of the participants). To understand the manner how they publish a question was asked which categories about choosing a journal are most important to them.

Table 21. Factors for choosing a journal (%)

Rank	Category	1	2	3	4	5	6
1	Reputation of the journal	58.02	39.51	1.23	–	–	1.23
2	Impact factor	38.27	44.44	7.41	6.17	2.47	1.23
3	Fast publishing process	20.99	51.85	12.35	11.11	2.47	1.23
4	Costs	14.81	27.16	25.93	19.75	11.11	1.23
5	Open Access	9.88	33.33	29.63	22.22	3.70	1.23
6	Linkage to data	4.94	30.86	30.86	23.46	7.41	2.47

Note: Hyphen means that there was no answers in that category. 1 = Very important; 2 = Important; 3 = Neither important nor unimportant; 4 = Not that important; 5 = Not at all important; 6 = No answer. n=81.

The most important factors are reputation of the journal and impact factor while open access and linkage to data are considered to be of lesser importance (see Table 21). This is not good news for research data centres as these two factors are considered to be supporting arguments for using such facilities in the manner of enhanced publications. Obviously the advantages of such processes have to be advertised more to the scientific communities.

To create enhanced publications the scientist must be willing to share their data as they will be connected to the journal article by landing page or direct access. 55.55% of researchers to share their data, the results were slightly better than expected. The participants in the survey show in comparison to other studies a higher willingness to share data. Nevertheless still almost half of them are not willing to give data to other researchers (44.44% of the participants). This filter was used to first ask the data sharing researchers with whom they share their data and later the non-sharing researchers were asked about the motives of their unwillingness. Nevertheless first the results of the question with whom they share their data.

Table 22. Recipients of data

Sharing data	Frequency (%)
I deliver to data to research data center or archive.	44.44
I hand over the data to a designated person at my organization.	33.33
I share my data only with researchers I know.	35.56
I share my data with other researchers from the same field.	37.78
I share my data with other researchers from other fields.	28.89
I share my data only with researchers from my organization.	22.22
I share my data with other persons who are interested in my research.	40.00

Note: The participants could mark multiple responses. N=46.

The answers are completely mixed with no clear preference (see Table 22). Some of the researchers (40%) share their data with all persons who are interested while some (22.22%) only with other persons in their organization or they personally know (35.56%). A slight majority (44.44%) is willing to deliver their data to research data centres. It might therefore make sense to explain all other modes could be made possible in such a facility as well.

In the survey almost half of the participants are not willing to share the data. Table 23 contains their opinion on different reasons not to do this. Dominating are the organizational or resource reasons like “too much effort” and “no support from employer”. This means those researchers could be reached with better data management software which takes over a majority of the tasks or better support processes from research data centres. Interesting nevertheless is the reason “somebody could use my data before me”. Though the researchers do not strongly agree on this point (making it the last position in the ranking) still 14 researchers agreed on this view. Though fears of somebody else “stealing the research” are less dominant than organizational problems, they still exist and have to be addressed by research data facilities and archives to motivate more researchers to store their data.

Table 23. Reasons to not share data (%)

Rank	Opinion	1	2	3	4	5	6
1	Preparation of data needs to much effort	22.22	41.67	22.22	2.78	5.56	5.56
2	I have no resources or support from my employer to hand in the data	22.22	38.89	27.78	5.56	2.78	2.78
3	My organization or I have security concern	13.89	27.78	25.00	11.11	19.44	2.78
4	I do not know where to hand in my data for archiving	8.33	22.22	38.89	11.11	16.67	2.78
5	Somebody could misinterpret my results	8.33	13.89	33.33	19.44	19.44	5.56
6	Somebody could use my data before me	2.78	38.89	13.89	22.22	16.67	5.56

1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree; 6 = No answer. n=36.

Last but not least all researchers were asked what could be motivating factors for handing in their data (see Table 24). This also includes the researchers who before stated they do not hand in their data. In the advanced section this will be analysed by groups, but here the overall result first. The clear motivating factor for researchers is the citation as a secondary publication as this is the value which gives them scientific reputation and therefore a career benefit. This means services like DataCite clearly hit a demand from researchers and should therefore be considered. Nevertheless also the other factors like “list of publications” or “information about data access” still generate enough interest to consider them for data management software. The only category with less interest is the “financial compensation”. Researchers are obviously not too interested in this.

Table 24. Motivating factors to share data (%)

Rank	Category	1	2	3	4	5	6
1	Data is cited as secondary publication	37.04	41.98	8.64	4.94	2.47	4.94
2	Support from the employer	32.10	38.27	12.35	7.41	4.94	4.94
3	Co-authorship in secondary publications based on the data	20.99	43.21	14.81	6.17	9.88	4.94
4	Archive provides list of publications based on the data	18.52	53.09	16.05	2.47	6.17	3.70
5	Archive informs data producer who accessed the data	14.81	38.27	29.63	8.64	4.94	3.70
6	The data producer receives a financial compensation	8.64	14.81	34.57	19.75	16.05	6.17

1 = Very important; 2 = Important; 3 = Neither important nor unimportant; 4 = Not that important; 5 = Not at all important; 6 = No answer. n=81.

To check the hypotheses of more senior researchers being reluctant to hand over their data the Kruskal Wallis H test was also used for analyzing the opinion of the groups towards data sharing. Other than expected the cross tabulation shows no differences in the frequencies between the groups (see Table 25). In general it can be stated the willingness to share data is not too prominent in all four groups. Research

data centers and archives therefore need more convincing towards researchers if they want to acquire data for secondary analysis.

*Table 25. Cross tabulation on willingness for data sharing – example of data delivery to research data centres (%)*

	Academic status			
	PhD students (n=25)	No PhD (n=14)	PhD (n=24)	Professor (n=16)
Not chosen	22.8	12.65	24.05	17.72
Yes	8.86	5.06	6.33	2.53

### **3.9. Knowledge of the metadata standards' software products**

The next chapter asks the researchers which software products they know as this indirectly also gives some input about metadata standards as some of the products support them (e.g. TAO and Questify support QTI). Table 26 shows the ranking which software products are known to the researchers. The most known products are the web survey tools Limesurvey, Unipark and Surveymonkey followed by the computer-based assessment systems CBA Itembuilder, TAO and eDia. TAO is actually the only product in the first five places to support a metadata standard. It is therefore advisable for the producers of the other software to consider import and export routines for metadata standards as they are in demand from researchers to make data management processes easier.

Table 26. Ranking of software products (%)

Rank	Software	I know this software well	I have a basic understanding of what the software does	I have only heard of this software	I do not know this software	No answer
1	Limesurvey	15.00	26.67	21.67	25.83	10.83
2	Globalpark / Unipark	12.50	16.67	14.17	43.33	13.33
3	Surveymonkey	10.83	31.67	16.67	29.17	11.67
4	CBA Itembuilder	10.83	17.50	5.83	53.33	12.50
5	Testing Assiste par Ordinateur (TAO)	10.83	10.83	9.17	55.83	13.33
6	eDia	5.83	3.33	4.17	74.17	12.50
7	Blaise	5.00	10.00	14.17	58.33	12.50
8	Nipo / Odin / Compass	2.50	8.33	13.33	61.67	14.17
9	Rogatus Survey / Repository	1.67	12.50	10.83	61.67	13.33
10	CASES	1.67	3.33	8.33	74.17	12.50
11	Questify	0.83	5.00	8.33	72.50	13.33
12	Questionmark	0.83	4.17	9.17	73.33	12.50
13	Voxco	0.83	2.50	8.33	75.00	13.33
14	Testdrive	–	2.50	11.67	71.67	14.17
15	MyItemwriter	–	2.50	3.33	80.83	13.33

Note: Hyphen means that there was no answers in that category. n=120.

### 3.10. Conclusions from the first data collection

The results from the first data collection are very promising and fit to the practical experience of the author from years of data management. While the descriptive statistics about the overall sample are the results from the group-based analytics can only be viewed as a preview as the field strength according to G\*Power is too low or the sample sizes in the individual groups too small. Therefore the decision was taken to strengthen the results by more field work. The following two tables 27 and 28 illustrate this problem. The sample sizes are too small to calculate significant results.



Table 27. Sample by academic status – only first dataset

	PhD students	No PhD	PhD	Professors	Total	Missing
1 <sup>st</sup> survey	25	14	24	16	79	51

Table 28. Sample by research area – all datasets

	Social scientist	Educational scientist	Both	Total	Missing
1 <sup>st</sup> survey	23	30	24	77	53

Nevertheless, as shown in the sections before a preliminary analysis of these groups has been performed and the following items showed differences:

- Metadata standards play an important role in data archives. We are interested in how often researchers use them in their surveys and projects. Please specify to what extent do you use these metadata standards? (MS\_01)
- How much do you agree on the following statements regarding metadata standards? (MS\_06)
- How often do you use the following data sources for your research? (DA\_01)
- How often do you provide these different types of metadata with a questionnaire or cognitive item? (IT\_02)
- How often do you use the following data sources for your research? (DC\_14)
- With whom do you share your data? (PU\_04)
- Please indicate the reasons why you do not share your data. (PU\_05)

This led to the idea to create a second data collection with less items (only the seven items which showed differences plus additional items for the identification including the participation in the first study) which should strengthen the descriptive statistics with more cases while offering also enough participants to form the desired groups.

*“You can have data without information, but you cannot have information without data.”*  
(Daniel Keys Moran)

## **4. Results from the second data collection**

As described in the section before a second data collection was needed to strengthen the results of the first data collection especially for deeper analytical methods like cluster analysis or data mining as the sample sizes were not sufficient. Though the advertisement effort for participation in the study was high the response rate was very low. The exact number cannot be evaluated as most of the mailing lists do not provide information on the number of recipients, but from the known numbers the first data collection has a response rate of well under 1% rather going down to less than 0.1%. Educational researchers and scientists seem to be a group which is hard to acquire in comparison to educational assessments at schools where convincing a headmaster or headmistress or a lead teacher provides for a whole block of new participants.

### **4.1. Improving sample sizes**

To improve the numbers in the second data collection the following paradigms have been considered:

- The survey should be much shorter as the first data collection had a lot of break-offs
- The number of items should be reduced only to the ones which showed significant results in the first data collection
- The number of mailing lists and recipients has to be improved
- More reminders for participation
- A longer data collection period
- If it is possible attach the data collection to a running large-scale project

Except for the last goal the other goals could be reached. The second data collection took place from March 9th, 2015 to July, 19th 2015 which is a much longer period. The number of mailing lists has been doubled and the amount of reminders for participation tripled.

The questionnaire from the first data collection has been shortened to seven questions in total which showed the most promising results. Therefore the survey time shrunk from 30-35 min to 5-7 min. This fact has been outlined specifically in the invitation to the survey. To enable a later merging of the results from both data collections recipients have been asked to identify in an additional question if they already took part in the first data collection so the double entries can be cleaned out of a potential combined dataset.

As expected the second data collection had a much higher participation rate due to the changes. In total 298 persons participated of which n=198 remained after data cleaning. Data cleaning consisted of eliminating people who identified themselves as not being a researcher by a filter question and persons who answered less than half of the questions. The second data collection provided a higher number of finished interviews than the first data collection but nevertheless in regards to the almost tripled efforts plus heavily reduced workload not satisfactory. The data collection was stopped in the moment reminders and new lists did not result in any new participation. As an example in the last phase of the

data collection a reminder for the study was sent to a prominent educational conference with several thousand persons on the mailing list which resulted in one new interview which was broken off after two questions out of seven.

The acquisition process therefore proved to be quite extensive and very difficult in comparison to school studies. The interesting question is if this non participation of scientists bases on pure time constraints or if the overall topic of data management is not very interesting to them.

#### 4.2. Intermediate results

To illustrate any changes between both data collections here an overview of the results from the second one classified as intermediate results. Due to the again low response rate advanced analytics will be performed on a combined dataset from both cycles.

Here the results of the data collection with n=198 participants. Unlike the results from the first data collection in Chapter 3 most of the following tables will not list any n-values in the notes as the second data collection did not contain any routing except for the question on not sharing data.

Table 29. Frequency (%) of Metadata Standards – second data collection

Standards	Usage “often”	Usage “rarely”	Usage “never”	Usage “Do not know”
Data Documentation Initiative (DDI)	16.7	11.6	24.7	47.0
Dublin Core	8.2	8.7	20.0	63.1
Learning Object Model (LOM)	1.0	4.2	21.4	73.4
Learning and Test Interoperability (LTI)	0.5	1.6	19.8	78.1
Machine Readable Cataloguing (MARC)	2.1	5.8	19.4	72.8
Metadata Encoding and Transmission Standard (METS)	2.1	5.2	16.8	75.9
QueDex	0.5	1.6	17.6	80.3
Questionnaire and Text Interoperability (QTI)	2.6	4.2	17.3	75.9
Shareable Content Object Reference Model (SCORM)	1.6	4.1	22.3	72.0
Statistical Data and Metadata Exchange (SDMX)	3.1	2.1	24.9	69.9
Text Encoding Initiative (TEI)	2.1	3.6	19.1	75.3

Table 29 shows the second data collection has a higher variety in usage of different metadata standards. Standards like QueDex which did not have any users in the first data collection at least gain some usage. Nevertheless the ranking of the prominence of certain standards like DDI or Dublin Core does not change. Also, the number of unknowns stay high even with the more popular standards. Metadata standards still are not a field of expertise for researchers. The next question therefore asked about the importance of metadata standards.

Table 30. Distribution of the statements regarding metadata standards – second data collection

Statement	1	2	3	4	5	6
Identification of metadata standards (MS0601)	3.3	15.3	19.0	13.6	7.1	41.8
Metadata standards and domains (MS0602)	6.0	10.9	9.2	20.1	24.5	29.3
Importance of metadata standards (MS0603)	24.6	32.8	8.2	1.6	0.5	32.2
Complexity of studies in regards to metadata standards (MS0604)	1.6	3.3	13.0	27.7	20.1	34.2

1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree; 6 = Do not know

Table 30 shows no big deviation from the first data collection. Researchers still recognize the importance of metadata standards and feel them to be adequate for their purposes. Nevertheless there are still difficulties in choice and handling.

Table 31. Frequencies of data sources (%) – second data collection

Data Source	Always	Often	Sometimes	Rarely	Never
Structure cognitive test (e.g. computer-based assessment)	11.7	24.6	22.9	11.7	29.1
Structured questionnaire (e.g. multiple choice)	31.1	46.1	12.8	3.9	6.1
Qualitative text data (e.g. narrative interviews, oral history)	6.1	17.3	18.4	27.9	30.2
Experimental designs	5.1	20.2	26.4	19.1	29.2
Observational data (e.g. recordings, transcripts)	5.6	23.5	25.1	21.2	24.6

Table 31 still shows the preference of data users towards quantitative designs like questionnaires and cognitive items. Qualitative datasets or observational data are much less common. Also there is not too much emphasis on experimental designs. This does not mean they can be neglected in regards to data management, but it shows solving some of the issues in regards to questionnaires and cognitive items will affect more users and thus might be selected for the first stages of improvements.

Table 32. Frequencies of the different metadata provided (%) – second data collection

Category	Always	Often	Sometimes	Rarely	Never	Do not know
Analytic scripts	12.7	20.5	17.5	19.9	20.5	9.0
Answer categories	39.4	25.5	12.7	8.5	7.3	6.7
Concepts to be measured	23.3	23.9	22.7	11.0	10.4	8.6
Flow logic	11.6	21.3	11.6	11.6	26.8	17.1
Fragments of software	4.3	14.0	13.4	20.7	36.6	11.0
Interviewer instructions	23.1	24.4	20.6	6.3	15.6	10.0
Layout definitions	5.0	16.3	13.8	20.6	29.4	15.0
References to research literature	27.0	18.4	20.9	13.5	14.1	6.1
Scoring rules	14.2	23.5	16.7	10.5	21.6	13.6
Statistical values from previous studies	5.5	21.3	23.2	14.0	25.6	10.4
Stimuli in reusable formats	4.3	13.0	13.0	18.6	34.2	16.8
Variable definitions	29.6	20.4	15.4	12.3	11.7	10.5

Table 32 confirms the results from the first data collection. Answer categories, concepts, variable definitions, interviewer instruction and references to research literature are more likely to be preserved, while the more complex technical components like software fragments, layout definitions and stimuli are less likely. Spoken from the perspective of a data archive this means calculations on the dataset might be possible while the reuse of items for later studies is not possible after some years as the means are not preserved. Items therefore have to be reconstructed if they are supposed to be used in later studies. Another question for preservation would be: are datasets still understandable for secondary researchers if the original item cannot be seen anymore or the definition only exists in an explanation or screen shot?

Table 33. Frequency of the access to datasets (%) – second data collection

Category	Always	Often	Sometimes	Rarely	Never	Do not know
Public Use Files	7.9	18.3	25.6	14.0	32.9	1.2
Scientific Use Files	5.5	24.4	22.6	18.9	27.4	1.2
Raw datasets	10.4	34.8	17.7	12.8	22.6	1.8
Data enclaves, secure data services or Virtual Research Environments	0.6	8.0	7.4	17.8	58.3	8.0
Relational databases	3.7	11.0	15.9	11.6	53.0	4.9
Data Warehouses or analytical databases	1.2	4.3	11.0	19.5	54.3	9.8
Remote calculation or job submission	0.6	1.8	4.9	7.9	71.3	13.4
Personal extracts created by portal websites	0.6	1.8	12.2	14.6	60.4	10.4

Table 33 shows how researchers access datasets for secondary usage. Still very surprising the prime access is also in the second data collection the huge amount of answers for the raw datasets. Research data centres are normally not allowed to hand out those files due to privacy laws. Normally they have to go through multiple stages of data cleaning and anonymization. This might be an indicator the scientists are not aware what they consider to be raw datasets are actually cleaned versions.

The next question was only a yes/no variety. Therefore a table was not used. On the question if the researchers share metadata, paradata or data from their research with others the answers were the following:

- 62.6% Yes
- 37.4% No

The ratio is a little better than in the first data collection (55% Yes), but still a third of the researchers is not sharing their data with fellow researchers. Therefore the next question asks with whom researchers are willing to share their data with multiple answer possibilities.

Table 34. Recipients of data (%) – second data collection

Statement	Share data	Don't share
I deliver my data to research data center or archive.	26.8	73.2
I hand over the data to a designated person at my organization.	18.7	81.3
I share my data only with researchers I know.	16.2	83.8
I share my data with other researchers from the same field.	20.2	79.8
I share my data with other researchers from other fields.	11.1	88.9
I share my data only with researchers from my organization.	14.1	85.9
I share my data with other persons who are interested in my research.	21.7	78.3

As can be seen the notion not to share data with different groups of peers is dominating. Still handing out research data to a specialized data centre has the highest value. The next question therefore specifically filtered all researchers who do not want to share data and asked for the reasoning. The number of participants in this specific part was N=59.

Table 35. Reasons to not share data (%) – second data collection

Opinion	1	2	3	4	5
I do not know where to hand in my data for archiving	15.3	22.0	28.8	23.7	10.2
I have no resources or support from my employer to hand in the data	17.2	22.4	29.3	20.7	10.3
Preparation of data needs too much effort	16.9	33.9	35.6	6.8	6.8
My organization or I have security concern	13.6	33.9	23.7	15.3	13.6
Somebody could misinterpret my results	6.8	13.6	25.4	32.2	22.0
Somebody could use my data before me	15.5	24.1	24.1	17.2	19.0

1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree

A main reason is still workload in form of resources or too much effort while also security concerns play a role as a secondary reason. Misinterpretation of the data does not seem to be a big concern.

### 4.3. Intermediate conclusion

The second data collection analysed on its own does not produce surprising results and can be used to be combined with the first data collection to strengthen the sample size. Deeper analysis will therefore be performed rather on this dataset as it contains more cases. The sample size is too small for methods like cluster analysis or data mining. Thus this section only gives a short overview of the results.

*“You can use all the quantitative data you can get,  
but you still have to distrust it and use your own intelligence and judgment.”*  
(Alvin Toffler)

## **5. Final data analytics on the combined dataset**

As described in the sections before the first data collection did not have enough participants to allow for more sophisticated data analytics. For this reason a second data collection took place with a shorter questionnaire and more advertisement to generate a higher response rate. This section describes the results of the combined dataset from the first and second data collection and includes some more advanced data analysis procedures.

### **5.1 Description of the combined dataset**

The combined dataset consists from the 120 completed interviews from the first data collection plus the 198 completed interviews from the second data collection. Both datasets were attached only for question items appearing in both questionnaires. In the second questionnaire there was a question to identify persons who have also participated in the first data collection. These individuals have been cleaned out of the dataset resulting in n=272 completed interviews for the combined dataset. Like in Chapter 4 most questions have been presented to all participants as the combined dataset does not feature the extensive use of routing like in the first data collection. Therefore the value of participants is always n=272 except for the question on not sharing data. Although this amount is still far from being impressive according to G\*Power the field strength was good enough to allow for separation by two areas:

- Academic status

It was possible to separate the participants into four groups (PhD students, researchers without PhD, researchers with PhD and Professors). There were not enough bachelor or master students to form a students' group.

- Research area

Also in this area three groups could be formed (researchers from educational sciences, social sciences or both). There were not enough participants from other disciplines (e.g. computer science as third largest group) to form further groups.

These groups are the basis for the cluster analysis in Section 5.3 and the Data Mining in Section 5.4, but here is first an overview of the descriptive statistics. The following two tables show the composition of the groups used in the combined dataset.



Table 36. Sample by academic status – all datasets

	PhD students	No PhD	PhD	Professors	Total	Missing
1 <sup>st</sup> survey	25	14	24	16	79	51
2 <sup>nd</sup> survey	30	24	29	22	105	37
Total	55	38	53	38	184	88

Table 37. Sample by research area – all datasets

	Social scientist	Educational scientist	Both	Total	Missing
1 <sup>st</sup> survey	23	30	24	77	53
2 <sup>nd</sup> survey	43	39	22	104	38
Total	69	66	46	181	91

The reliability of the newly formed questionnaire and the resulting combined dataset has been checked as well. Cronbach- was used for this purpose. The resulting value is 0.82 which is an acceptable limit for a questionnaire.

## 5.2 Descriptive statistics on the combined dataset

As to be expected from the first and second data collection the combined results also show the same results as the individual datasets which differed only in minor details. This can be proved already in the first table about knowledge on metadata standards.

Table 38. Frequency (%) of Metadata Standards – combined dataset

Standards	Usage “often”	Usage “rarely”	Usage “never”	Usage “Do not know”
Data Documentation Initiative (DDI)	15.0	10.9	27.1	47.0
Dublin Core	7.6	8.7	19.3	64.4
Learning Object Model (LOM)	1.2	3.9	19.4	75.6
Learning and Test Interoperability (LTI)	1.2	1.6	17.8	79.5
Machine Readable Cataloguing (MARC)	2.3	6.2	16.0	75.5
Metadata Encoding and Transmission Standard (METS)	1.6	3.5	17.9	77.0
QueDex	0.4	0.8	15.4	83.4
Questionnaire and Text Interoperability (QTI)	3.5	2.7	16.9	76.9
Shareable Content Object Reference Model (SCORM)	0.8	5.4	17.4	76.4
Statistical Data and Metadata Exchange (SDMX)	2.7	2.3	23.5	71.5
Text Encoding Initiative (TEI)	1.5	3.1	17.8	77.6

The results from the second data collection allowed for a little more variety than the first data collection. As both datasets are now combined this effect is moderated by the lower spread in the first dataset. Nevertheless the combination of datasets does not change the results from the first and second data collection which states a high level of uncertainty and lack of details in regards to metadata standards. Metadata standards are largely unknown to researchers and the ranking stays stable in all three datasets.

Though metadata standards are mostly unknown to researchers they still keep their importance which can be shown from Table 39:

Table 39. Distribution of the statements regarding metadata standards – combined dataset

Statement	1	2	3	4	5	6
Identification of metadata standards (MS0601)	3.5	15.7	15.4	14.6	9.4	41.3
Metadata standards and domains (MS0602)	4.7	11.0	10.2	18.9	25.2	29.9
Importance of metadata standards (MS0603)	21.7	38.3	9.9	1.2	1.2	27.7
Complexity of studies in regards to metadata standards (MS0604)	2.0	4.3	12.6	30.3	19.7	31.3

1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree; 6 = Do not know

Metadata standards are recognized to be important, but researchers feel their knowledge and handling not to be adequate. This again leads to the overall question if researchers need to be the persons to be in charge or the driver of these processes. To identify which kinds of data sources the researchers use they were asked to specify their most common research design.

Table 40. Frequencies of data sources (%) – combined dataset

Data Source	Always	Often	Sometimes	Rarely	Never
Structure cognitive test (e.g. computer-based assessment)	10.7	22.2	17.9	15.9	33.3
Structured questionnaire (e.g. multiple choice)	35.2	42.7	10.7	3.6	7.9
Qualitative text data (e.g. narrative interviews, oral history)	6.3	19.0	21.8	26.2	26.6
Experimental designs	5.6	16.7	24.7	23.9	29.1
Observational data (e.g. recordings, transcripts)	5.6	18.7	29.0	21.8	25.0

Not surprisingly Table 40 shows a heavy preference towards questionnaires and structured cognitive tests with the first data collection tipping the scale even more towards quantitative research methods. This result can also explain why metadata standards and data description are underdeveloped in these areas. As stated before these kinds of research design should not be ignored in data management, but it leads to the question if an archive can provide a service structure which can accommodate the needs for managing quantitative and qualitative datasets. Nevertheless this question is out-of-bounds for this dissertation as it cannot be discussed on basis of these datasets.

As in the previous data collections the researchers were asked which kind of different metadata they usually provide.

*Table 41. Frequencies of the different metadata provided (%) – combined dataset*

Category	Always	Often	Sometimes	Rarely	Never	Do not know
Analytic scripts	16.8	18.8	18.3	17.3	22.1	6.7
Answer categories	44.4	20.8	16.9	6.8	5.8	5.3
Concepts to be measured	23.9	27.3	23.4	9.3	10.2	5.9
Flow logic	15.5	18.4	17.5	9.7	24.3	14.6
Fragments of software	3.9	13.1	15.0	18.9	38.3	10.7
Interviewer instructions	26.1	25.1	17.7	10.3	12.3	8.4
Layout definitions	9.9	18.2	18.2	18.2	24.1	11.3
References to research literature	26.3	18.0	23.9	9.8	17.1	4.9
Scoring rules	20.1	18.6	17.6	8.3	20.6	14.7
Statistical values from previous studies	6.8	22.0	22.0	16.1	26.3	6.8
Stimuli in reusable formats	5.4	12.7	14.2	19.6	34.8	13.2
Variable definitions	28.4	22.1	16.7	7.8	12.7	12.3

As Table 32 from the second data collection originally confirmed the order and values from the second data collection it is no surprise Table 41 contains almost the same values. Answer categories, concepts, variable definitions, interviewer instruction and references to research literature are more likely to be preserved while the technically more advanced metadata are not. The more technically oriented metadata seem to be out-of-scope for researchers.

Another important question for preservation next to metadata is the usage of data in different environments or how the researchers access the data itself as this defines preservation strategies.

Table 42. Frequency of the access to datasets (%) – combined dataset

Category	Always	Often	Sometimes	Rarely	Never	Do not know
Public Use Files	8.8	22.8	25.7	12.3	29.3	1.2
Scientific Use Files	9.3	25.6	23.8	18.6	21.5	1.2
Raw datasets	8.8	31.6	18.1	15.2	25.1	1.2
Data enclaves, secure data services or Virtual Research Environments	1.7	8.1	8.7	18.0	54.7	8.7
Relational databases	2.9	11.6	15.7	12.8	51.8	5.8
Data Warehouses or analytical databases	0.6	4.1	11.6	16.9	54.1	12.8
Remote calculation or job submission	0.6	1.8	6.4	8.2	65.5	17.5
Personal extracts created by portal websites	0.6	2.9	9.9	14.5	59.3	12.8

Table 42 shows a tendency towards file-based approaches (Public Use Files, Scientific Use Files and Raw Datasets) which is again no deviation from the results of the individual data collections. Remote calculation, Virtual Research Environments or personal extracts by variable shopping basket systems are virtually unknown. The result is especially interesting as in international studies the datasets for dissemination are provided by data collection or survey organizations and undergo cleaning processes before release. Either the researchers only analyse small internal datasets or they are not aware their “raw data” is actually not directly from the field as they assumed.

The next question asked if the researchers shared metadata, paradata or data from their research with others. Like in the individual data collections almost half of the participants did not want to participate in such processes. Here the combined result:

- 59.2% Yes
- 40.8% No

The participants in the second data collection were a little more open with almost two-thirds sharing their data, but here the result from the first data collection (55%) mediated the result. From a data archive perspective all the results are actually disappointing. The next question therefore asked with whom they usually share or do not share their data.

Table 43. Recipients of data (%) – combined dataset

Statement	Share data	Don't share
I deliver my data to a research data center or archive.	20.2	79.8
I hand over the data to a designated person at my organization.	14.0	86.0
I share my data only with researchers I know.	14.3	85.7
I share my data with other researchers from the same field.	15.4	84.6
I share my data with other researchers from other fields.	9.6	90.4
I share my data only with researchers from my organization.	10.7	89.3
I share my data with other persons who are interested in my research.	16.5	83.5

As can be seen sharing the data (59.2%) does not necessarily mean delivering it to a research data facility or data archive (20.2%). It can be expected the values might even go down further if there was a separation between institutional archive and external archive. This of course raises the question why researchers do not share their data. The following question only was asked by filter to the n=79 researchers (40.8%) who answered not to share their data.

Table 44. Reasons to not share data (%) – combined dataset

Opinion	1	2	3	4	5
I do not know where to hand in my data for archiving	12.7	19.0	36.7	17.7	13.9
I have no resources or support from my employer to hand in the data	19.2	29.5	29.5	14.1	7.7
Preparation of data needs too much effort	17.9	38.5	29.5	6.4	7.7
My organization or I have security concerns	13.9	32.9	24.1	11.4	17.7
Somebody could misinterpret my results	7.7	15.4	30.8	25.6	20.5
Somebody could use my data before me	12.8	28.2	19.2	20.5	19.2

1 = Strongly agree; 2 = Agree; 3 = Neither agree nor disagree; 4 = Disagree; 5 = Strongly disagree

Obviously preparing the data for an external usage needs too much effort and there are also trust issues with other organizations. Research data centres and archives might use these results to improve their processes. The hurdles for researchers to hand in their data might simply be too big.

### 5.3 Differences between groups

Though the overall results provide some interesting insights there were researchers from different backgrounds participating as well as persons who have been in academia for different amounts of time. The question is if scientific field or status in academia have any impact on the results. To answer those questions has been performed mainly by the Mann-Whitney U test (Mann & Whitney, 1947) or Kruskal-Wallis H Test (Kruskal & Wallis, 1952) as they are considered to be the best fit for questionnaires like they were used in both data collections.

### 5.3.1 Academic status

As described before the combined response rate allowed for a separation into four groups – PhD students, researchers without PhD, researchers with PhD (post-docs) and Professors.

Here are the results to the individual questions based on groups

- Usage of the different type of metadata standards

There were no significant differences between the groups.

- Opinion regarding metadata standards

There were significant differences between the groups in regards to importance of metadata standards ( $\chi^2(3)=7.88, p<.05$ ). In detail there were significant differences between students and the researchers without PhD ( $U=1030, p<.05$ ) plus PhD Students and researchers with PhD ( $U=712, p<.05$ ). The PhD students evaluated the importance of the metadata standards lower than the other two groups. In other cases there were no significant differences between the groups.

Table 45. Averages and standard deviations of the statements regarding metadata standards

Statement	PhD students (n=53)	No PhD (n=38)	PhD (n=51)	Professors (n=38)
Identification of metadata standards (MS0601)	4.77(1.64)	4.16(1.59)	4.14(1.63)	4.47(1.57)
Metadata standards and domains (MS0602)	4.21(1.69)	4.66(1.02)	4.57(1.32)	4.05(1.75)
Importance of metadata standards (MS0603)	3.43(2.07)	2.34(1.59)	2.51(1.64)	3.13(2.07)
Complexity of studies in regards to metadata standards (MS0604)	4.66(1.3)	4.02(1.37)	4.43(1.3)	4.71(0.97)

- Usage of data sources

There were no significant differences between the groups.

- Different kinds of metadata provided by researchers

There were significant differences between the groups in analytic scripts ( $\chi^2=8.63, p<.05$ ), concepts to be measured ( $\chi^2=8.82, p<.05$ ), fragment of software ( $\chi^2=8.73, p<.05$ ) and scoring rules ( $\chi^2=8.95, p<.05$ ).

Analytic scripts: The researchers without PhD, researchers with PhD and Professors provide more often the analytic scripts than the PhD students ( $U_{\text{noPhD}}=545.5, p<.05; U_{\text{PhD}}=797, p<.05, U_{\text{prof}}=592.5, p<.05$ ). In the other cases there were no significant differences.

Concepts to be measured: The researchers with PhD and Professors provided more often concepts to be measured than PhD students ( $U_{\text{PhD}}=912, p<.05, U_{\text{prof}}=597, p<.05$ ). In the other cases there were no significant differences.

Fragments of software: The researchers without PhD and Professors provide more often the fragments of software than PhD students ( $U_{\text{noPhD}}=504, p<.05; U_{\text{prof}}=610.5, p<.05$ ). In the other cases there were no significant differences.

Scoring rules: The researchers with PhD and Professors provide more often the scoring rules than PhD students ( $U_{\text{phd}}=765.5, p<.05, U_{\text{prof}}=583.5, p<.05$ ). In the other cases there were no significant differences.

In summary the Professors delivered more often additional metadata while the PhD students at least in these data collections do not seem very prone to the practices of data documentation.

Table 46. Averages and standard deviations of the different metadata provided

Category	PhD students (n=45)	No PhD (n=30)	PhD (n=42)	Professors (n=36)
Analytic scripts	3.89(1.49)	3.16(1.46)	3.15(1.58)	2.97(1.62)
Answer categories	2.53(1.74)	2.09(1.23)	2.26(1.64)	1.97(1.34)
Concepts to be measured	3.29(1.65)	2.74(1.49)	2.52(1.5)	2.32(1.36)
Flow logic	4.1(1.65)	3.22(1.84)	3.43(1.81)	3.62(1.52)
Fragments of software	4.53(1.14)	3.78(1.31)	4.11(1.39)	3.79(1.42)
Interviewer instructions	3.51(1.87)	2.72(1.51)	2.91(1.81)	2.67(1.31)
Layout definitions	4.04(1.56)	3.47(1.67)	3.52(1.66)	3.57(1.41)
References to research literature	3.23(1.74)	3.06(1.58)	2.54(1.54)	2.54(1.48)
Scoring rules	3.98(1.88)	3.25(1.63)	3.07(1.74)	2.89(1.63)
Statistical values from previous studies	3.98(1.51)	3.38(1.16)	3.46(1.49)	3.19(1.43)
Stimuli in reusable formats	4.34(1.45)	3.97(1.38)	3.98(1.59)	3.76(1.34)
Variable definitions	3.42(1.81)	2.39(1.33)	3.02(1.96)	2.68(1.67)

- frequency of the access to datasets

There were significant differences between the groups in the usage of relational databases ( $\chi^2=9.28, p<.05$ ). The researchers without PhD, researchers with PhD and Professors used more often relational databases than the PhD students ( $U_1=430, p<.05; U_2=618, p<.05; U_3=442.5, p<.05$ ). In the other cases there were no significant differences between the groups.

Table 47. Averages and standard deviations of the access to datasets

Category	PhD students (n=40)	No PhD (n=31)	PhD (n=40)	Professors (N=30)
Public Use Files	3.85(1.31)	3.13(1.26)	3.29(1.45)	3.4(1.33)
Scientific Use Files	3.49(1.47)	2.9(1.22)	3.02(1.37)	3.67(1.18)
Raw datasets	3.39(1.39)	3.22(1.38)	3.07(1.44)	3.23(1.33)
Data enclaves, secure data services or Virtual Research Environments	4.51(1.19)	4.19(1.14)	4.21(1.33)	4.73(0.64)
Relational databases	4.66(1.02)	3.77(1.48)	4.05(1.31)	4.03(1.25)
Data Warehouses or analytical databases	4.88(0.68)	4.35(1.28)	4.62(1.13)	4.67(0.88)
Remote calculation or job submission	5.0(0.87)	4.87(0.67)	4.83(0.97)	5.0(0.56)
Personal extracts created by portal websites	4.85(0.76)	4.45(0.93)	4.59(1.19)	4.87(0.68)

### 5.3.2 Research area

Similarly to the academic status the researchers were also asked about their field. From the amount of participants it was possible to create three groups – researchers who come from the field of educational sciences, social sciences or both. Here their answers on the different items.

- Usage the different type of metadata standards

An interesting question is which metadata standards are more common in the different disciplines.

Table 48 - Averages and standard deviations of usage of metadata standards

Standards	Social scientist (n=65)	Educational scientist (n=59)	Both (n=46)
Data Documentation Initiative (DDI)	3.61(0.72)	2.46(1.14)	3.08(1.01)
Dublin Core	3.75(0.53)	3.03(1.11)	3.52(0.84)
Learning Object Model (LOM)	3.75(0.66)	3.73(0.48)	3.57(0.72)
Learning and Test Interoperability (LTI)	3.73(0.64)	3.77(0.42)	3.76(0.52)
Machine Readable Cataloguing (MARC)	3.84(0.51)	3.49(0.72)	3.65(0.82)
Metadata Encoding and Transmission Standard (METS)	3.84(0.41)	3.54(0.74)	3.67(0.76)
QueDex	3.86(0.38)	3.67(0.51)	3.83(0.53)
Questionnaire and Text Interoperability (QTI)	3.62(0.78)	3.74(0.44)	3.48(0.96)
Shareable Content Object Reference Model (SCORM)	3.69(0.63)	3.69(0.49)	3.67(0.7)
Statistical Data and Metadata Exchange (SDMX)	3.76(0.63)	3.51(0.64)	3.76(0.57)
Text Encoding Initiative (TEI)	3.83(0.48)	3.52(0.69)	3.74(0.61)

Table 49. Usage of different kinds of metadata – Kruskal-Wallis test

Category	2	p
Data Documentation Initiative (DDI)	38.20	<.01
Dublin Core	19.73	<.01
Machine Readable Cataloguing (MARC)	12.93	<.05
Metadata Encoding and Transmission Standard (METS)	7.68	<.05
QueDex	9.28	<.05
Statistical Data and Metadata Exchange (SDMX)	12.18	<.05
Text Encoding Initiative (TEI)	12.15	<.05

As Table 49 shows there were significant differences between the groups in these categories. The social scientists used more often metadata standards than educational scientists. There were no significant differences between the educational scientists and the combined group, except in the usage of the metadata standard DDI ( $U=1082.5, p<.05$ ). The combined group used more often the rest of the metadata standards than social scientists with two exceptions: MARC and METS. In these last two categories there were no significant differences between them.



Table 50. Usage of different kinds of metadata – Mann-Whitney test

Category	Educational – social sciences	Social sciences – both group
Data Documentation Initiative (DDI)	946.0(p<.01)	1052.5(p<.05)
Dublin Core	1385.0(p<.01)	1125.0(p<.05)
Machine Readable Cataloguing (MARC)	1485.0(p<.01)	1153.5(p>.05)
Metadata Encoding and Transmission Standard (METS)	1673.5(p<.05)	1249.0(p>.05)
QueDex	1698.5(p<.05)	1185.0(p<.05)
Statistical Data and Metadata Exchange (SDMX)	1578.0(p<.05)	1112.5(p<.05)
Text Encoding Initiative (TEI)	1567.0(p<.05)	1189.0(p<.05)

- Opinion regarding metadata standards

As table 51 shows there were significant differences between the groups in identification ( $\chi^2=16.02$ ,  $p<.01$ ) and importance ( $\chi^2=19.49$ ,  $p<.01$ ) of metadata standards. The social scientist and the combined group agreed more these statements than educational scientist. There were no significant differences between the social scientist and the combined group.

Table 51 - Averages and standard deviations of the statements regarding metadata standards

Statement	Social scientist (n=67)	Educational scientist (n=64)	Both (n=46)
Identification of metadata standards (MS0601)	4.97(1.46)	3.84(1.7)	4.33(1.48)
Metadata standards and domains (MS0602)	4.45(1.71)	4.25(1.46)	4.41(1.26)
Importance of metadata standards (MS0603)	3.61(1.99)	2.31(1.71)	2.69(1.72)
Complexity of studies in regards to metadata standards (MS0604)	4.69(1.35)	4.48(1.13)	4.17(1.3)

Table 52. Opinion regarding metadata standards – Mann-Whitney test

Category	Educational – social sciences	Educational sciences – both group
Identification of metadata standards	1342.5(p<.01)	1152.5(p<.05)
Importance of metadata standards	1260.0(p<.01)	1143.5(p<.05)

- Usage of data sources

Another important question was which data sources are used by the different groups. Here a summary of the findings.

Table 53 - Averages and standard deviations of data sources

Data Source	Social scientist (n=66)	Educational scientist (n=64)	Both (n=46)
Structure cognitive test (e.g. computer-based assessment)	2.76(1.46)	4.15(1.0)	2.78(1.38)
Structured questionnaire (e.g. multiple choice)	2.1(1.19)	2.04(0.98)	1.76(1.05)
Qualitative text data (e.g. narrative interviews, oral history)	3.15(1.18)	3.47(1.26)	3.8(1.16)
Experimental designs	3.37(1.29)	3.54(1.17)	3.61(1.22)
Observational data (e.g. recordings, transcripts)	2.98(1.27)	3.59(1.12)	3.69(1.03)

There were significant differences between the groups in usage of structure cognitive tests ( $\chi^2=37.54$ ,  $p<.01$ ), qualitative text data ( $\chi^2=8.19$ ,  $p<.05$ ) and observational data ( $\chi^2=11.81$ ,  $p<.05$ ).

Structure cognitive test: Not surprisingly the educational scientist and the both group used significantly more times the cognitive test than social science group ( $U_1=1034.0$ ,  $p<.01$ ;  $U_2=682.0$ ,  $p<.01$ ). There was no significant difference between the educational sciences and the combined group.

Qualitative text data: The combined group used more often the qualitative text data than the educational scientists ( $U=1052.0$ ,  $p<.05$ ). In other cases there were no significant differences.

Observational data: Educational scientists used more often observational data than social scientist ( $U=1599.0$ ,  $p<.05$ ) and the combined group used it more than the educational scientists ( $U=1039.5$ ,  $p<.05$ ). There was no significant difference between the combined group and the social scientists.

- Different kinds of metadata provided by researchers

Table 54 shows the different kinds of metadata and how they relate to the different groups.

Table 54 - Averages and standard deviations of the different metadata provided

Category	Social scientist (n=56)	Educational scientist (n=56)	Both (n=40)
Analytic scripts	3.39(1.69)	3.37(1.53)	3.28(1.52)
Answer categories	2.68(1.72)	1.87(1.27)	2.07(1.35)
Concepts to be measured	2.85(1.65)	2.71(1.49)	2.55(1.33)
Flow logic	4.17(1.43)	3.27(1.77)	3.09(1.76)
Fragments of software	4.34(1.19)	4.01(1.44)	3.86(1.35)
Interviewer instructions	3.08(1.71)	2.91(1.67)	2.78(1.62)
Layout definitions	3.93(1.42)	3.59(1.59)	3.52(1.58)
References to research literature	2.76(1.79)	2.75(1.48)	2.93(1.42)
Scoring rules	2.98(1.84)	3.85(1.68)	3.0(1.61)
Statistical values from previous studies	3.37(1.47)	3.82(1.39)	3.02(1.37)
Stimuli in reusable formats	3.74(1.55)	4.44(1.35)	3.88(1.33)
Variable definitions	3.1(1.91)	2.85(1.73)	2.64(1.56)

Table 55. Metadata provided by researchers – Kruskal-Wallis test

Category	$\chi^2$	P
Answer categories	8.14	<.05
Flow logic	12.11	<.05
Scoring rules	8.83	<.05
Statistical values from previous studies	7.85	<.05
Stimuli in reusable formats	7.82	<.05

There were significant differences between the groups in these categories (Table 55). As the Mann-Whitney test shows there were significant differences between educational and social scientists (Table 56), except in statistical values from previous studies. The social scientists provided more often answer categories and flow logic than educational scientists, but in the remaining categories the situation is reversed. There were no significant differences between the combined group and the educational scientists, except in flow logic. The combined group provided more often flow logic than educational scientists ( $U=809.5$ ,  $p<.05$ ). The combined group provided more often these metadata types than the social scientists (Table 56), except answer categories and flow logic, because in these cases there were no significant differences.

Table 56. Metadata provided by researchers – Mann-Whitney test

Category	Educational – social sciences	Social sciences – both group
Answer categories	1276.5( $p<.05$ )	1110.5( $p>.05$ )
Flow logic	1274.0( $p<.05$ )	1202.5( $p>.05$ )
Scoring rules	1286.5( $p<.05$ )	886.5( $p<.05$ )
Statistical values from previous studies	1457.0( $p>.05$ )	860.0( $p<.05$ )
Stimuli in reusable formats	1289.5( $p<.05$ )	925.5( $p<.05$ )

- Access to datasets

There were significant differences between the groups in access of scientific use files ( $\chi^2=37.54$ ,  $p<.01$ ) and data enclaves ( $\chi^2=8.19$ ,  $p<.05$ ). Social scientists and the both group access more often these data products than educational scientists, and there was no significant difference between both group and social scientist (Table 58).

Table 57. Averages and standard deviations of the access to datasets.

Category	Social scientist (n=49)	Educational scientist (n=56)	Both (n=33)
Public Use Files	3.56(1.43)	3.33(1.42)	3.24(1.19)
Scientific Use Files	3.78(1.13)	2.84(1.35)	2.91(1.23)
Raw datasets	3.37(1.44)	3.43(1.33)	3.12(1.43)
Data enclaves, secure data services or Virtual Research Environments	4.84(0.76)	4.29(1.22)	3.97(1.26)
Relational databases	4.47(1.17)	4.05(1.33)	4.06(1.25)
Data Warehouses or analytical databases	4.82(0.89)	4.61(1.03)	4.55(0.9)
Remote calculation or job submission	4.88(0.94)	4.93(0.88)	4.94(0.7)
Personal extracts created by portal websites	4.76(0.95)	4.58(0.98)	4.79(0.86)

Table 58. Access to datasets – Mann-Whitney test

Category	Educational – social sciences	Educational sciences – both group
Scientific use file	880.0(p<.01)	509.0(p<.05)
Data enclaves	1134.5(p<.05)	486.0(p<.01)

This result is remarkable as scientific use files should be very common also for educational scientists. The question is what they use instead as raw files or primary research data are less common especially in international large scale assessment.

## 5.4 Cluster analysis

Using advanced statistical methods like the Kruskal-Wallis and Mann-Whitney test provided more insight how different groups use data management in their respective disciplines or how academic status changes the behavior. To get more insight the different questions were also analyzed using cluster analysis methodology. Here some results and their interpretation.

As a clustering method nearest neighbor with single linkage was used meaning the distance between clusters is the smallest distance between two data points. As ordinal scales were used chi-squared was the selection for distance measure. The results were plotted as dendrograms which can be found in the following section.

The first cluster analysis is concerned which metadata standards form clusters based on distance. The result is shown in Figure 47.

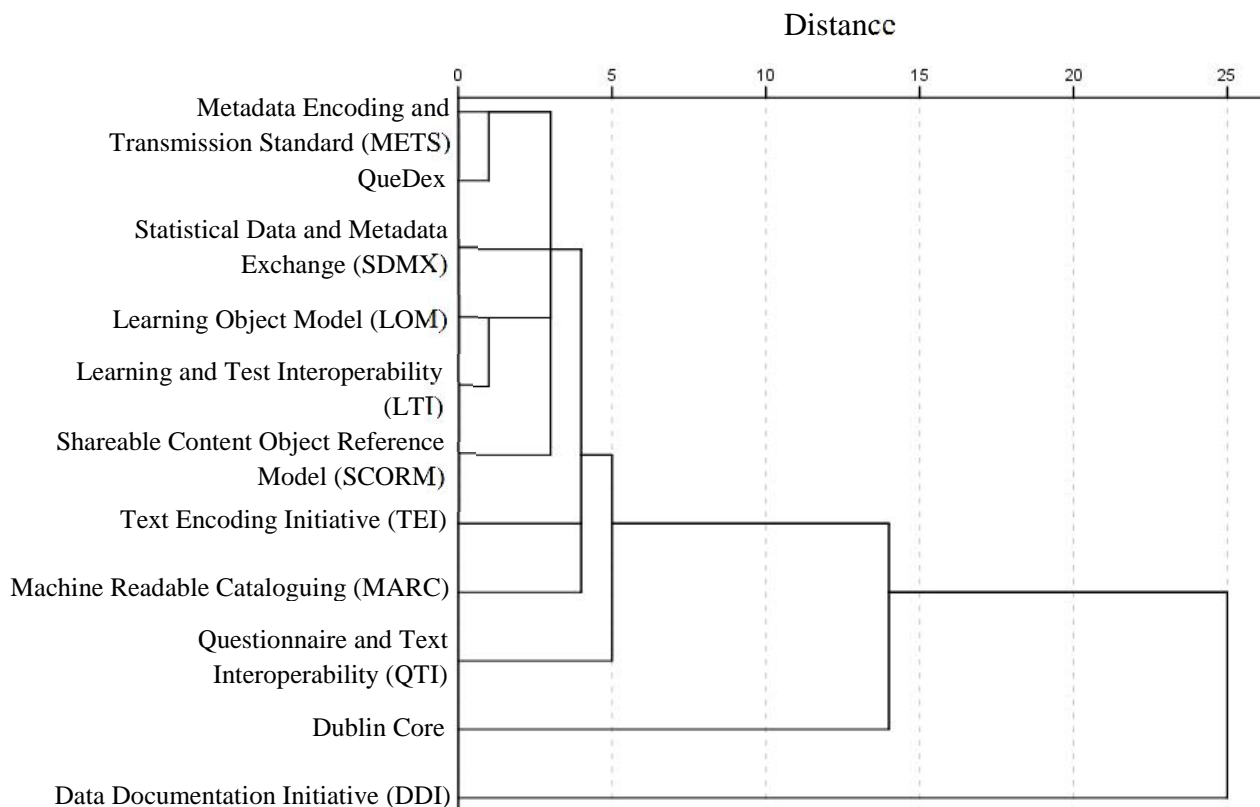
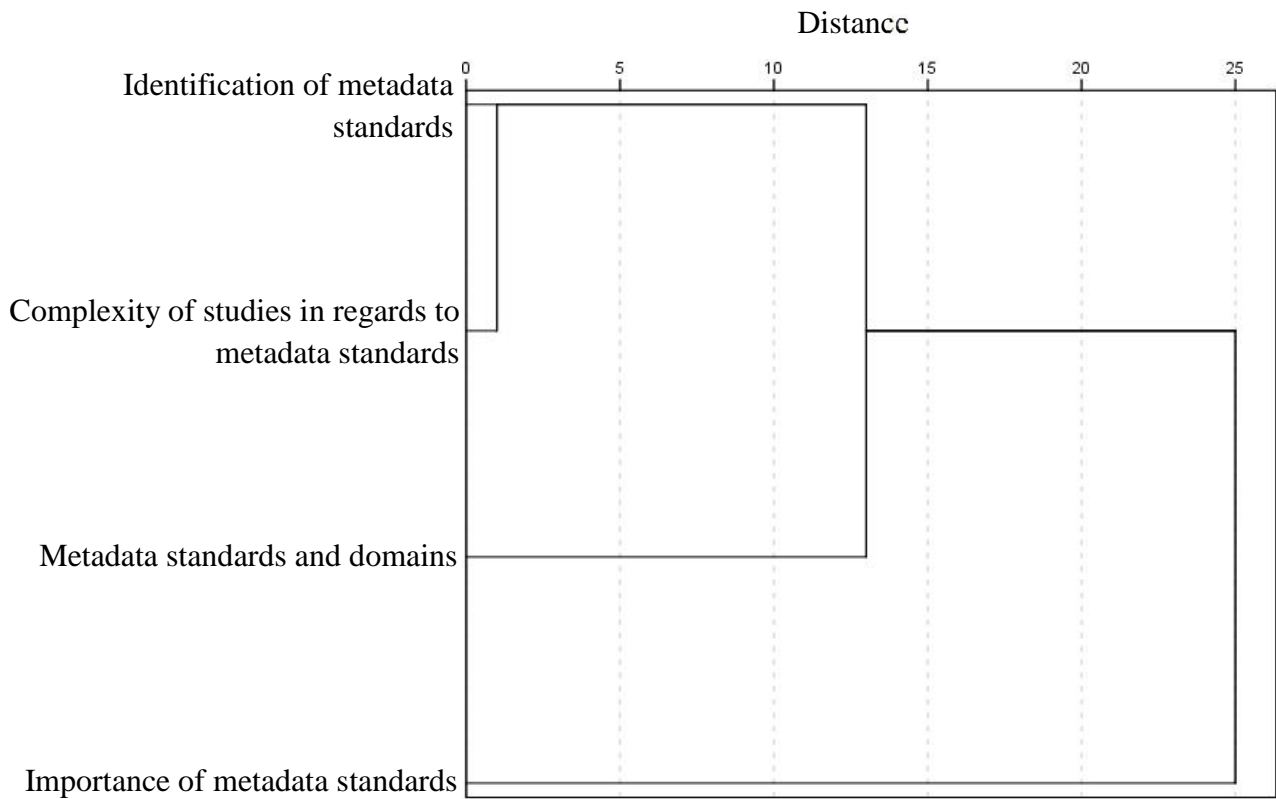


Figure 46. Cluster analysis of the usage of Metadata Standards

There are low level clusters combining METS and QueDex which is not surprising as both relate to describing qualitative data like audio files, video files or text. The same can be said about LOM and LTI and with more distance SCORM which all relate to educational data especially files derived from e-learning platforms. Surprising is the great distance between Dublin Core and DDI versus the rest of the standards. Both standards seem to have a different standing in their respective communities as well as fully different levels of usage which separates them.

Figure 48 shows the cluster analysis about the statements researchers had about the usage of metadata standards.



*Figure 47. Cluster analysis of statements regarding metadata standards*

Researchers seem to struggle with identifying the right metadata standards and the complexity of their studies which they feel they cannot express in a standardized form thus forming a cluster with close proximity. Metadata and their domains are not that close and quite distanced is the importance of metadata standards which we know from the descriptive statistics is very highly regarded.

The next question in the questionnaire for the second data collection was concerned with the types of data sources researchers use. Figure 49 shows the respective cluster analysis.

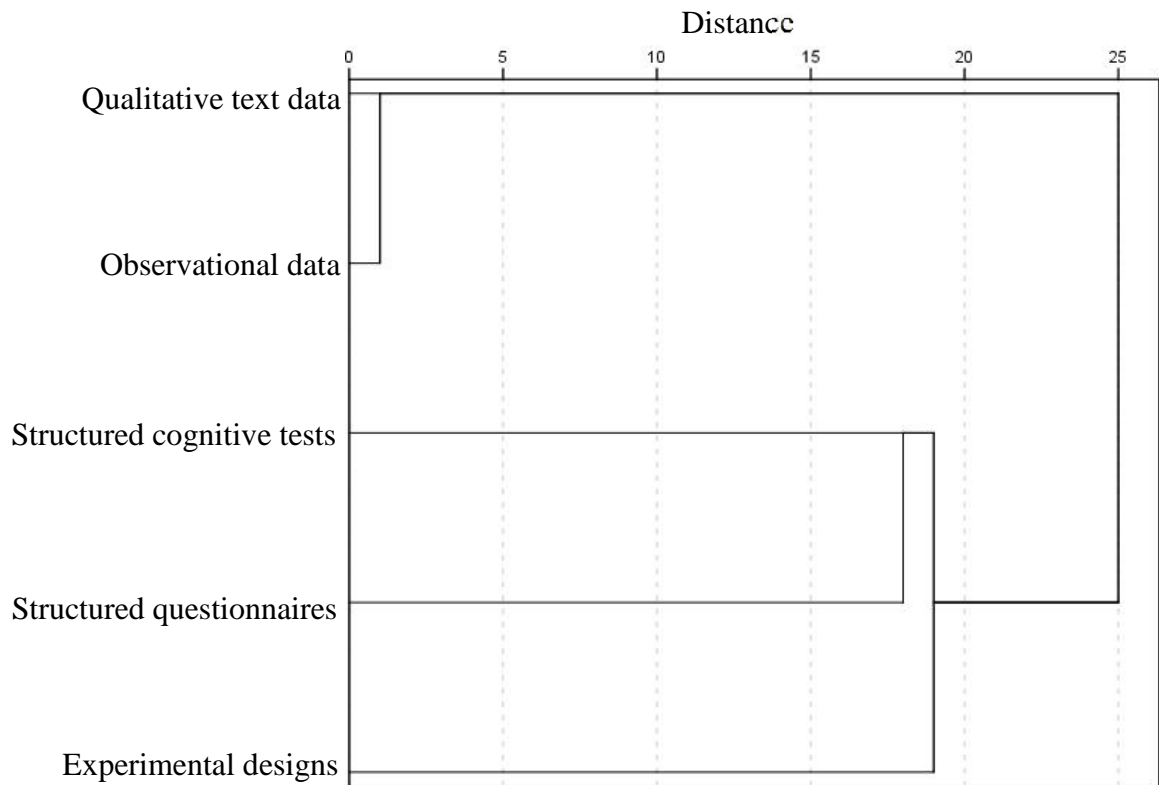


Figure 48. Cluster analysis of data sources

As can be seen in Figure 49 qualitative text data and observational data form a cluster with narrow proximity as both can be regarded to be used in a similar fashion e.g. for ethnographic research. From a data center perspective this also matches quite well as both can be considered to be stored as large files into a data repository. Structured cognitive tests and structured questionnaires form the next cluster which does not have the same narrow distance, but still combine. Experimental designs nevertheless can be seen in an outsider role as they form a cluster with the two other clusters but only with a large distance.

The next question tried to identify which metadata researchers provide with their datasets. Here the cluster analysis to find out which of the different categories form clusters.

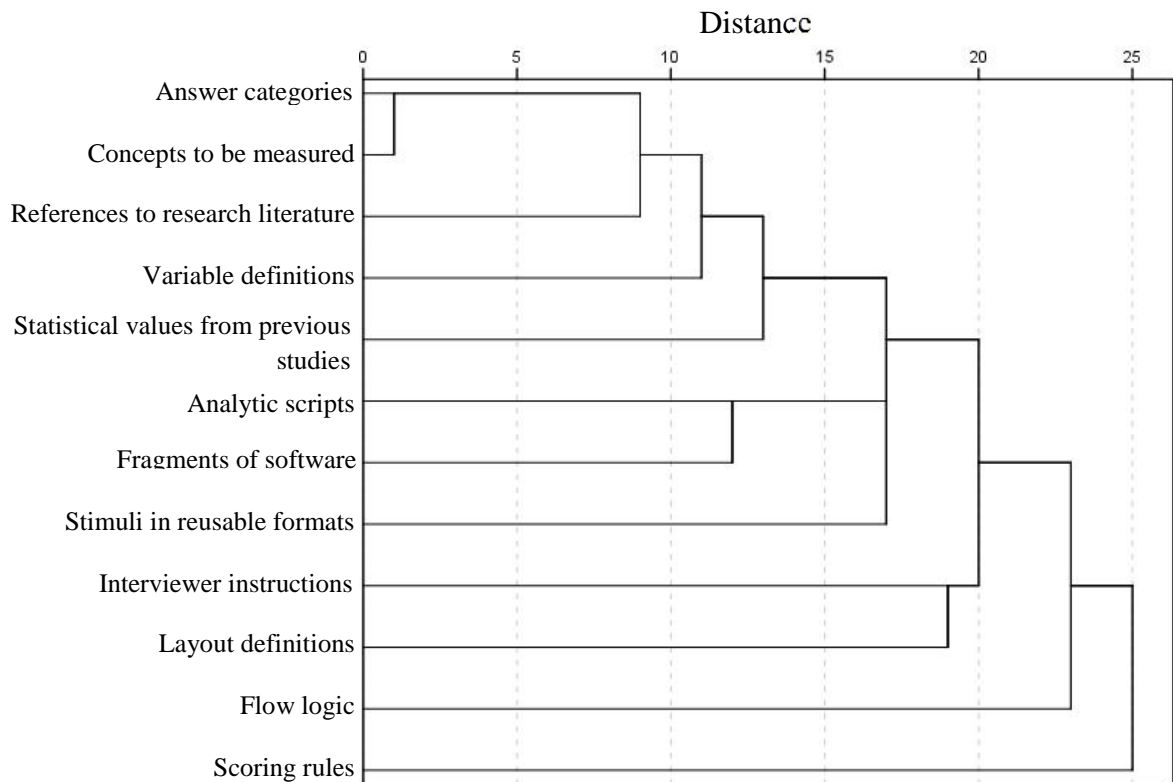


Figure 49. Cluster analysis of provided different metadata

Figure 50 almost shows at the top a stair like combination of clusters. The first cluster is formed between answer categories and concepts which are also quite prominent in the descriptive statistics. This cluster forms the next cluster with research literature, this cluster again combines with variable definitions to finally reach statistical values from previous studies. It seems to be a clustering from the common metadata to the less common metadata. Analytic scripts and fragments of software also form the next cluster but with a much higher distance. Flow logic and scoring rules appear to be the distanced categories as they form connections only at a very high distance.

The next question asked the researchers which data products they use for their research. The results from the cluster analysis are shown in Figure 51.



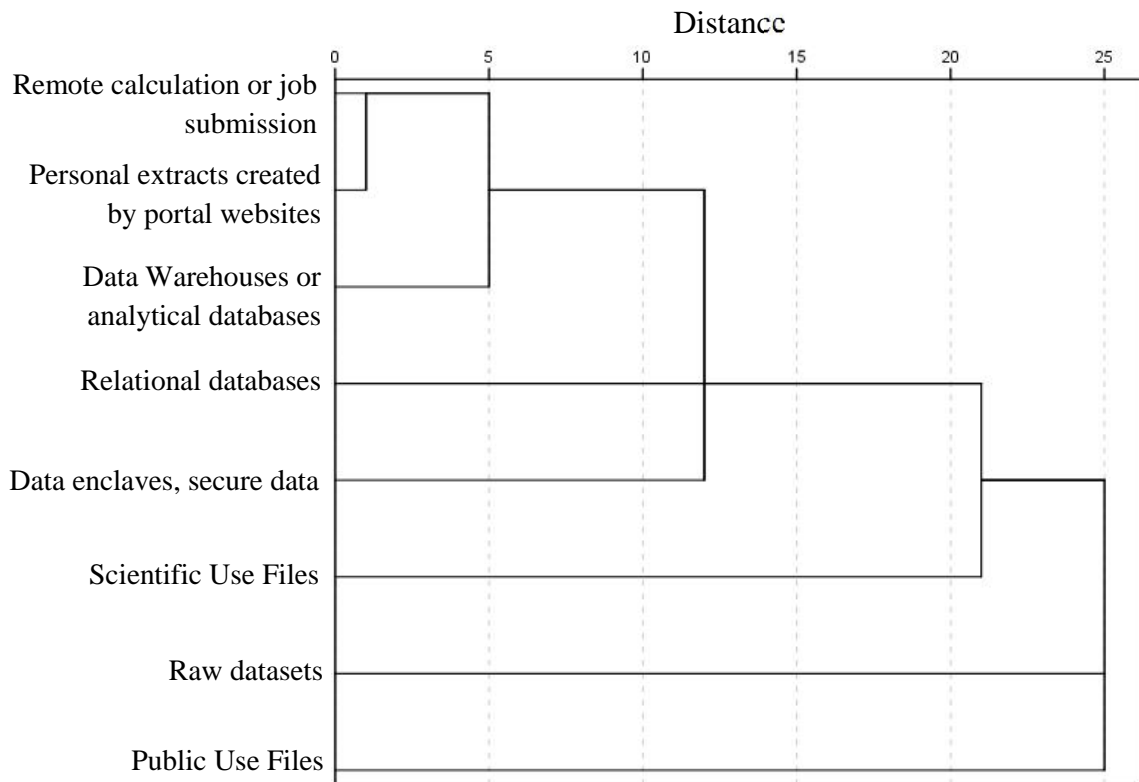


Figure 50. Cluster analysis of data products

Not very surprisingly the more modern techniques like remote calculation, personal extracts, data warehouse and data enclaves form clusters with low distances. Nevertheless the descriptive statistics in the previous show these data access methods to be very uncommon across different groups. On the other hand raw datasets and Public Use Files (PUFs) do not form clusters and are quite high in the overall ranking. It seems the more advanced techniques are used in combination by users with a higher affinity towards IT while the majority uses a very file-based approach. The result in general is quite surprising as Public Use Files do not allow for deeper data analytics due to their high amount of anonymization. Raw datasets on the other hand are at least in large-scale study not available for secondary analysis. This means normally these raw data files are the ones the researcher might have gotten by own data collection. Raw data files which were processed by a data collection agency are normally cleaned due to internal data quality regulations and thus cannot be considered “raw”.

The last question in the questionnaire was only shown to researchers who admitted they do not share their data and tries to find out the reason behind it. Figure 52 shows the cluster analysis of this question in the resulting graph.

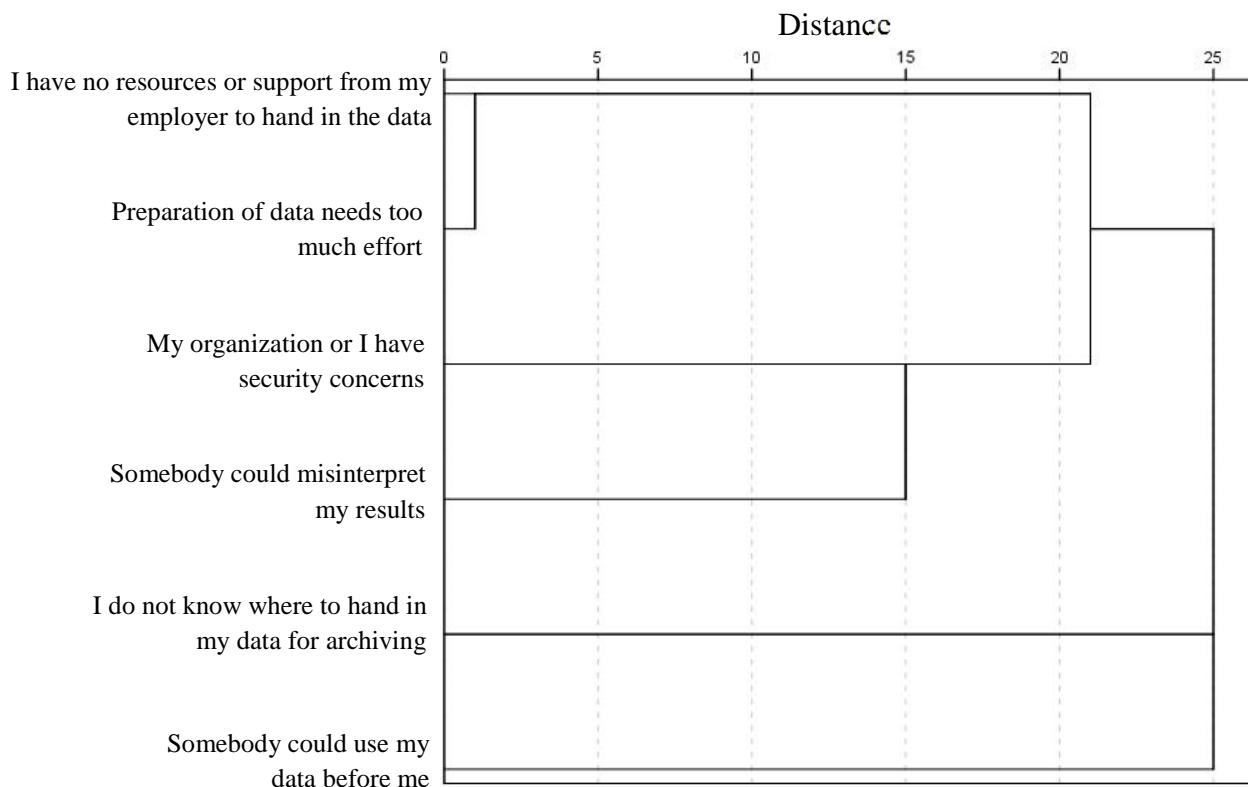


Figure 51. Cluster analysis of the reasons not to share data

Figure 49 shows with the very first cluster of low proximity a resource problem. No resources from the employer combines very well with the notion the preparation of data needs to much effort. This is not really surprising. The next cluster is formed between security concerns and misinterpretation of results. Not very deeply connected are the two statements about not knowing the location for handing in the data and the fear of somebody using the data before the original researcher.

Overall the cluster analysis strengthens the findings which were already visible in the descriptive statistics with displaying some of the underlying connections. To elaborate on more statistically advanced methodology as a final step the dataset was subjected to data mining procedures.

## 5.5 Data Mining

To better understand the results shown by the cluster analysis the idea was to perform a deeper assessment e.g. with the help of advanced data mining techniques. In late 2015 after the analysis of the second data collection and the combination of the datasets the author experimented with the dataset from both data collections e.g. by applying Structural Equation Modelling (SEM) (e.g. Ullman & Bentler, 2003) to it which did not provide interesting results as the data derived from the questionnaire did not show any latent variables except for measuring opinion. Another candidate for a more explorative approach could have been Educational Data Mining, but for applying techniques from this field a proper audit trail (e.g. see the PIAAC TAO BLACK and Itembuilder results in Chapter 1) is necessary. Audit trail means all interactions like clicking buttons or mouse movements are recorded as well as changing of answers including a timestamp in milliseconds. Limesurvey as a questionnaire system does not offer

this high level of granularity as in surveys the timing is not of the essence in an interview. It only records the total interview time and answer time for individual items. This is also not uncommon for educational studies. In PIAAC the interview system only records the individual time of items to detect cheating by the interviewer (i.e. faking interviews) not for data analysis purposes (see Zabal et. al. 2012). Real exact time measurement only starts in the cognitive assessment. This is also the reason why the technical standards and guidelines of this study allow a break or coming back on another day during the Background Questionnaire, but not during the Computer-Based Assessment. Timing of interactions simply does not matter during the interview. This is also the reason why most web survey systems like Limesurvey do not possess this functionality.

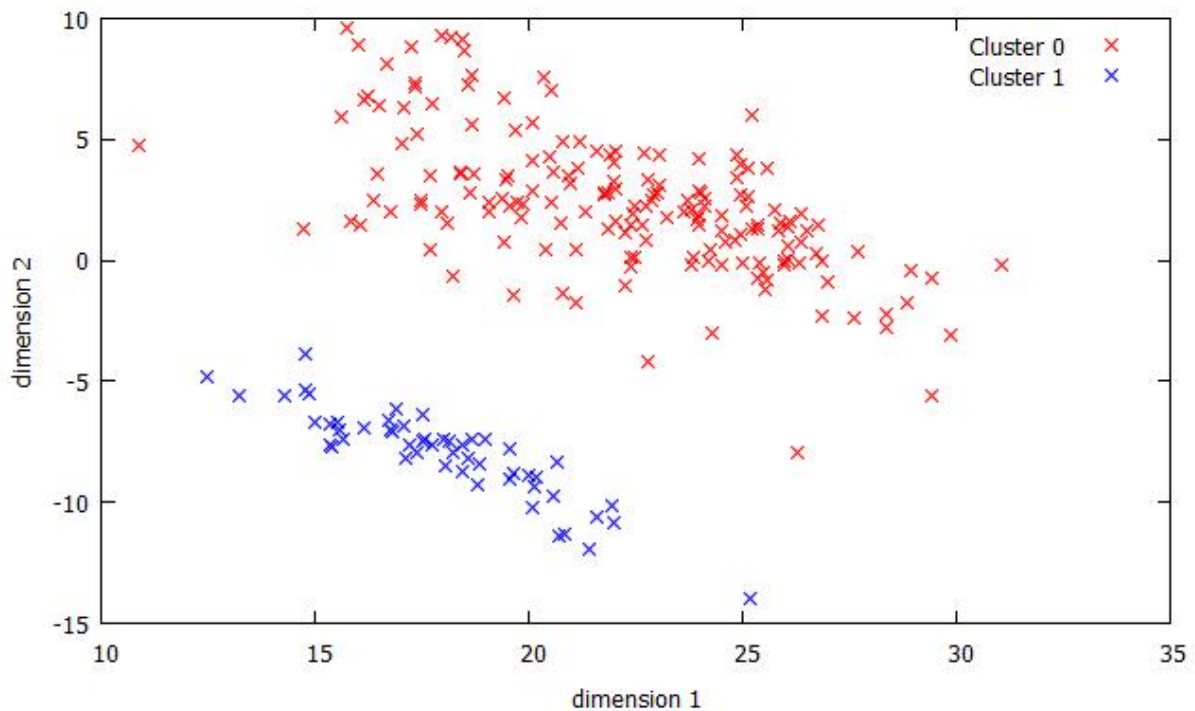
Nevertheless it is possible to apply data mining directly to the dataset in an exploratory approach to detect more groups or clusters than the ones based on categories. The basic technique is to combine all participants with all possible variables as dimensions and start reduction processes so clusters might show up. In this case the dimensions of the data were set by manually selecting 46 variables as the required memory and time for processing the data increase exponentially with the growth of the number of dimensions. As described before the database had instances with missing values and unfilled variables. Therefore instances with more than 20 undefined features had to be removed from the dataset as the processing steps would be meaningless (too many zeros). After this filtering step the sample size was reduced to 224 participants in the database for further analysis. The remaining missing values in the dataset were as hinted at before replaced by zeros, since this value has no effect on the result of the matrix-vector multiplication which is in the upcoming dimension reduction step. Instead of clustering the 46-dimensional data and analyzing the result of the method (e.g. measuring the distances between the cluster centroids or the compactness of the clusters) the dimension of the data was reduced to a two-dimensional plot as human understanding cannot grasp the complexity of a 46-dimensional hypercube (Bellman defined this phenomenon as the “curse of dimensionality” – Bellman, 1956). This allows to plot the instances to a plane and try to identify patterns. This plot can be seen in Figure 50 and it indicates there are two clusters in the dataset. The technique used to reduce the dimensions is called Singular Value Decomposition (SVD) (Golub & Reinsch, 1970) which is very commonly used in the area of machine learning. It searches for the bases of a subspace which maximizes the variance in the dataset.

Another option to plane the dataset could have been the Principal Component Analysis (PCA) method (e.g. Jolliffe, 2002) which is one of the most commonly used dimensionality reduction technique. It has the same basics as Singular Value Decomposition). The difference is PCA uses the eigenvectors calculated on the covariance matrix of the data and SVD uses the singular values which are the left and right eigenvectors of the dataset. The advantages of the feature extraction methods against the selection ones are the following:

- Preserves the relative distances among the data elements
- All of the resulting dimension is a linear combination of the original feature space.
- More information can be preserved by transforming the data into the two dimensional space instead of just select the two best features in the original space.

To analyze the datasets the following vectors were used. Let be  $A \in \mathbb{R}^{m \times n}$  the matrix which has the data samples row-by-row, where  $m$  is the number of instances and  $n$  is the number of features. The SVD method finds the decomposition  $A = U\Sigma V^T$ , where matrices  $U \in \mathbb{R}^{m \times m}$ ,  $V \in \mathbb{R}^{n \times n}$  have the left and right singular vectors, the matrix  $\Sigma \in \mathbb{R}^{m \times n}$  contains the singular values of the matrix  $A$ , respectively. To transform the matrix and the data to the two-dimensional space the multiplication  $\hat{A} = AV_2$  has to be performed. As a result  $V_2 \in \mathbb{R}^{n \times 2}$  is the matrix which has these two columns of the matrix  $V$  corresponding to the maximal singular values. The  $F$ -norms of the matrices  $\|A\|_F = 353.5166$  and  $\|\hat{A}\|_F = 330.6563$  mean only 6.5% of the information were lost after the dimension reduction.

Using these methods creates the following scatter plot of the planed clusters shown in Figure 50 including the two clusters marked in red (0) and blue (1) for convenience.



*Note: Dimensions 1/2 (x/y axis) correspond to the direction of the first/second singular vector, which means these directions have the largest variance.*

*Figure 52. Scatter plot of planed clusters*

These two clusters become visible by slicing the hypercube into different two-dimensional planes. Nevertheless what do they mean and where do they differ? As described before they do not base on human defined categories like academic status or research discipline which were used in the previous sections. Looking at the results it shows the two clusters differ in the following variables according to Table 59:

Table 59. Variables with significant differences between Cluster 0 and Cluster 1

	Cluster	N	Mean	Std. Deviation	Std. Error Mean
MS01_6	1	53	3.87	.341	.047
	0	171	3.55	.882	.068
MS01_7	1	53	3.98	.137	.019
	0	171	3.69	.697	.053
MS01_10	1	53	3.79	.454	.062
	0	171	3.53	.842	.064
MS01_11	1	53	3.83	.509	.07
	0	171	3.60	.793	.061
DA01_4	1	53	3.13	1.287	.177
	0	171	3.54	1.233	.094
DC14_1	1	53	0	0	0
	0	171	3.32	1.391	.106
DC14_2	1	53	.00	0	0
	0	171	3.20	1.324	.101
DC14_3	1	53	0	0	0
	0	171	3.18	1.4	.107
DC14_4	1	53	0	0	0
	0	171	4.43	1.137	.087
DC14_5	1	53	0	0	0
	0	171	4.16	1.271	.097
DC14_6	1	53	.00	0	0
	0	171	4.58	1.028	.079
DC14_7	1	53	0	0	0
	0	171	4.87	.945	.072
DC14_8	1	53	0	0	0
	0	171	4.68	.962	.074

The results are a bit surprising as they differ in relatively uncommon metadata standards mainly in the area of qualitative research (METS, QueDex, SDMX and TEI), use of experimental designs, use of fragments of software and virtually all of the different kinds of data access from Public Use Files to Data Enclaves. If human categories should be used the representation could be interpreted as clusters of more qualitative and experimental oriented scientists vs. empirically driven researchers in social sciences and educational sciences.

The even more surprising result is the differences between Cluster 0 and Cluster 1 in regards to data access. None of them (with n=53) uses any kind of data access to secondary research data. As this represents the qualitative and experimental oriented persons which fall out of scope of this dissertation the topic will not be discussed further, but it certainly has a lot of potential to be researched in upcoming papers.

*“A conclusion is the place where you get tired of thinking.”*  
(Arthur Bloch)

## **6. Conclusion**

The last chapter showed the results from the survey in detail, but what does this mean in operational terms? Before conducting the data collection research questions and hypotheses were formulated which can now be compared with the data. Basing on those results it might be possible to build new research processes with embedded support for data management or better software. This will be shown in some models. At the end of the chapter there will be an outlook into the future how data management for computer-based assessment could be improved.

### **6.1 Results of the analysis**

#### **6.1.1 Hypotheses revisited**

In this section the hypotheses formulated in Section 1.3.7 will be revisited and compared with the data if they were correct or not.

- *We expect more familiarity with data management and metadata standards from social scientist or researchers working in both fields than from pure educational scientist.*

This hypothesis proved to be true. Social scientist know more metadata standards than their counterparts from the educational sciences. Nevertheless the distinction might not be that clear as the analysis show a big overlap in the usage of cognitive instruments by social scientists and questionnaires by educational sciences. The categories in general might not be that clear.

- *We expect a general positive attitude towards metadata management as researchers understand the advantages, but a mediocre or even negative attitude towards current standards and their implementation.*

This hypothesis proved to be true. The importance of metadata standards was evaluated highly by all groups regardless of discipline or academic status. On the other hand most metadata standards are unknown to the groups and there are a lot of missing elements which were identified by the participants. Especially in the handling of metadata for log file analysis there is confusion about standards and formats. More work in specification is therefore necessary.

- *We expect most metadata standards will only be known by name or not at all by most researchers. There will not be many experts.*

This hypothesis proved to be true as pointed out in the first bullet point. Metadata standards are for researchers more like an opportunity than a reality. They only know very few standards in detail and the software products they are using are exactly the ones which do not support anything except their own propriety formats.

- *We expect significant interest of educational researchers in secondary analysis of datasets.*

This hypothesis proved to be true though there is no significant difference between the educational researchers and the other groups. All the researchers in the study were interested in the secondary analysis of datasets and a large portion of them are also actively doing this. It can be expected from the other values this trend will continue and get even stronger. This is also in line with Corti et. al. (2014) or Vartanian (2011).

- *We expect younger researchers to have a more positive attitude towards sharing of their research data than more senior researchers.*

This hypothesis proved to be false. There was no significant difference between academic status or years in research and the willingness to share data. Though researchers are seeing the benefits of metadata standards and do secondary analysis the willingness to share the own data especially with persons outside the own organization is rather low.

- *We expect the average time for data management and documentation in studies and surveys not to be sufficient for generating high level metadata.*

This hypotheses proved to be true. The time specified for data management in an 18 month survey is about six weeks. This can be cross-referenced with the metadata elements data producers from both disciplines offer. In end data users can consistently expect codebooks with variable information and data files. Every further information is optional and it is not clear if it is produced by the original research team or the staff at research data centres performing the process of enhancing metadata.

## **6.1.2 Answering the research questions**

Here are the research questions formulated in Section 1.3.6 and the comparison with the respective data.

- *Which categories of different metadata standards do scientist know?*

The survey shows scientists neither in the social sciences nor the educational sciences have deeper knowledge about the metadata standards and their internal structure themselves. DDI and QTI are known by name and scientists are also aware of the advantages of metadata standards, but cannot really use them without the help of data management personnel. Unfortunately also the software basing on metadata standards is less used than software using propriety formats. So there is also not much indirect usage of metadata standards.

- *How can those metadata standards be integrated into their daily work?*

As the awareness about metadata standards is low especially in the educational sciences which is the main focus of this dissertation more impulses will have to come from research data centres towards researchers or IT personnel to change current procedures. This can be either by consulting researchers on data management procedures and helping them to embed them into their processes with the least impact on time. Furthermore also IT personnel can be trained from data management side to embed the creation of standardized metadata into software development so they become a by-product of the research process and therefore do not produce additional time constraints for researchers (ideally new software products should even save time for researchers). This topic will be further discussed in Section 6.2.2.

- *What are the requirements of educational scientists towards metadata standards?*

The requirements do not differ from those of social scientists as long as both are using questionnaires and for this purpose DDI seems to fulfil most demands.

Scientists from both disciplines would like to store:

- Answer categories
- Concepts to be measured
- Scoring rules
- Variable definitions
- Flow logic
- Interviewer instructions

The situation changes when the focus is moved towards computer-based testing with cognitive instruments. The requirements on the survey lifecycle are very similar and have extended demands in regards to instruments (e.g. scoring, stimuli, complexity), but the metadata standards like QTI or APIP are much inferior. Currently there is no metadata standard for the educational sciences which fulfils those demands.

- *What is missing to provide a high degree of re-usability of items and instruments for computer-based assessment?*

Current metadata standards for computer-based testing in the educational sciences like QTI only focus on the design and delivery of simple item types. The processes of “Survey Design”, “Data Processing”, “Distribution”, “Discovery”, “Analysis” and “Re-Purposing” in comparison to DDI Lifecycle are missing. QTI can store the design of a simple item without layout information for re-use.

- *What is missing for educational researchers who want to do a secondary analysis of the data?*

Data producers offer a least a basic metadata set along with the data and the data users seem to be satisfied with this. For pure analysis this might be good enough. Nevertheless this can only be accounted for internal data as most researchers do not hand in their data to third parties. This means in many cases the data is not available externally though the scientists would be willing to share the data. A result which is also in line with Tenopir et. al. (2011).

- *How could a model for storing a computer-based assessment including all metadata and paradata in the long term look like?*

Based on the survey results it looked similar to the GLBPM or DDI Lifecycle model, but with an additional focus on scoring, stimuli and complexity in the area of instrument development and delivery. This topic will be further discussed in Section 6.1.3.

- *Do researchers in the educational sciences need a work-around in the meantime?*

Yes, as the current metadata standards do not fulfil the demands completely a mix of metadata standards from different domains plus manual documentation is needed even for simple items and instruments. This topic will be further discussed in Section 6.2.1.

- *How can especially complex item types like simulations in complex problem solving (e.g. MicroFIN, see Greif and Funke 2010) be modelled in a meaningful way?*

With the current metadata standards complex problem solving cannot be modelled as neither QTI nor APIP deliver answer modes, scoring mechanisms or support for graphical elements which can express



this level of complexity. For those items a completely new development in form of an abstraction layer or even standardized modelling language is necessary. Proper data modelling as expressed e.g. in Simsion & Witt (2005) cannot be fulfilled.

- *How can the results be stored for long time preservation or archiving?*

The results from studies are the smallest problem encountered in the survey. The formats for statistical packages are quite stable over the years and known to the researchers. In a worst case scenario the tabular files can be converted into CSV files – a standard which is unchanged since the 1970s and supported as the lowest demeanour also in the coming decades. Nevertheless there might be data loss involved e.g. if the original data was stored in a relational database or analytical database with additional information and in a more meaningful way. The answer therefore is implementing data management, data archiving or data stewardship processes which means dedicating personnel which preservation and understanding of these data. An example is given in Section 6.2.2 and can be seen as progress towards the vision of a digital library (Altman et. al., 2001) as recommended in Germany by the Scientific Council (Wissenschaftsrat, 2012).

### **6.1.3 Impact of the results on current metadata standards**

As the research questions prove currently handling metadata for computer-based surveys and assessment is a problem as the metadata standards do not cover all the demands from a researcher's perspective. A dissertation should provide mainly new insights but it is also not a mistake to provide answers towards the applicability of those results. Therefore this section will contain a discussion how the situation can be improved using current standards. At the end there will also be a topic about creating a completely new standard.

#### **6.1.3.1 Extending Data Documentation Initiative (DDI) with educational content**

In the survey DDI proved to be the most prominent of the metadata standards and is used by social scientists and educational scientists as well. DDI covers also most of the lifecycle tasks specified in the GLBPM, GSBPM and GSIM. Furthermore DDI Lifecycle is currently undergoing a change from a XML scheme towards an ontology using multiple representations (version shift from DDI Lifecycle 3.2 to DDI Lifecycle 4.0 in the DDI Moving Forward project). The new DDI is supposed to be modular so it can be extended with other domains outside of the core set (e.g. healthcare metadata). With this modular structure it would be possible to add an extension especially for educational content.

To illustrate this point here a Figure 54 shows a model of the upcoming modular DDI.

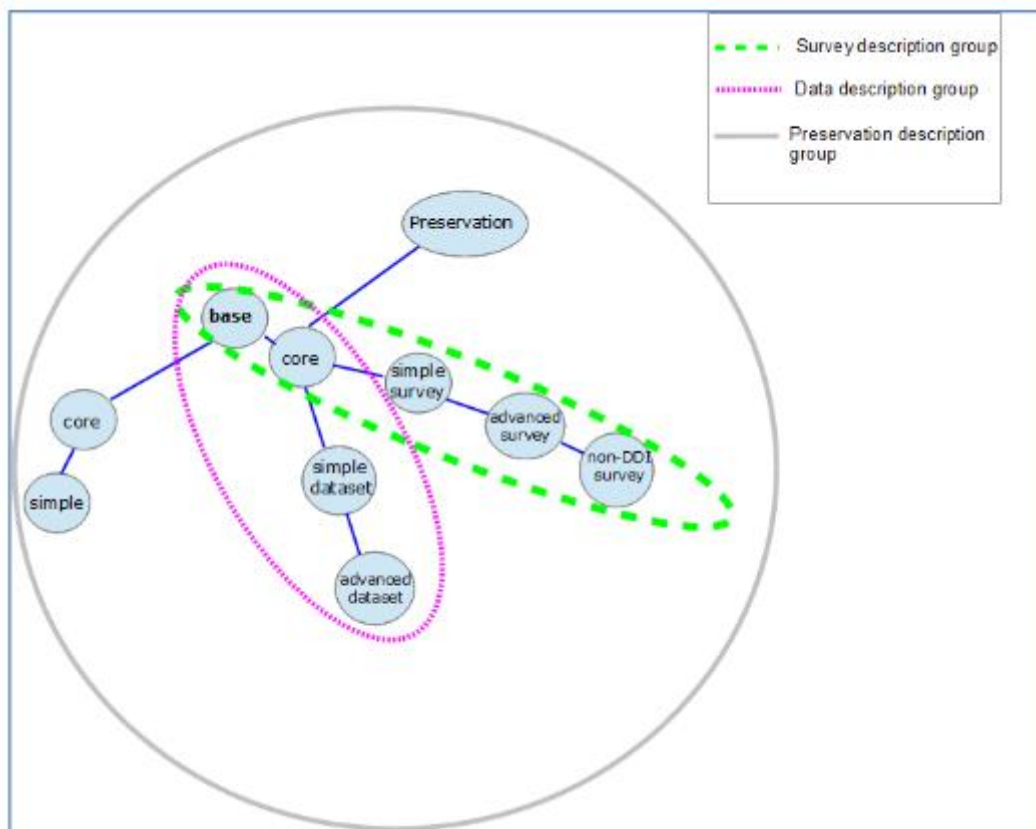


Figure 53. The modularized DDI model

Source: DDI, 2012 p. 5.

The module „educational survey“ could be a branch of „simple survey“ or „advanced survey“ depending how these packages will look once they have been fully specified by the DDI Alliance. The current development process regarding the core packages for DDI4 is running since 2012 and a first release candidate is scheduled for 2015, but the actual fully released and tested version 4.0 of DDI might take more time than that.

An involvement for educational researchers in DDI for creating an own module is a useful enterprise for several reasons:

- Educational researchers also heavily use questionnaires and can therefore get the advantages of handling both types of instruments within the same standard
- DDI 3.2 already contains some extensions derived from the educational sciences (e.g. stimuli, item batteries) as some institutions from this sector already were part of the development process in the past (e.g. DIPF)
- DDI contains a lot of support for statistical packages and publications which were also topics in high value by educational scientists

Nevertheless if educational scientists start their involvement now it will take some years until the results will be visible in the metadata standard itself and even more time until they will influence software products. Counting on DDI as an extended educational standard is therefore promising but only a long-term option.

Furthermore it should be mentioned the German Institute for Educational Progress (IQB) started this discussion of embedding educational content into DDI already once in form of the EduDDI initiative (Mechtel, 2009), but this did not lead to a lasting success at this time. This can be explained by the very early version of DDI Lifecycle (version 3.0 at this time) and the lack of awareness in the educational sector. Maybe the process of developments around DDI4 prove a more successful ground.

### **6.1.3.2 Extending Questionnaire and Test Interoperability (QTI) with content from lifecycle models**

The other option to create a better metadata standard for computer-based surveys and assessments for the educational sciences would be to extend the most prominent standard there. Similar to DDI also QTI is currently under further development due to the merging process with APIP. Nevertheless this process might be more difficult than extending DDI as there are some challenges:

- QTI does not contain any objects outside of the instrument design domain, so processes regarding survey design, data collection, dissemination and publication have to be added. This is a huge endeavour as several hundreds of object classes and types have to be specified.
- The development cycles in QTI are long. The further development from QTI 2.0 to 2.1 took ten years from 2003 to 2013. This process might speed up due to the APIP involvement.
- QTI is not well-known even among educational researchers as the survey has shown.

In summary hoping for an improvement of QTI into a direction of more lifecycle-oriented workflows might take more time than the developments surrounding the DDI Moving Forward project. It might nevertheless be an idea for educational researchers interested in metadata management to join the community at IMS Global surrounding QTI and influence those developments.

### **6.1.3.3 Developing a fully new metadata standard for computer-based assessment**

Another idea might be a fully new creation of a metadata standard especially for computer-based assessment specifying missing parts like complex item types, log file standards or scoring rules. Nevertheless the organizational and political dimensions of such a project are huge. The development cycles for versions of Dublin Core, DDI or QTI take years with a multitude of organizations being present in these alliances. As an example the DDI Alliance currently consists of 39 large member organizations. A list can be found in the appendix.

Building up structures like these with an appropriate number of members who then also commit to use the standard in their respective organizations takes very long. In the case of DDI it started beginning of the 1990s with the first version. Specifying a completely new standard which also has to compete with existing standards might therefore not be the ideal way to go.

## **6.2 Development of a model for software development based on the results**

Unfortunately the further development of standards will not provide a quick solution for the lack of metadata management in the educational sciences. Therefore the question is how these problems can be handled in the meantime while the standards are progressing. The section will therefore discuss some scenarios to improve the current situation of a gap between demands from researchers, processes available in data management and software developed by a multitude of vendors.

### **6.2.1 Combining metadata standards for educational research as an interim solution**

As the literature review has shown the current metadata standards are not sufficient to represent content from computer-based assessment. Furthermore until DDI, QTI or other standards have filled the gaps and white spaces more time will be lost. Interim solutions are therefore necessary to bridge the development time for the metadata standards.

To show a possible matrix of metadata standards the following two axis have been built. On the one hand processes from the Generic Longitudinal Business Process Model (GLBPM) have been taken as they represent the typical workflows which are also common to surveys from the educational sector. On the other axis the following areas according to the previous survey have been identified:

- Survey information (describing the research project from a scientist's perspective)
- Education-systematic information (metadata information necessary for educational institutions like schools and universities)
- Questionnaires (as they are also very common with educational researchers)
- Simple item types (e.g. multiple choice items)
- Complex item types (e.g. simulations)

The preferred standards to be used in this system are DDI Lifecycle 3.2 and QTI v2.1 as they were the most familiar to educational researchers in the survey though in general the knowledge about metadata standards themselves can be considered to be low.

DDI Lifecycle 3.2 provides another advantage which can be used for combining it with other standards. Elements which do not exist within DDI but are necessary to represent content for an agency can be defined as a user-attribute pair (DDI, 2009) for a custom extension. It is therefore possible to embed another standard into DDI without violations. Nevertheless these user-attribute pairs are only understood by the agency which originally created them or other agencies using the same layout. A completely unaware DDI instance will not be able to parse user-attribute pairs defined by another agency without implementing the same layout.

Figure 55 shows the model of organizing current metadata standards into the process model.

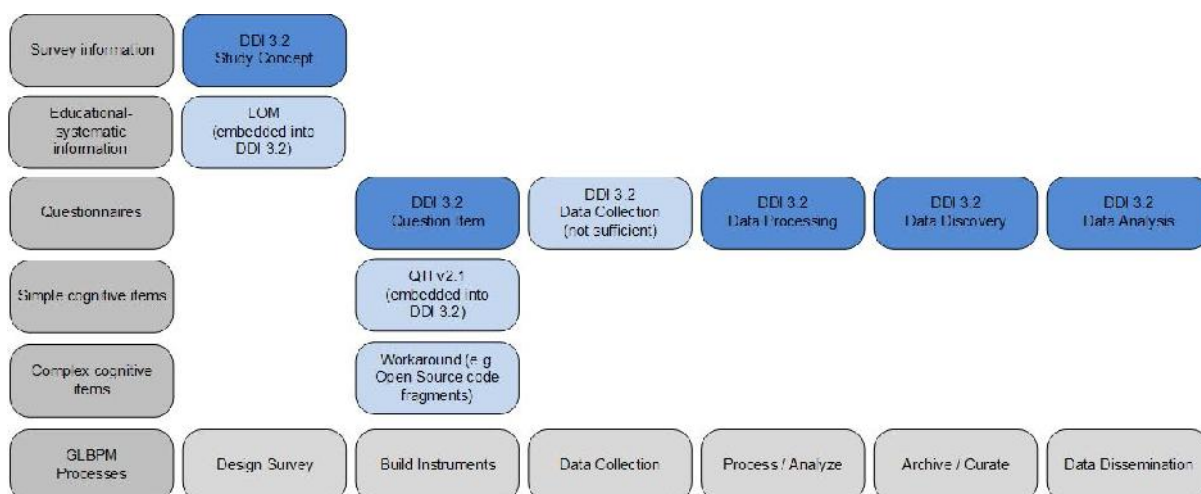


Figure 54. Hybrid metadata model for computer-based assessment

As can be seen the majority of processes can be handled within DDI 3.2 while some extension can be done by using other standards:

- Educational-systematic information can be handled by using the Learning Object Model though it has to be observed LOM is actually adapted to the educational system of different countries (e.g. UK-LOM, LOM-CH). This means the correct “flavour” of LOM has to be chosen. Furthermore it would be possible to embed LOM into a DDI 3.2 structure using user-attribute pairs.
- Simple cognitive items can be handled by using QTI v2.1 as a standard and similar to the procedure for LOM it can be represented by user-attribute pairs within DDI 3.2 enabling the integration from one standard into the other.
- There is no good solution for complex items as none of the standards is able to handle them. A workaround could be to store additional information about the item (e.g. screenshots, videos, graphical elements, documentation about the workflow) in archival formats and try to preserve the original software code by using Open Source repositories thus preserving as much as possible. Hopefully the challenge of complex cognitive items will be addressed by a metadata standard for education in the future.
- Unfortunately data collection procedures and the paradata derived from it are also not handled very well within DDI 3.2. It is therefore necessary to document information like sampling, disposition codes, interviewer information or field reporting into archival formats which can easily be imported and exported between databases. A recommendation would be specifying a propriety XML-based scheme including a description of the layout as long as metadata standards do not offer any support for paradata.

Furthermore there are additional challenges derived from the details of the survey.

- There is no format to handle scoring rules. QTI offers a limited set of rule definitions but they are not sufficient for complex workflows. If there are rules which cannot be expressed in QTI they have to be documented externally used procedures like described above.

- The same problem exists for log file metadata. There is no unified format for these information. It is therefore necessary to document the layout of the log files so they can be analysed in the future.

A metadata and data repository following these rules will be able to store computer-based instruments and the datasets derived from it in a meaningful way. Of course the information in this paragraph is not detailed to develop such a system from scratch. This would also not be possible as the content has to be adapted to the requirements of the researchers within the facility. It can nevertheless serve as a blue print for first design considerations which then have to be extended by an internal requirement analysis between the researchers, data management personnel and computer-scientists if the software is developed in-house.

### **6.2.2 Embedding data management personnel into the software development cycle**

As the last paragraph of the previous sections describes embedding data management into an organization involves three players. From the survey we know neither educational researchers nor computer-scientists on their own have sufficient knowledge about data management processes to properly embed them into their research or software development cycles. Thus data management personnel has to be embedded into those processes.

This can be done by the following ways:

- Data Managers can act as advisors to researchers for choosing the proper software products, metadata standards and processes to support their survey work
- Data Managers can act as advisors to software developers who develop survey tools, assessment software, data collection software or repository software to support processes from the very beginning
- All three groups can be involved in the development of new processes and new standards at certification bodies or alliances with their expertise thus creating an data management supported research process

For this vision to come true it is needed researchers see the benefit of sharing their data and enriching them with metadata in the first place. For the researchers stating in the survey they are willing to support these processes this needs convincing arguments and a change of paradigm. In a world where financial resources for conducting surveys gets less and less there might be no other option then strengthening the usage of high quality datasets from research data centres with additional research questions and analysis.

### 6.3 An outlook into the future

To close the dissertation here are two views at the subject matter and a projection into the future. First a general view how data management will influence computer-based assessment in the future and then a more specific look at current issues at one of the home organizations of the author.

#### 6.3.1 Data management and computer-based assessment

A clear result of the dissertation is the necessity of data management for computer-based assessment. The question should therefore not be if data management has to be implemented but rather how this task can be done. Unfortunately it can be stated educational scientists are in the normal day-to-day work not aware of the problems the lack of data management produces surrounding their items, instruments and datasets. Given the current trend towards computer-based assessment the problem is very likely to increase over time. This means in a worst case scenario in 15-20 years there will be a dramatic loss of metadata and data as nobody is able to reproduce the items and results from studies which were not big enough to have own data management procedures or were considered to be sufficiently important for a research data centre to actively inquire for it.

It can only be hoped this dissertation or similar movements from data managers will result in software which supports data management procedures as a by-product of the item development and data collection process as it can be considered unlikely stakeholders will start extra funding for more advanced data management procedures. Nevertheless it can be expected stakeholders will pressure scientists more via funding regulations to embed data management procedures, e.g. formulating the necessity to have a data management plan for a survey. Currently more and more funding bodies are moving towards this direction. Unfortunately this does not necessarily lead to a certain quality as the data management plan might not be sufficient e.g. if it was designed without the help or data management personnel. The author has seen in his professional careers data management plans like “*at the end of the study we hand over the codebooks and the datasets to a research data centre*”. Though it can be stated positively at least the survey is willing to share the data externally we have no idea which quality the data documentation will reach if the research data centre is not involved during the process. Actually a valid description of the current situation is metadata and data is handed over to RDCs at the end of the study on stakeholder pressure. The datasets are very often not documented sufficiently to be handled by the repository systems and portal solutions a RDC uses and also not prepared for long-term preservation. Very often a process which could be described as “metadata forensics” starts where a scientist not involved in the original survey tries to reconstruct information about it. This process is time-consuming and might provide errors as the work is not done by an insider but an external scientist who has to resort to interpretation.

Software and metadata standards have thus to become more user friendly and support the full survey process ideally in one software package which can handle the whole workflow or at least the interfaces have to be designed to support a seamless migration from tool to tool.

### **6.3.2 Recommendations for data management and metadata standards for the Institute for Educational Sciences at the University of Szeged**

At the very end here is a very specific section in regards to the Institute for Educational Sciences at the University of Szeged, so more or less a “closing section after the closing chapter”. The author is currently involved with two different organizations – the German Institute for International Educational Research (DIPF) and the University of Szeged. While DIPF is already running an own research data centre in form of the RDC “Forschungsdaten Bildung” at the information centre of education (ICE) department and therefore is already concerned with those issues the Institute for Educational Sciences at the University of Szeged does not have any data management processes in place. Therefore here are some recommendations how the results of this dissertation can be used in a pragmatic manner to build up such structures. These results might also be interesting for other institutions with similar goals.

Currently the Institute for Educational Sciences is involved in several international and national projects where a multitude of assessment instruments are created and used mainly in schools for data collection. The main project is a TAMOP-funded national diagnostic school monitoring program which started in 2008 and is projected to run until autumn of 2015. In this project thousands of items and datasets are created by using the in-house produced eDia testing system. Due its design the instruments and results this project creates are highly interesting for secondary research. Thus this dissertation could be seen from an internal perspective as a requirement analysis to start the development of those data management processes.

The following processes can be recommended:

- Awareness of data management processes by building up own personnel for these tasks

As we have seen data management does not come naturally to educational scientists and neither does it to computer scientist who are involved in creating software for instrument design and data collection processes. Therefore own data management personnel is needed for those tasks with a designated role. The process can start by assigning the data manager role to one of the senior scientists and let this person start with the creation of a concept for a research data facility or data archive. This concept can be used to apply for further funding from national or international sources.

- Use of metadata standards for own software products (e.g. eDia)

Currently eDia does not use any metadata standard. The new data manager can serve as an advisor to the software developers which standards could be implemented into eDia in the middle and long run. The model created in section 4.2.1 between DDI and QTI can serve as a basis, but usually it has to be adapted towards institutional restrictions by performing a requirement analysis. This requirement analysis might also include partner institutions of Szeged where instruments might be re-used like the University of Helsinki, the University of Heidelberg or the University of Luxembourg.

- Legal framework

If datasets from Szeged are available to external researchers the legal requirements in regards to Hungarian law have to be evaluated. This is especially important if they are supposed to be available also to international researcher while here another separation also has to be accounted for – EU and non-EU residents. The recommendation would be to hire a lawyer in this field to provide an expertise on this subject.



- Access methods

Should the data be available as Public Use Files, Scientific Use Files or will a data enclave be available? Does Szeged want to offer guest scientists seats for visiting researchers? These questions can only be answered properly in regards to funding as all of these mechanisms base on costs. Nevertheless the recommendation would be to consider those different methods from a pragmatic perspective (e.g. start with Scientific Use Files and plan beyond).

- Visibility

To enhance the attractiveness of the results from the Hungarian school monitoring the research data unit should be visible to the community.

This means the following tasks:

- Generating a web presence for this unit describing the services offered
- Documenting the instruments and data also in foreign languages
- Taking part in the international research data community (e.g. conferences, alliances)
- Formal certification or accreditation as a research data centre or data archive

In summary the instruments and data generated by a proper data management will provide the Institute for Educational Sciences at the University of Szeged a higher visibility in international research processes and serve as a basis for more collaboration. Nevertheless a starting point has to be set for this purpose.

If the institute successfully starts a research data centre this unit can also be used for neighbouring disciplines like the social sciences, psychology or economy as the datasets from a storage perspective are quite similar.

## References

- Allen, L., Brand, A., Scott, J., Altman, M., & Hlava, M. (2014). Credit where credit is due. *Nature*, 508(7496), 312-313. Retrieved August 19, 2016, from <http://www.nature.com/news/publishing-credit-where-credit-is-due-1.15033>.
- Altman, M., Andreev, L., Diggory, M., King, G., Sone, A., Verba, S., & Kiskis, D. L. (2001). A digital library for the dissemination and replication of quantitative social science research. The virtual data center. *Social Science Computer Review*, 19(4), 458-470.
- Amin, A., Barkow, I., Kramer, S., Schiller, D., & Williams, J. (2012). Representing and utilizing DDI in relational databases. *RatSWD Working Paper Series*, 191, 1-17. Retrieved August 19, 2016, from <http://dx.doi.org/10.2139/ssrn.2008184>
- Bagley, P. R. (1968). *Extension of programming language concepts*. Philadelphia: University City Science Center.
- Baldoni, M., Baroglio, C., Patti, V., & Torasso, L. (2004). Reasoning about learning object metadata for adapting SCORM courseware. *Engineering the Adaptive Web., CS-Report*, 04-18. Retrieved August 19, 2016, from <http://www.rewerse.net/publications/download/REWERSE-RP-2004-35.pdf>
- Baker, T. (2000). A grammar of Dublin Core. *D-lib magazine*, 6(10), 47-60.
- Bambey, D., & Rittberger, M. (2013). Das Forschungsdatenzentrum (FDZ) Bildung des DIPF. Qualitative Daten der empirischen Bildungsforschung im Kontext. In D. Huschka, H. Knoblauch, C. Oellers, & H. Solga (Eds.), *Forschungsinfrastrukturen für die qualitative Sozialforschung* (pp. 63-71). Berlin: Scivero-Verlag, S. Retrieved August 19, 2016, from [http://ratswd.de/dl/downloads/forschungsinfrastrukturen\\_qualitative\\_sozialforschung.pdf](http://ratswd.de/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf).
- Bambey, D., Reinhold, A., & Rittberger, M. (2012). Pädagogik und Erziehungswissenschaft. In H. Neuroth, S. Strathmann, A. Oßwald, J. Klump, & J. Ludwig, (Eds.), *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*. (pp. 111-135). Boizenburg: Hülsbusch, S.
- Barker, P. (2005). What is IEEE learning object metadata/ims learning resource metadata. *Cetis standards briefings series*. Retrieved August 19, 2016, from <http://publications.cetis.ac.uk/wp-content/uploads/2011/02/WhatIsIEEELOM.pdf>
- Barkow I., Block B., Greenfield J., Gregory A., Hebing M., Hoyle L., & Zenk-Möltgen W. (2012). Generic longitudinal business process model. *DDI Working Paper Series, Longitudinal Best Practices II*, 5, 1-26. Retrieved August 19, 2016, from <http://ddionrails.org/glbpm/GLBPM.pdf>
- Barkow, I., Leopold, D. S. T., Raab, D. S. M., Schiller, D., & Rittberger, M. (2011). RemoteNEPS: data dissemination in a collaborative workspace. *Zeitschrift für Erziehungswissenschaft*, 14(2), 315-325.
- Barkow, I. (2007). Statistische Analyse momentaner IT-Bildungsabschlüsse auf ihre Relevanz in der Wirtschaft.
- Barsky, E., Abel, J., Deardorff, A., Henderson, L., Mitchell, M., Soper, D., & Taylor, S. (2014). *Research Data Management DataGuide v2.5*. University of British Columbia, Vancouver. Retrieved August 19, 2016, from [http://researchdata.library.ubc.ca/files/2014/10/RDM\\_DataGuide\\_V02.5\\_20141209.pdf](http://researchdata.library.ubc.ca/files/2014/10/RDM_DataGuide_V02.5_20141209.pdf)
- Barz, H. (2011). Der PISA-Schock. Über die Zukunft von Bildung und Wissenschaft im Land der "Kulturnation". In G. Besier (Ed.), *20 Jahre neue Bundesrepublik. Kontinuitäten und Diskontinuitäten*. Göttingen: Vandenhoeck & Ruprecht.
- Bellman, R. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences*, 42(10), 767-769.
- Bender, S. & Heining, J. (2011). The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing. *IASSIST Quarterly*, 35, 10-16. Retrieved

- August 19, 2016, from <http://iassistdata.org/iq/research-data-centre-research-data-centre-approach-first-step-towards-decentralised-international>
- Berg, A. M., & Korcuska, M. (2009). *Sakai courseware management: The official guide*. Olton: Packt Publishing Ltd.
- Bethlehem, J., & Biffignandi, S. (2012). *Handbook of Web Surveys*. New Jersey: John Wiley & Sons.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5-31.
- Blank, G., & Rasmussen, K. B. (2004). The Data Documentation Initiative The Value and Significance of a Worldwide Standard. *Social Science Computer Review*, 22(3), 307-318.
- Blossfeld, H. P., von Maurice, J., & Schneider, T. (2011). The National educational panel study: need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft*, 14(2), 5-17.
- Bohl, O., Scheuhase, J., Sengler, R., & Winand, U. (2002). The sharable content object reference model (SCORM)-a critical review. In B. Werner (Ed.), *Proceedings International Conference on Computers in Education* (pp. 950-951). Los Alamitos: The Institute of Electrical and Electronics Engineers, Inc.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., & Walther, G. (2005). *IGLU - Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Bosch, T., Cyganiak, R., Wackerow, J., & Zapilko, B. (2012). Leveraging the DDI model for linked statistical data in the social, behavioural, and economic sciences. In *Proceedings of the 2012 International Conference on Dublin Core and Metadata Applications* (pp. 46-55). Dublin: Dublin Core Metadata Initiative. Retrieved August 19, 2016, from <http://dcpapers.dublincore.org/pubs/article/view/3654/1877>
- Borgman, C. L. (2012a). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63,(6) 1059-1078.
- Borgman, C. (2012b). Why Are the Attribution and Citation of Scientific Data Important? In P. E. Uhlir, (rapporteur), *For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Washington, DC: The National Academies Press. Retrieved August 19, 2016, from [http://nap.edu/catalog.php?record\\_id=13564](http://nap.edu/catalog.php?record_id=13564)
- Bandilla, W., Bosnjak, M., Mohler, P. P., & Lüttinger, P. (2000). Online-Surveys als Herausforderung für die Umfrageforschung. *Festschrift für Max Kaase*, 9-28.
- Bradburn, N., Horton, R., Lane, J., & Tilkin, M. (2006, March). *Developing a data enclave for sensitive microdata*. Paper presented at the NSF SBE/CISE workshop, Airlie House, Virginia. Retrieved August 19, 2016, from [https://www.academia.edu/17769494/Developing\\_a\\_Data\\_Enclave\\_for\\_Sensitive\\_Microdata](https://www.academia.edu/17769494/Developing_a_Data_Enclave_for_Sensitive_Microdata)
- Brandt, M., & Schiller, D. (2013, October). *Safe centre network – need for a safe centre to enrich European research*. Paper presented at the Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa, Canada. Retrieved August 19, 2016, from [http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic\\_3\\_Brandt\\_Schiller.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_3_Brandt_Schiller.pdf)
- Brase, J. (2010). DataCite - A global registration agency for research data. *RatSWD Working Paper Series*. 149, 1-6. Retrieved August 19, 2016, from <http://dx.doi.org/10.2139/ssrn.1639998>
- Bryla, B., & Loney, K. (2008). *Oracle database 11g DBA handbook*. United States: Mcgraw-Hill.

- Büttner, S., Hobohm, H-C., & Müller, L. (2011). Research Data Management. In B. Stephan, H-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement*. (pp. 13-24). Bad Honnef: Bock und Herchen Verlag.
- Capadisli, S., Auer, S., & Ngonga Ngomo, A. C. (2015). Linked SDMX data. *Semantic Web*, 6(2), 105-112.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, (2<sup>nd</sup> Edition). Hillsdale, N.J. : Lawrence Erlbaum Associates.
- Corti, L. (2006). Editorial. *Methodological Innovations Online*, 1(2), 1-9. Retrieved August 19, 2016, from <http://www.esds.ac.uk/news/publications/MIOCortiEditorial-pp1-9.pdf>
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.
- Corti, L., & Thompson, P. (2004). Secondary analysis of archived data. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative Research Practice: Concise Paperback Edition*, (pp. 297-313). London: SAGE Publications Ltd.
- Couper, M. P. (2000). Review: Web surveys: A review of issues and approaches. *The Public Opinion Quarterly*, 64(4), 464-494.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012). Technological issues of computer-based assessment of 21st century skills. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 143-230). New York: Springer.
- Csapó, B. (2011, April). *Developing an online assessment system: Results of the first phase and future plans*. Paper presented at the Third Szeged Workshop on Educational Evaluation (SWEE), Szeged. Abstract retrieved August 19, 2016, from <http://www.edu.u-szeged.hu/swee/>
- Csapó, B., L. rincz, A., & Molnár, G. (2012). Innovative Assessment Technologies in Educational Games Designed for Young Students. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 235-254). New York: Springer.
- Csapó, B., Molnár, Gy., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, 106(2), 639-650.
- DataCite (2014). *DataCite Metadata Scheme V3.1*. Retrieved August 19, 2016, from [http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel\\_v3.1.pdf](http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf)
- Data Documentation Initiative (2009). *Technical Specification, Part I: Overview, Version 3.1*. Retrieved August 19, 2016, from <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/>
- Data Documentation Initiative (2012). *Developing a Model-Driven DDI Specification*. Retrieved August 19, 2016, from [http://www.ddialliance.org/system/files/DevelopingaModel-DrivenDDISpecification2013\\_05\\_15.pdf](http://www.ddialliance.org/system/files/DevelopingaModel-DrivenDDISpecification2013_05_15.pdf)
- Data Documentation Initiative (2014). *What is DDI?*. Retrieved August 19, 2016, from <http://www.ddialliance.org/what>
- Deutsche Forschungsgemeinschaft (DFG) (2013). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. Weinheim: Wiley-VHC. Retrieved August 19, 2016, from [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf).

- Digital Curation Centre (n.d.) Data Management Plans. Retrieved August 19, 2016, from <http://www.dcc.ac.uk/resources/data-management-plans>
- Dublin Core Metadata Initiative (1994). *DCMI Abstract Model*. Retrieved August 19, 2016, from <http://dublincore.org/documents/abstract-model/>
- European Commission (2010). *Riding the Wave. How Europe Can Gain from the Rising Tide of Scientific Data*. Final report of the high level Expert Group on Scientific Data. European Union. Retrieved August 19, 2016, from <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- European Commission (2012). *Online Survey on Scientific Information in the Digital Age*. Luxembourg: Publications Office of the European Union. Retrieved August 19, 2016, from [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/survey-on-scientific-information-digital-age\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf).
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.
- Generic Statistical Business Process Model (2013). *Introduction to GSBPM v5.0*. Retrieved August 19, 2016, from <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>
- Generic Statistical Information Model (2013). *Introduction to GSIM v1.1*. Retrieved August 19, 2016, from <http://www1.unece.org/stat/platform/display/gsim/Understanding+GSIM>
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608-626.
- Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische mathematik*, *14*(5), 403-420.
- Gregory, A., Heus, P., & Ryssevik, J. (2009). Metadata. *RatSWD Working Paper Series*, *57*, 1-22. Retrieved August 19, 2016, from [http://www.ratswd.de/download/workingpapers2009/57\\_09.pdf](http://www.ratswd.de/download/workingpapers2009/57_09.pdf)
- Gregory, A., & Heus, P. (2007). DDI and SDMX: Complementary, not competing, standards. *Open Data Foundation Paper*. Retrieved August 19, 2016, from [http://www.opendatafoundation.org/papers/DDI\\_and\\_SDMX.pdf](http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf).
- Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. *Zeitschrift für Pädagogische Psychologie*, *56*, 216-227.
- Hankinson, A., Roland, P., & Fujinaga, I. (2011). The music encoding initiative as a document-encoding framework. In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, (pp. 293-298). Retrieved August 19, 2016, from <http://ismir2011.ismir.net/papers/OS3-1.pdf>
- Hagerstown Community College (2014). Student learning outcomes assessment Guide. Retrieved August 19, 2016, from [http://www.hagerstowncc.edu/sites/default/files/documents/HCC%20SLOA%20Guide%201-13-15\\_0.pdf](http://www.hagerstowncc.edu/sites/default/files/documents/HCC%20SLOA%20Guide%201-13-15_0.pdf)
- Harman, K. & Koohang, A. (Eds.). (2007) *Learning Objects: Standards, Metadata, Repositories, and LCMS*. Santa Rosa, California: Informing Science Press.
- Helic, D. (2006). Template-based Approach to Development of Interactive Tests with IMS Question & Test Interoperability. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vol. 1*, (pp. 2075-2081). Retrieved August 19, 2016, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.1661&rep=rep1&type=pdf>
- Huschka, D., Oellers, C., Ott, N., & Wagner, G. C. (2011). *Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften*. In S. Büttner, H-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 35-48). Bad Honnef: Bock und Herchen, S.

- Kimball, R. (2008). *The data warehouse lifecycle toolkit. Second Edition*. New York: Wiley.
- IBM SPSS Statistics (n.d.). Retrieved August 19, 2016, from <http://www.spss.com>
- IMS Global Learning Consortium (2012). *IMS accessible portable item protocol (APIP) Overview - Version 1.0*. Retrieved August 19, 2016, from [http://www.imsglobal.org/apip/apipv1p0/APIP\\_OVW\\_v1p0.html](http://www.imsglobal.org/apip/apipv1p0/APIP_OVW_v1p0.html)
- IMS Global Learning Consortium (2006). *IMS question and test interoperability overview - Version 2.1 Public Draft (revision 2) Specification*. Retrieved August 19, 2016, from [http://www.imsglobal.org/question/qti\\_v2p0/imsqti\\_oviewv2p0.html](http://www.imsglobal.org/question/qti_v2p0/imsqti_oviewv2p0.html)
- Jahnke, T., & Meyerhöfer, W. (Eds.). (2006, revised 2nd edition 2007). *PISA & Co: Kritik eines Programms*. Hildesheim, Berlin: Franzbecker.
- Jesukiewicz, P. (2009). SCORM 2004 4th Edition. Content Aggregation Model. Retrieved August 19, 2016, from [http://www.adlnet.org/wp-content/uploads/2013/09/SCORM\\_2004\\_4ED\\_v1\\_1\\_CAM\\_20090814.pdf](http://www.adlnet.org/wp-content/uploads/2013/09/SCORM_2004_4ED_v1_1_CAM_20090814.pdf)
- Jolliffe, I. (2002). *Principal component analysis*. Springer.
- Kirsch, I., & Yamamoto, K. (2013). PIAAC assessment design. In *Technical Report of the Survey for Adult Skills (PIAAC)* (Ch. 1.). OECD. Retrieved August 19, 2016, from [http://www.oecd.org/site/piaac/Technical%20Report\\_17OCT13.pdf](http://www.oecd.org/site/piaac/Technical%20Report_17OCT13.pdf)
- Klieme, E., & Beck, B. (2007). *Sprachliche Kompetenzen Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim und Basel: Beltz Verlag.
- Krautz, J. (2012). *Ware Bildung: Schule und Universität unter dem Diktat der Ökonomie*. Diederichs Verlag.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.
- Lane, J. (2009). Administrative transaction data. *RatSWD Working Paper Series*, 52, 1-17. Retrieved August 19, 2016, from <http://dx.doi.org/10.2139/ssrn.1445349>
- Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), 2-16.
- Latour, T., & Martin, R. (2007). TAO, an open and versatile computer-based assessment platform based on semantic web technology. *ERICIM News*, 71, 32-33.
- Latour, T., Jadoul, R., & Wagner, M. (2013). Development of the CAPI questionnaire system. In *Technical Report of the Survey for Adult Skills (PIAAC)* (Ch. 7.). OECD. Retrieved August 19, 2016, from [http://www.oecd.org/site/piaac/Technical%20Report\\_17OCT13.pdf](http://www.oecd.org/site/piaac/Technical%20Report_17OCT13.pdf)
- Learning Technology Standards Committee of the IEEE (2002). *IEEE 1484.12.1-2002, Learning Object Metadata Standard*. Retrieved August 19, 2016, from [http://www.imsglobal.org/metadata/mdv1p3pd/imsmd\\_bestv1p3pd.html](http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html)
- Leighton, V. (2002). Developing a new data archive in a Time of maturing standards. *IASSIST Quarterly*, 26, 5-9. Retrieved August 19, 2016, from <http://www.iassistdata.org/downloads/iqvol26lleighton.pdf>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Erlbaum.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 18(1), 50-60.
- Mechtel, M. (2009, May). *EduDDI - an application of DDI 3.0 for large-scale assessments in education*. Paper presented at the annual meeting of the International Association of Social Science Information

- Service and Technology (IASSIST), Tampere, Finland. Retrieved August 19, 2016, from <http://www.ddialliance.org/node/191>
- Medjedovi, I. (2010). Sekundäranalyse. In G. Mey, K. Mruck, (Eds.), *Handbuch Qualitative Forschung in der Psychologie*. (pp. 304-319). Wiesbaden: VS, Verlag für Sozialwissenschaften, S.
- Mistry, R., & Misner S. (2014). *Introducing Microsoft SQL Server 2014*. Washington: Microsoft Press.
- Molenaar, M. & Garrard, T. (2014, September). *Using Standards to Create Best of Breed Assessment Solutions*. Paper presented at the European Association of Test Publishers 2014 (E-ATP 2014), Budapest, Hungary. Retrieved August 19, 2016, from <http://www.slideshare.net/Cito/using-standards-to-createbestofbreed-assessment-solutions>
- Moodle (2014). *Introduction to Moodle 2.8.1*. Retrieved August 19, 2016, from [https://docs.moodle.org/28/en/Main\\_page](https://docs.moodle.org/28/en/Main_page)
- Murray, T., Blessing, S., & Ainsworth, S. (Eds.). (2003). *Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive and intelligent educational software*. Dordrecht: Kluwer Academic Press.
- National Information Exchange Model (2014). *NIEM 3.0 summary*. Retrieved August 19, 2016, from <http://release.niem.gov/>
- National Information Standards Organisation (2004). *Understanding metadata*. Bethesda: NISO Press. Retrieved August 19, 2016, from [http://www.niso.org/publications/press/Understanding Metadata.pdf](http://www.niso.org/publications/press/Understanding_Metadata.pdf)
- Punch, K. F. (2009). *Introduction to research methods in education*. London: Sage.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Ras, E., Swietlik, J., Plichart, P., & Latour, T. (2010). TAO - A versatile and open platform for technology-based assessment. In M. Wolpers, P.A. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Sustaining TEL: From Innovation to Learning and Practice* (pp. 644-649). Heidelberg: Springer.
- Reeff, J. P. (2007). Technische Lösungen für ein computer- und internetbasiertes Assessment-System. In J. Hartig (Eds.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (pp. 81-91). Berlin: BMBF.
- Reißmann, O. (2013). *Überwachung: Wer hat uns verraten? Metadaten!* Retrieved August 19, 2016, from <http://www.spiegel.de/netzwelt/netzpolitik/vorratsdatenspeicherung-wer-hat-uns-verraten-metadaten-a-909942.html>
- Rölke, H. (2012). The ItemBuilder: A Graphical authoring system for complex item development. In T. Bastiaens, & G. Marks (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (pp. 344-353). Chesapeake: VA: AACE.
- Rölke, H. (2013, September). Integrating and extending TAO – Technical background of the EDK Itembank. Paper presented at the TAO Days 2013, Bern, Switzerland. Retrieved August 19, 2016, from <http://www.taotesting.com/wp-content/uploads/2014/09/TAODAYS-2013-R%C3%B6lke.pdf>
- SAS Institute (n.d.). *The SAS System*. Retrieved August 19, 2016, from <http://www.sas.com>
- Schiller, D. (2013). *Proposal for a European Remote Access Network (EuRAN) – main components*. Paper presented to the Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa, Canada. Retrieved August 19, 2016, from [http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic\\_3\\_Schiller.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_3_Schiller.pdf)

- Schneider, H., & Wolf, C. (2008). Die Datenservicezentren als Teil der informellen Infrastruktur. In G. Rolf, M. Zwick, & G.G. Wagner (Eds.), *Fortschritte der informationellen Infrastruktur in Deutschland* (pp. 1-19). Baden-Baden: Nomos Verlagsgesellschaft.
- Schmitz, C. (2012). LimeSurvey: An open source survey tool. LimeSurvey Project Hamburg, Germany. URL <http://www.limesurvey.org>.
- Schuster, A., & Bambey, D. (2014). Potenziale einer zentralen Verfügbarkeit von Daten und Instrumenten im Forschungsdatenzentrum (FDZ) Bildung. *Zeitschrift für interpretative Schul- und Unterrichtsforschung*, 3,(1), 152-153.
- Severance, C., Hanss, T., & Hardin, J. (2010). LMS learning tools interoperability: Enabling a mash-up approach to teaching and learning tools. *Technology, Instruction, Cognition and Learning*, 7(3-4), 245-262.
- Siegert, O. (2010). Speicherung und Publikation von Forschungsdaten. Der Beitrag der Deutschen Zentralbibliothek für Wirtschaftswissenschaften. *RatSWD Working Paper Series*, 158. Retrieved August 19, 2016, from [http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_158.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_158.pdf).
- Simsion, G. C., & Witt, G. C. (2005). *Data Modeling Essentials. 3rd edition*. San Francisco, CA: Morgan Kaufmann.
- Smit, E. (2011). Abelard and Héloïse: Why Data and Publications Belong Together. *D-Lib Magazine*, 17(1-2). Retrieved August 19, 2016, from <http://www.dlib.org/dlib/january11/smit/01smit.html>
- Smythe, C., & Roberts, P. (2000). An overview of the IMS question & test interoperability specification. In *Proceedings of the 4th CAA Conference*. Loughborough: Loughborough University. Retrieved August 19, 2016, from <https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1784/1/smythec00.pdf>
- Solga, H., & Wagner, G. G. (2007). A Modern statistical infrastructure for excellent research and policy advice: Report on the German Council for Social and Economic Data during its first period in office (2004–2006). *RatSWD Working Paper Series*, 2. Retrieved August 19, 2016, from <http://dx.doi.org/10.2139/ssrn.1417451>
- Söderberg, L.-J. (2005, November). *MONA-Microdata ON-line access at statistics Sweden*. Paper presented at Joint UNECE/Eurostat work session on statistical data confidentiality, Geneva, Switzerland. Retrieved August 19, 2016, from <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.3.e.pdf>
- Statacorp Stata. (n.d.) Retrieved August 19, 2016, from <http://www.stata.com>
- Statistics Netherlands (2005). *Blaise for Windows 4.5 Developer's Guide*. Heerlen: Methods and Informatics Department. Retrieved August 19, 2016, from <http://www.dfpp.univr.it/documenti/OccorrenzaIns/matdid/matdid077301.pdf>
- Tenopir, C., Palmer, C. L., Metzger, L., van der Hoeven, J., & Malone, J. (2011). Sharing data: Practices, barriers, and incentives. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-4.
- UK Data Archive (n.d.) *Create and manage data: Planning for sharing*. Retrieved August 19, 2016, from <http://www.data-archive.ac.uk/create-manage/planning-for-sharing>.
- Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. In I. B. Weiner (Ed.), *Handbook of Psychology, Second Edition* (pp. 607-634). John Wiley & Sons, Inc.
- Upsing, B., Gissler, G., Goldhammer, F., Rölke, H. & Ferrari, A. (2011). Localisation in international large-scale assessments of competencies: Challenges and solutions. *Localisation Focus*, 10(1), 42-57.



- Vale, S. (2010). Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model. *DDI Working Paper Series*. Retrieved August 19, 2016, from <http://dx.doi.org/10.3886/DDIOtherTopics01>
- Vardigan, M., Pascal H., & Wendy T. (2008). Data documentation initiative: Toward a standard for the social sciences. *The International Journal of Digital Curation*, 3(1), 107-113.
- Vartanian, T. P. (2011). *Secondary Data Analysis*. New York: Oxford University Press.
- Villar, A., Callegaro, M., & Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review*, 0894439313497468.
- W3C (2009). *Hypertext Markup Language 4.01 Specification*. Retrieved August 19, 2016, from <https://www.w3.org/TR/2009/REC-html401-20090303/>
- Weermann, B., Oriens, B. & Colvin, T. (2011). *MMIC Manual*. Retrieved August 19, 2016, from <https://mmic.rand.org/research/rand/demo/manual.pdf>
- Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). *Dublin core metadata for resource discovery*. Retrieved August 19, 2016, from <https://tools.ietf.org/html/rfc2413>
- Weichselgartner, E., Günther, A., Dehnhard, I. (2011). Archivierung von Forschungsdaten. In S. Büttner, H-C. Hobohm, L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 191-202). Bad Honnef: Bock und Herchen.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Wijnand, A. & De Bruijne, M. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers. An Experiment with a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review*, 31, 482-504.
- Wissenschaftsrat (2012). *Empfehlungen der Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*. Berlin: Wissenschaftsrat. Retrieved August 19, 2016, from <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.
- Wittenberg, R., Cramer, H., & Vicari, B. (2014). *Datenanalyse mit IBM SPSS Statistics. Eine syntaxorientierte Einführung*. Konstanz und München: UVK Verlagsgesellschaft.
- Wuttke, J. (2009). PISA: Nachträge zu einer nicht geführten Debatte. *Mitteilungen der Gesellschaft für Didaktik der Mathematik* 87, 22-34.
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3). Retrieved August 19, 2016, from <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00259.x/full>
- Zabal, A., Silke, M., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I. & Rammstedt, B. (2014). *PIAAC Germany 2012: Technical report*. Münster: Waxmann.

## Appendices

### I. Main Study Questionnaire in detail

#### - Chapter 1 – Main Study - “Your research”

Code	RO01
Filter	None
Question Text	Are you involved in survey design, experimental design or instrument development?
Response Domain	Yes / No
Rationale	Separation of persons involved in survey design from other respondents thus shortening the questionnaire

Code	RO02
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	How often do you have to provide a data management plan (i.e. a plan that describes how to document or store the data for later usage)?
Response Domain	Option List
Options	Scale from “Always” to “Never” plus “Do not know”
Rationale	How familiar are researchers with data management plans

Code	RO03
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	How do you store your research data, e.g. survey or experimental data?
Response Domain	Multiple Choice
Choices	“Research data is stored in a research data facility or archive which belongs to the institution I work for.” “Research data is stored in a third party research data facility or archive.” “Research data stays with the Prime Investigator (PI) or project team.” “Research data stays on the IT infrastructure of the project.” “Do not know”
Rationale	How often do researchers use research data centres

Code	RO04
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	Approximately how many months did the last study, e.g. survey or experiment you were involved in take? This overall time may include survey design, data collection, analysis and publication.
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey

Code	RO05
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	How many persons were involved in your last study?
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey

Code	RO06
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	Approximately how much time, as in percent of the total project time, was used for documenting data in your last study?
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey

- Chapter 2 – Main Study - “Metadata standards”

Code	MS01
Filter	None
Question Text	Metadata standards play an important role in data archives. We are interested in how often researchers use them in their surveys and projects. Please specify to what extent do you use these metadata standards?
Response Domain	Matrix
Options	“I use this metadata standard often.” “I use this metadata standard rarely.” “I have never used this metadata standard.” “I do not know this metadata standard.”
Sub Questions	“Data Documentation Initiative (DDI)” “Dublin Core” “Learning Object Model (LOM)” “Learning Tools Interoperability (LTI)” “Machine Readable Catalogueing (MARC)” “Metadata Encoding and Transmission Standard (METS)” “QueDex” “Question and Test Interoperability (QTI)” “Sharable Content Object Reference Model (SCORM)” “Statistical Data and Metadata Exchange (SDMX)” “Text Encoding Initiative (TEI)”
Rationale	Which metadata standards are known to the respondents?

Code	MS02
Filter	MS01_MS101.NAOK = “I use the metadata standard often” “Data Documentation Initiative (DDI)”= “I use the metadata standard often”
Question Text	How satisfied are you with the metadata standard Data Documentation Initiative (DDI)?
Response Domain	Option List
Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the metadata standard DDI.

Code	MS03
Filter	MS01_MS101.NAOK = “I use the metadata standard often” “Data Documentation Initiative (DDI)”= “I use the metadata standard often”
Question Text	Which versions of DDI are you using?
Response Domain	Multiple Choice
Options	“DDI Lifecycle V3.2” “DDI Lifecycle V3.1” “DDI Lifecycle V3.0” “DDI Codebook V2.5” “DDI Codebook V2.1” “Earlier version than DDI Codebook V2.1”
Rationale	Which versions of DDI are used by the respondents? Combining satisfaction from the previous question with the version of the standard.

Code	MS04
Filter	MS01_MS108.NAOK = “I use the metadata standard often” “Question and Test Interoperability (QTI)”= “I use the metadata standard often”
Question Text	How satisfied are you with the metadata standard Question and Text Interoperability (QTI)?
Response Domain	Option List
Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the metadata standard QTI.

Code	MS06
Filter	None
Question Text	How much do you agree on the following statements regarding metadata standards?
Response Domain	Matrix
Options	Scale from “Strongly agree” to “Strongly disagree” plus “Do not Know”
Sub Questions	“I am always able to identify the fitting metadata standards for my needs.” “To my knowledge there are no metadata standards in my scientific area.” “There are a lot of benefits in using metadata standards.” “My studies are too complex to use metadata standards for data documentation.”
Rationale	MS05 (omitted) and MS06 were in the field test questions for respondents who said they do not use metadata standards. Unfortunately this combination did not work out as planned. In the main test all respondents therefore have to evaluate opinions on metadata standards. The reasons for not using metadata standards have been modified into positive answers to avoid double negatives. Somebody who does not use metadata standards would therefore answer “There are a lot of benefits in using metadata standards.” with “Disagree” or “Strongly disagree”.

- Chapter 3 – Main Study - “Data Sources”

Code	DA01
Filter	None
Question Text	How often do you use the following data sources for your research?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Structured cognitive tests (e.g. computer-based assessments)” “Structured questionnaires (e.g. multiple choice)” “Qualitative text data (e.g. narrative interviews, oral history)” “Experimental designs” “Observational data (e.g. recordings, transcripts)”
Rationale	Which kinds of data sources are the respondents using for their research?

- Chapter 4 – Main Study - “Research Documentation”

Code	SU01
Filter	None
Question Text	Which of the following categories do you think are important if a research project, e.g. a survey or study has to be documented?
Response Domain	Matrix
Options	Scale from “Very important” to “Not at all important”
Sub Questions	“Analysis Plan” “Concepts of the study” “Data cleaning methods” “Funding Sources” “Organizations involved” “Previous publications” “Quality control mechanisms” “Research team” “Sampling methods” “Structure of the study” “Target population / Universe” “Technical Infrastructure (e.g. data collection servers, interviewer laptops)”
Rationale	These are general terms derived from metadata standards (e.g. DDI) or models (e.g. GLBPM) describing the first phase of the survey cycle. If researchers have resource issues (time or money) and thus cannot invest much time into data documentation it would be important to know what they consider to be valuable information.

- Chapter 5 – Main Study - “Data Collection Instruments”

Code	IT01
Filter	None
Question Text	Are you involved in the creation of items for questionnaires or cognitive tests?
Response Domain	Yes / No
Rationale	Separation of item designers from the rest of the respondents.

Code	IT02
Filter	IT01.NAOK = “Yes” “Are you involved in the creation of items for questionnaires or cognitive tests?” = “Yes”
Question Text	How often do you provide these different types of metadata with a questionnaire or cognitive item?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Analytic scripts (e.g. SPSS or Stata code that generates variables)” “Answer categories” “Flow logic (e.g. routing, filtering, dynamic text)” “Fragments of software (e.g. programming code for a specific item)” “Information about the concepts to be measured” “Interviewer instructions” “Layout definitions” “References to research literature” “Scoring rules (e.g. information in a cognitive test which answers were correct or incorrect)” “Statistical values about the items from previous studies” “Stimuli in reusable form (e.g. pictures, videos, audio files)” “Variable definitions”
Rationale	These are general terms derived from metadata standards (e.g. DDI or QTI). As the filter questions before only lets respondents through who design items the question is which elements they deliver with an item if they would be considered to be data producers. This can be compared with the expectations of data users in question IT12. The question was asked the same way as matrix question in the field test, but it contained less examples.

Code	IT11
Filter	None
Question Text	Have you been re-using or are planning to re-use items or questions developed by other persons or institutions?
Response Domain	Yes / No
Rationale	Separation of secondary item users from the rest of the respondents.

Code	IT12
Filter	IT11.NAOK = “Yes” “Have you been re-using or are planning to re-use items or questions developed by other persons or institutions?” = “Yes”
Question Text	How often do you provide these different types of metadata with a questionnaire or cognitive item?
Response Domain	Matrix
Options	Scale from “Very important” to “Not at all important”
Sub Questions	“Analytic scripts (e.g. SPSS or Stata code that generates variables)” “Answer categories” “Flow logic (e.g. routing, filtering, dynamic text)” “Fragments of software (e.g. programming code for a specific item)” “Information about the concepts to be measured” “Interviewer instructions” “Layout definitions” “References to research literature” “Scoring rules (e.g. information in a cognitive test which answers were correct or incorrect)”

	<p>“Statistical values about the items from previous studies”</p> <p>“Stimuli in reusable form (e.g. pictures, videos, audio files)”</p> <p>“Variable definitions”</p>
Rationale	<p>These are general terms derived from metadata standards (e.g. DDI or QTI). The filter before filters people who are considering re-use or are re-using items. IT12 can be seen as mirror question to IT02 as it compares the importance of these concepts to the availability from the item designers thus leading to a basis for new models or new software products. Originally this question was designed as ranking question but was transformed into a matrix due to field test feedback. The disadvantage if a respondent is an item designer and a re-user of items the questions might be too similar and thus boring leading to break-offs.</p>

- Chapter 6 – Main Study - “Data Analysis”

Code	DC11
Filter	None
Question Text	Do you analyze data which is provided by a data collection agency, research data center or data archive?
Response Domain	Yes / No
Rationale	Separation of secondary data users from the rest of the respondents.

Code	DC12
Filter	DC11.NAOK = “Yes” “Do you analyze data which is provided by a data collection agency, research data center or data archive?” = “Yes”
Question Text	How often are the following metadata or paradata information included with the data sets delivered to you? Note: paradata are administrative metadata about the data collection process on a micro-level (e.g. how many days the interview took, contact attempts by the interviewer, timing of answers by the interviewee).
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	<p>“Codebook with variable definitions”</p> <p>“Cleaned data files”</p> <p>“Descriptions about data cleaning processes”</p> <p>“Field report (e.g. contact attempts, information about the interviewers)”</p> <p>“Log files from each collected case (e.g. changes of answers including time stamps)”</p> <p>“Quality control reports”</p> <p>“Raw data files”</p>
Rationale	These are files which are produced by data collection agencies during field work. The question is which of them are forwarded to the researcher to support the data analysis process.

Code	DC13
Filter	DC11.NAOK = “Yes” “Do you analyze data which is provided by a data collection agency, research data center or data archive?” = “Yes”

Question Text	In general, how satisfied are you with the metadata and paradata information you receive from data collection agencies, research data centres or data archives?
Response Domain	Option List
Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the metadata delivered by data collection agencies.

Code	DC14
Filter	DC11.NAOK = “Yes” “Do you analyze data which is provided by a data collection agency, research data center or data archive?” = “Yes”
Question Text	How often do you use the following data sources for your research?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Data enclaves, secure data services or Virtual Research Environments (e.g. NORC, MONA, RemoteNEPS)” “Data warehouses or analytical databases” “Personal data extracts created by portal websites (e.g. variable shopping baskets like in Questasy or IPUMS)” “Public Use Files (e.g. highly anonymized datasets which can be downloaded from data centres without registration)” “Scientific Use Files (e.g. moderately anonymized datasets which can be downloaded from data centres after an application process)” “Raw datasets (non-anonymized datasets)” “Relational databases (e.g. SQL Server, Oracle, MySQL)” “Remote calculation or job submission services (e.g. Joshua)”
Rationale	Research data centres offer different ways of accessing the data. This question tries to identify which of them are used by researchers.

Code	DC21
Filter	None
Question Text	Do you provide datasets to others?
Response Domain	Yes / No
Rationale	Separation of data producers from the rest of the respondents.

Code	DC22
Filter	DC21.NAOK = “Yes” “Do you provide datasets to others?” = “Yes”
Question Text	Are the users of your data mostly internal or external?
Response Domain	Option List
Options	Scale from “Only external” to “Only internal”
Rationale	To distinguish if the data collection happens for research teams in the own facility or rather for third parties (the latter would be a sign for a commercial data collection or market research agency)

Code	DC23
Filter	DC21.NAOK = “Yes” “Do you provide datasets to others?” = “Yes”
Question Text	Which metadata or paradata (i.e. metadata derived from data collection processes) information do you usually provide to the users of your data?



Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Codebook with variable definitions” “Cleaned data files” “Descriptions about data cleaning processes” “Field report (e.g. contact attempts, information about the interviewers)” “Log files from each collected case (e.g. changes of answers including time stamps)” “Quality control reports” “Raw data files”
Rationale	This question mirrors DC14 for data collection agencies by asking which additional files are delivered to the research teams for data analysis.

- Chapter 7 – Main Study - “Paradata”

Code	DI03
Filter	None
Question Text	Do you analyze paradata from log files collected during the survey (e.g. timestamps)? Note: paradata are administrative metadata about the data collection process on a micro-level (e.g. how many days the interview took, contact attempts by the interviewer, timing of answers by the interviewee).
Response Domain	Yes / No
Rationale	Identification of researchers who disseminate paradata

Code	DI04
Filter	DI03.NAOK = “Yes” “Do you analyze paradata from log files collected during the survey (e.g. timestamps)?” = “Yes”
Question Text	When you analyze paradata from log files, which formats do you prefer?
Response Domain	Multiple Choice
Options	“CSV file” “Excel file” “Mplus file” “R file” “Relational Database” “SAS file” “SPSS file” “Stata file” “TXT file” “Word document” “XML file”
Rationale	Log files come in different recording formats. This question tries to identify the preference of researchers.

Code	DI05
Filter	DI03.NAOK = “Yes” “Do you analyse paradata from log files collected during the survey (e.g. timestamps)?” = “Yes”
Question Text	In general, how satisfied are you with the documentation of paradata from log files you use for your research?
Response Domain	Option List

Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the paradata documentation.

- Chapter 8 – Main Study - “Publication”

Code	PU01
Filter	None
Question Text	Are you publishing scientific research in journals?
Response Domain	Yes / No
Rationale	Separation of researchers from the rest of the respondents.

Code	PU02
Filter	PU01.NAOK = “Yes” “Are you publishing scientific research in journals?” = “Yes”
Question Text	How important to you are the following factors in choosing a journal for publication?
Response Domain	Matrix
Options	Scale from “Very important” to “Not at all important”
Sub Questions	“Costs” “Fast publishing process” “Impact factor” “Linkage to data” “Open access” “Reputation of the journal”
Rationale	From a data documentation point of view this question tries to analyse how important data linkage is in comparison to typical quality factors of journals.

Code	PU03
Filter	None
Question Text	Do you share the metadata, paradata or data from your research with others?
Response Domain	Yes / No
Rationale	Separation between researchers who share their data and researchers who do not share their data.

Code	PU04
Filter	PU03.NAOK = “Yes” “Do you share the metadata, paradata or data from your research with others?” = “Yes”
Question Text	With whom do you share your data?
Response Domain	Multiple Choice
Options	“I deliver the data to a research data centre or archive.” “I hand over the data to a designated person at my organization.” “I share my data only with researchers I know.” “I share my data with other researchers from the same field.” “I share my data with other researchers from other fields.” “I share my data only with researchers from my organization.” “I share my data with other persons who are interested in my research.”
Rationale	These choices indicate how far researchers are willing to spread their data to third parties. If somebody says “I share my data only with researchers I

	know.” forwarding the data to a RDC and providing enhanced publications is not possible though the person in PU03 claims they are sharing their data.
--	---

Code	PU05
Filter	PU03.NAOK = “No” “Do you share the metadata, paradata or data from your research with others? = “No”
Question Text	Please indicate the reasons why you do not share your data.
Response Domain	Matrix
Options	Scale from “Strongly agree” to “Strongly disagree”
Sub Questions	“I do not know where to hand in my data for archiving.” “I have no resources or support from my employer to hand in the data.” “Preparation of the data for archiving needs too much effort.” “My organization or I have security concerns.” “Somebody could misinterpret my results.” “Somebody could use my data before me.”
Rationale	This question is for researchers who indicated in PU03 they are not willing to share their data and gives them reasons to evaluate why they do not do it.

Code	PU06
Filter	None
Question Text	Which of the following could be important motivating factors for you to share your data with others?
Response Domain	Matrix
Options	Scale from “Very important” to “Not at all important”
Sub Questions	“Archive informs data producer which users accessed the data.” “Archive publically provides a list of publications based on the data.” “Co-authorship in secondary publications based on the data” “Data is cited in secondary publications” “Support by the employer (e.g. financial resources, time, staff)” “The data producer receives a financial compensation”
Rationale	Originally this question was limited to researchers who do not share their data. Nevertheless the field test showed this question could be slightly changed and asked to the whole respondent group.

- Chapter 9 – Main Study - “Software”

Code	SO02
Filter	None
Question Text	Some software products support the use of metadata standards. We are interested to know which of these products are known to researchers. To what extent do you know the following software products for use in the area of questionnaires or computer-based assessment?
Response Domain	Matrix
Options	“I know this software well.” “I have a basic understanding of what this software does.” “I have only heard of this software, but have never seen it.” “I do not know this software.”
Sub Questions	“Blaise” “CASES”

	“CBA Itembuilder” “eDia” “Globalpark / Unipark” “Limesurvey” “MyItemwriter” “Nipo / Odin / Compass” “Questify” “Questionmark” “Rogatus Survey / Repository” “Surveymonkey” “Testdrive” “Testing Assistee par Ordinateur (TAO)” “Voxco”
Rationale	This question contains a mix of different software products from survey work and it tries to identify which products are known. Survey modes (CAPI, CATI, P&P) as well as research area (educational sciences, social sciences) are mixed here as this software products are in reality also used in multiple domains (e.g. TAO for questionnaires, Blaise for simple assessments) therefore the line cannot be easily drawn. In the field test there used to be a filter question “SO01” which asked if the respondents are using computer-based assessment or survey software but it has been omitted due to field test feedback (the respondents did not know what kind of software this is without seeing the actual list).

- Chapter 10 – Main Study - “Organization”

Code	OR01
Filter	None
Question Text	Who is your employer? If you have more than one employer please specify the one where you spend the most of your time.
Response Domain	Option List
Options	“University (including Colleges and Technical Universities)” “Public Research Institute” “Private Research Institute” “Data Collection Agency / Market Research Agency” “Software Development Company” “Other private sector organization”
Rationale	To identify the work place of the respondent.

- Chapter 11 – Main Study - “Information on Researcher’s background”

Code	RB01
Filter	OR01.NAOK = “OR101” OR OR01.NAOK = “OR102” OR OR01.NAOK = “OR103”  “Who is your employer?” = “University (including Colleges and Technical Universities)” OR “Who is your employer?” = “Public Research Institute” OR “Who is your employer?” = “Private Research Institute”

Question Text	What is your current academic status? Please choose the option which reflects best your daily work even if more than one role fits to you.
Response Domain	Option List
Options	“Undergraduate student (Bachelor level)” “Graduate student (Master level)” “Doctoral student / PhD Student” “Researcher (without PhD)” “Postdoctoral Researcher (with PhD)” “Assistant / Associate Professor” “Full Professor” “Not a researcher”
Rationale	To identify the academic status of the respondent. Code RB108 is necessary as there might be personnel working in academic institutions as non-scientists (e.g. administrative personnel).

Code	RB02
Filter	RB01.NAOK IS NOT “RB108” AND (OR01.NAOK = “OR101” OR OR01.NAOK = “OR102” OR OR01.NAOK = “OR103”)  “What is your current academic status?” IS NOT “Not a researcher” AND (“Who is your employer?” = “University (including Colleges and Technical Universities)” OR “Who is your employer?” = “Public Research Institute” OR “Who is your employer?” = “Private Research Institute”)
Question Text	How many years have you been working in academia?
Response Domain	Number (Integer)
Rationale	To identify academic working experience.

Code	RB03
Filter	RB01.NAOK IS NOT “RB108” AND (OR01.NAOK = “OR101” OR OR01.NAOK = “OR102” OR OR01.NAOK = “OR103”)  “What is your current academic status?” IS NOT “Not a researcher” AND (“Who is your employer?” = “University (including Colleges and Technical Universities)” OR “Who is your employer?” = “Public Research Institute” OR “Who is your employer?” = “Private Research Institute”)
Question Text	Which areas of research have you been working in?
Response Domain	Multiple Choice
Options	“Agricultural Sciences” “Arts” “Computer Science” “Educational Sciences” “Engineering” “Human medicine, Health Science or Veterinary Sciences” “Law and Economics” “Linguistics and cultural studies” “Medicine” “Social Sciences”

	“Sports”
Rationale	To identify the academic field. The question was originally an option field so researchers had to choose their main topic, but the field test respondents found this too limiting.

## Chapter 12 – Main Study - “Region”

Code	RE01
Filter	None
Question Text	Which region do you work in primarily?
Response Domain	Option List
Options	“Africa” “Asia” “Australia / Oceania” “Central and South America” “European Union (EU)” “Europe (Non-EU)” “North America”
Rationale	To identify the region of the respondent

Code	RE02
Filter	RE01.NAOK = “RE105” “Which region do you work in primarily?” = “European Union (EU)”
Question Text	In which country of the European Union (EU) do you work in primarily?
Response Domain	Option List
Options	“Austria” “Belgium” “Bulgaria” “Croatia” “Cyprus” “Czech Republic” “Denmark” “Estonia” “Finland” “France” “Germany” “Greece” “Hungary” “Ireland” “Italy” “Latvia” “Lithuania” “Luxembourg” “Malta” “Netherlands” “Poland” “Portugal” “Romania” “Slovakia” “Slovenia” “Spain” “Sweden” “United Kingdom”

Rationale	To identify the EU country where the respondent works in.
-----------	---

Code	RE03
Filter	RE01.NAOK = "RE107" "Which region do you work in primarily?" = "North America"
Question Text	In which country in North America do you work in primarily?
Response Domain	Option List
Options	"Canada" "United States of America (USA)"
Rationale	To identify the North American country where the respondent works in.

- Chapter 13 – Main Study - “Personal Information”

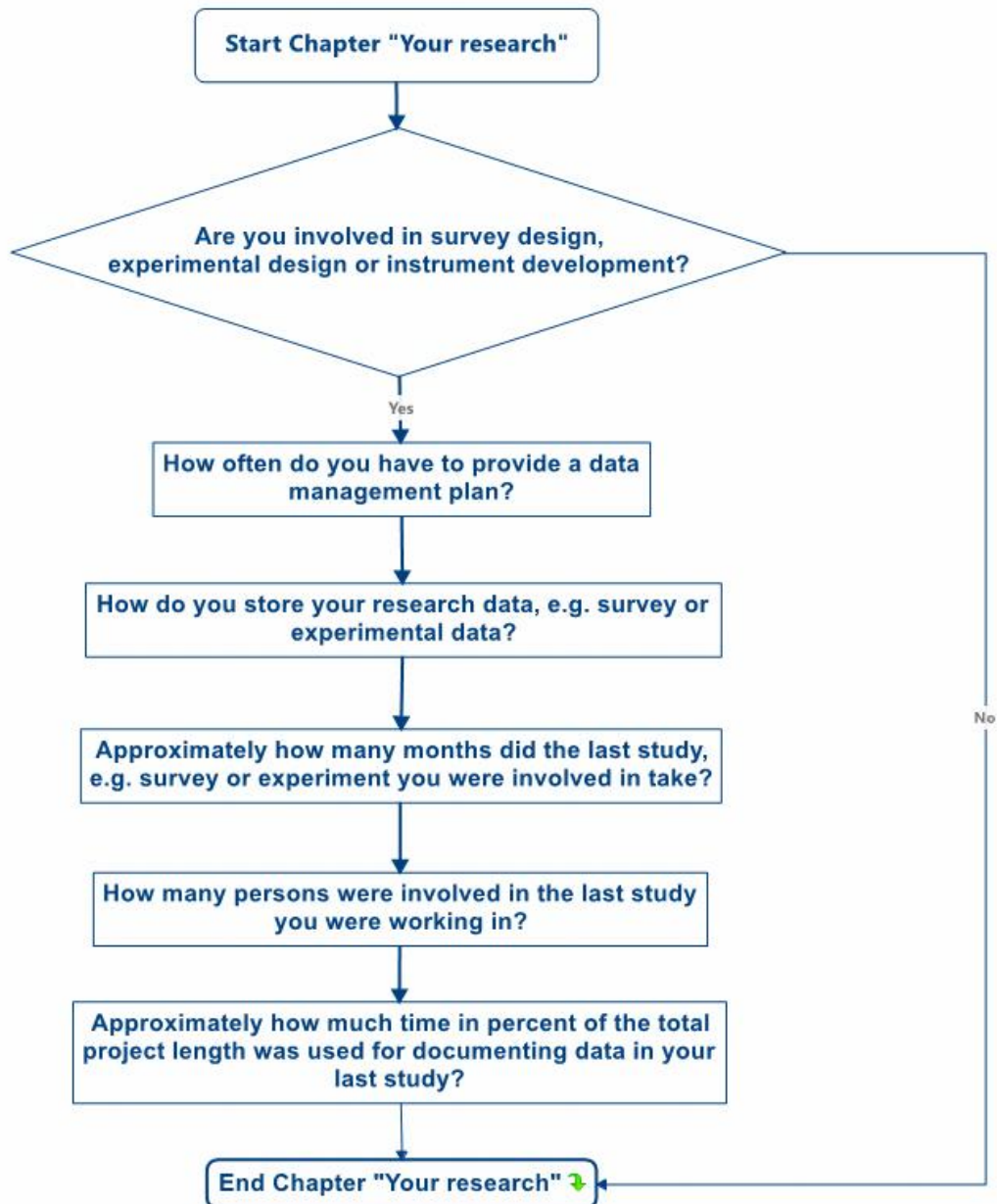
Code	PI01
Filter	None
Question Text	In which year were you born?
Response Domain	Number (Integer)
Rationale	To identify the respondents’ age.

Code	PI02
Filter	None
Question Text	What is your gender?
Response Domain	Gender
Rationale	To identify the respondents’ gender. The survey uses the standard scale of Limesurvey which uses “male” and “female” as these can be considered sufficient for this sample size (the use of “transgender” and other options have no analysis potential add due to very small numbers to be expected). As all questions are optional respondents can skip this question so there should not be a moral dilemma for persons who do not feel themselves comfortable with only two options.

## II. Main Study Questionnaire Flow Logic

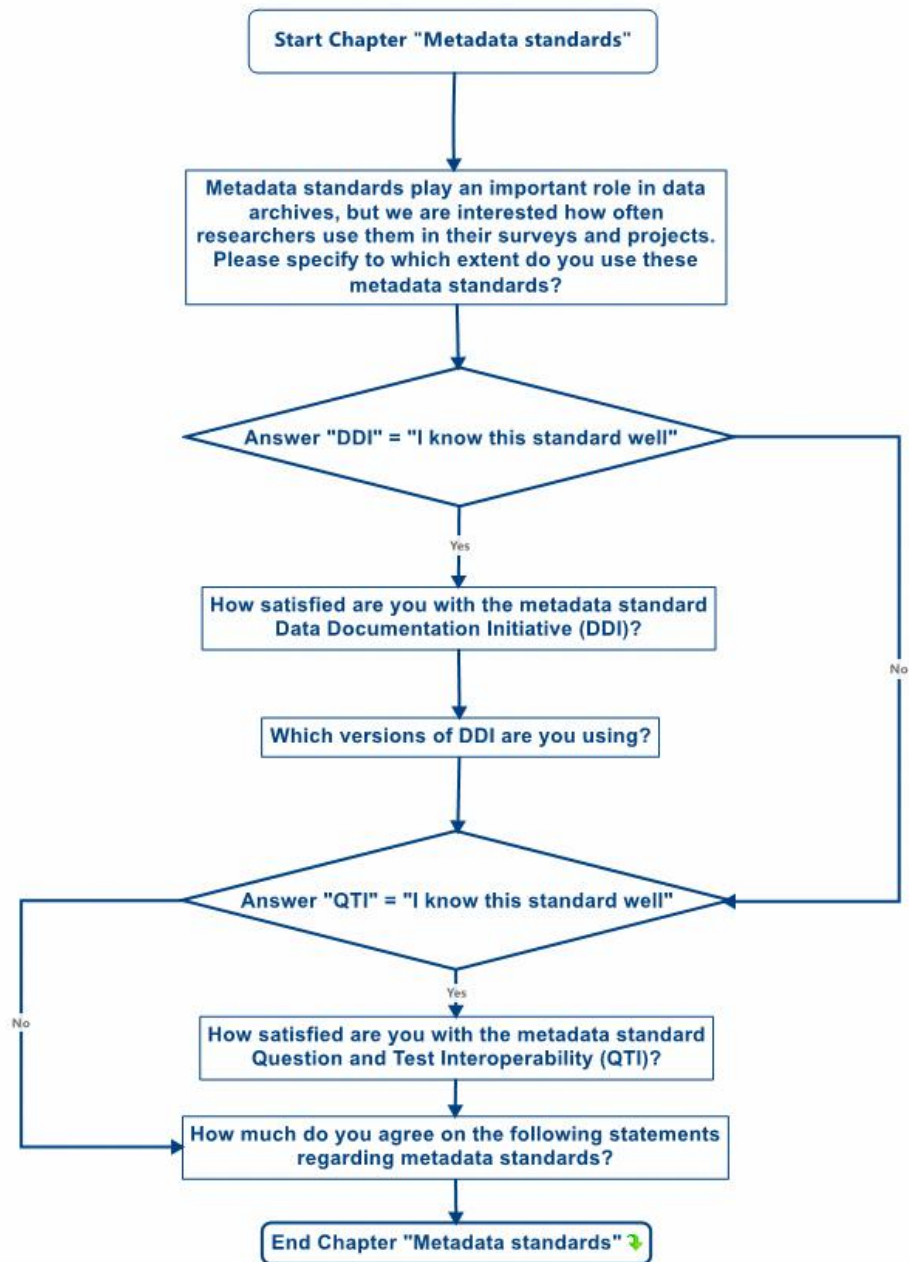
To give a better overview of the mechanics of the questionnaire here the flow logic. The flow charts have been divided by chapters. There is a routing between the two chapters “Organization” and “Information on Researcher’s background” but it is not too complex.

- Chapter 1 – Main Study - “Your research”

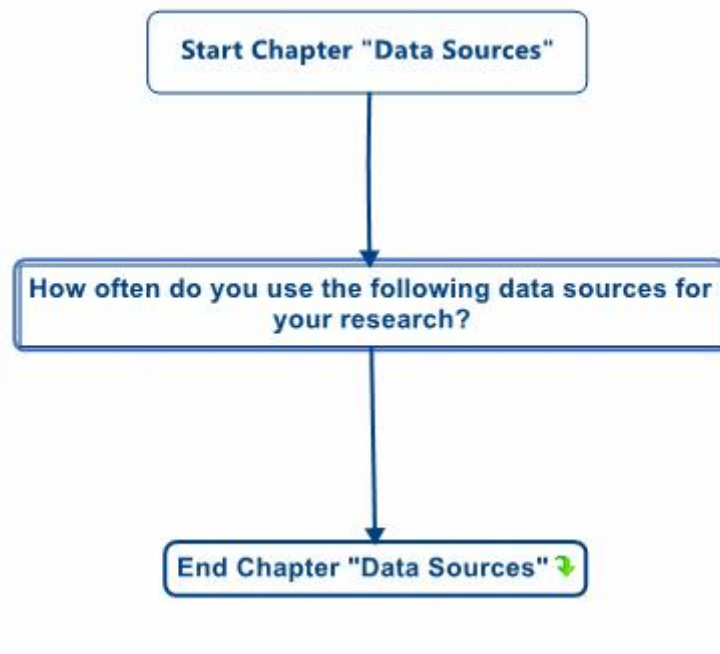




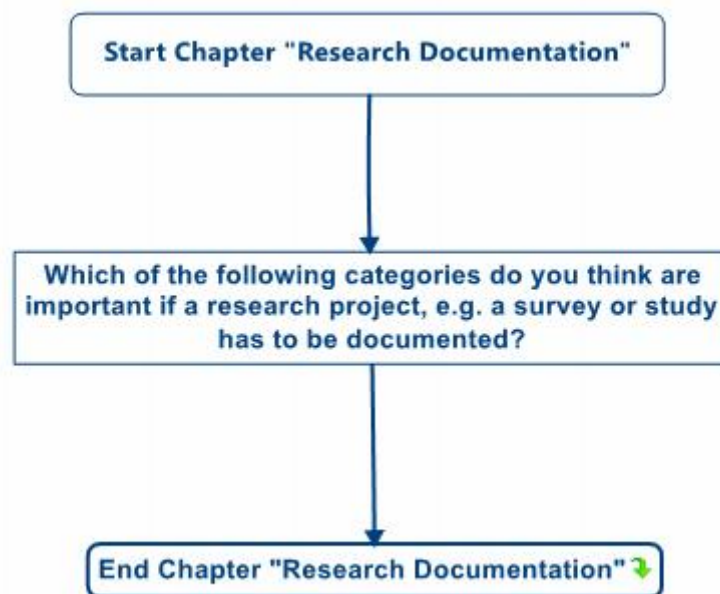
- Chapter 2 – Main Study - “Metadata standards”



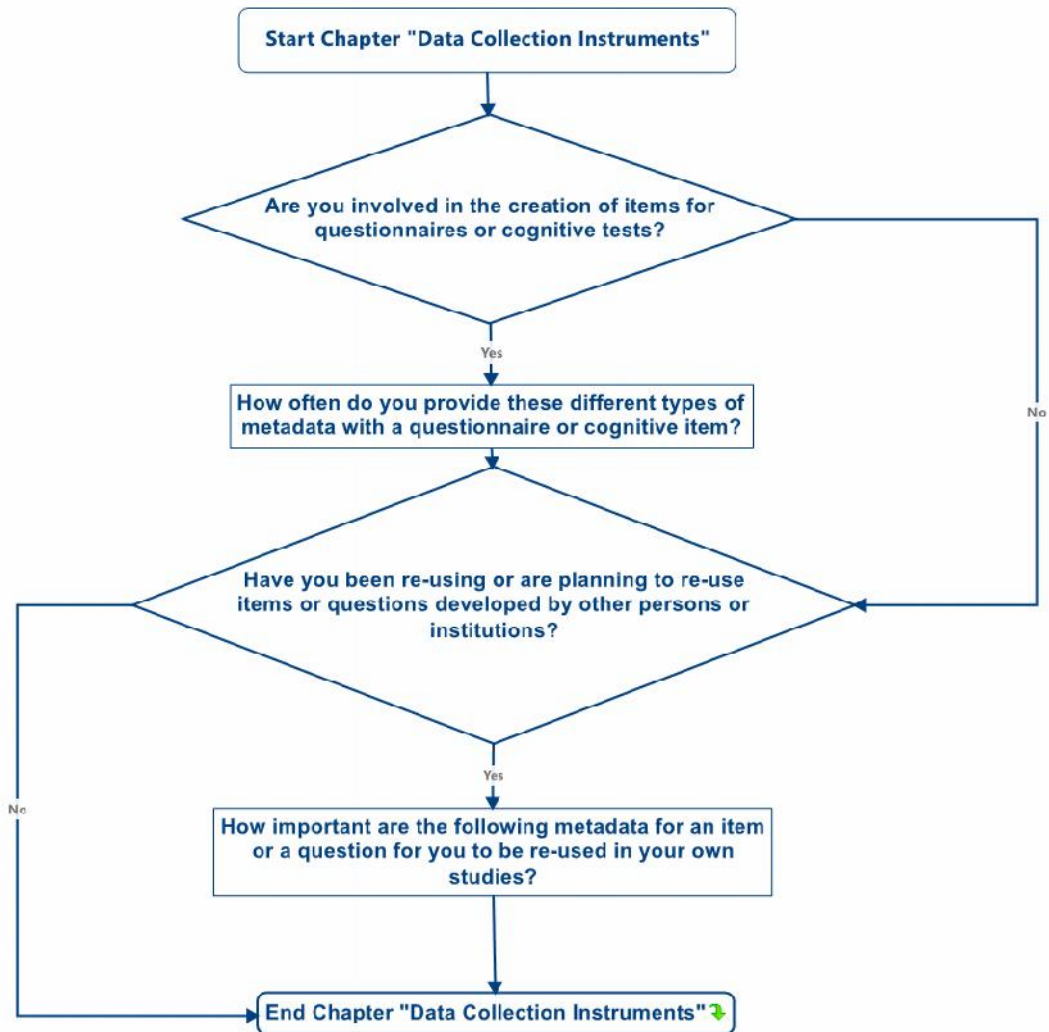
- Chapter 3 – Main Study - “Data Sources”



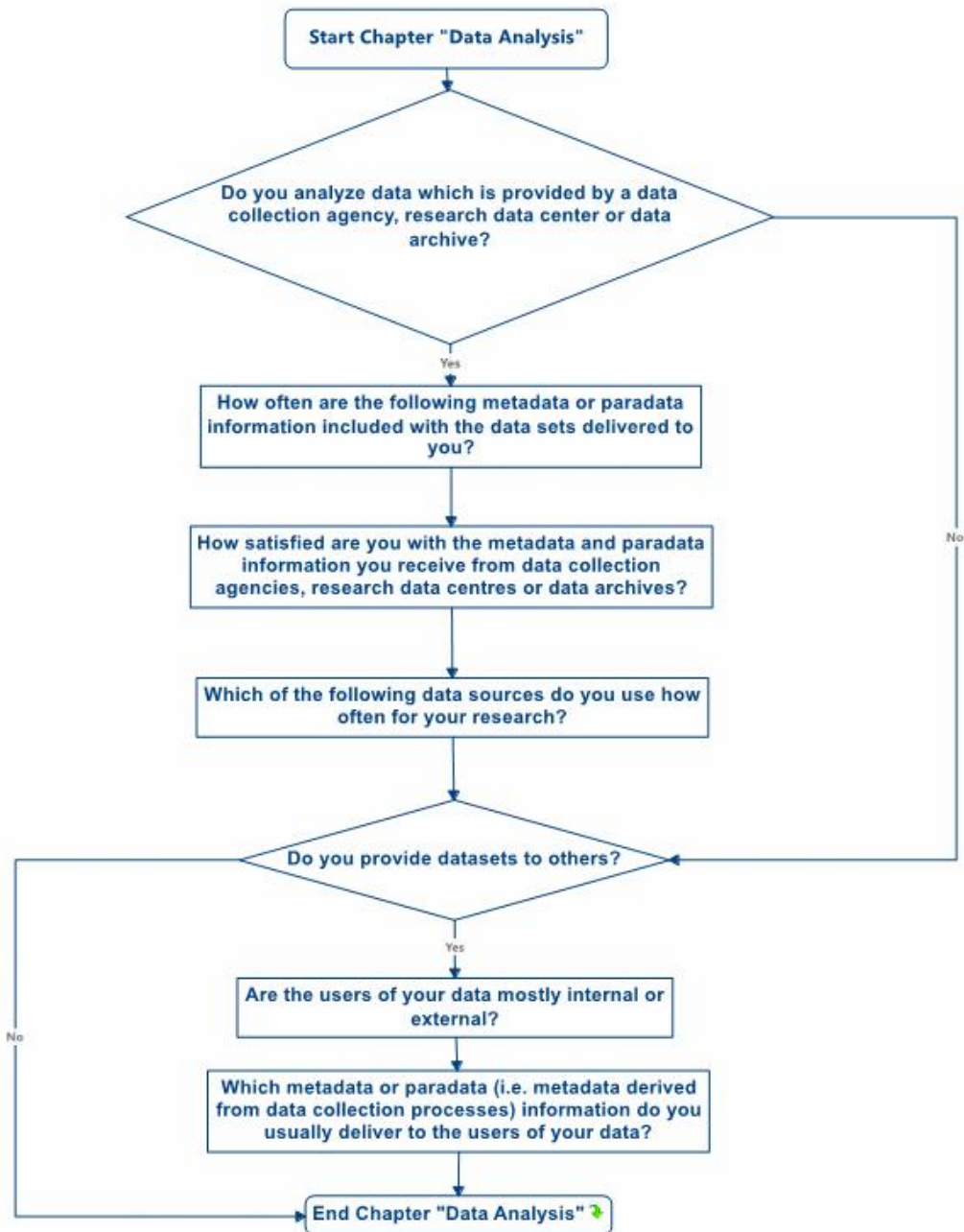
- Chapter 4 – Main Study - “Research Documentation”



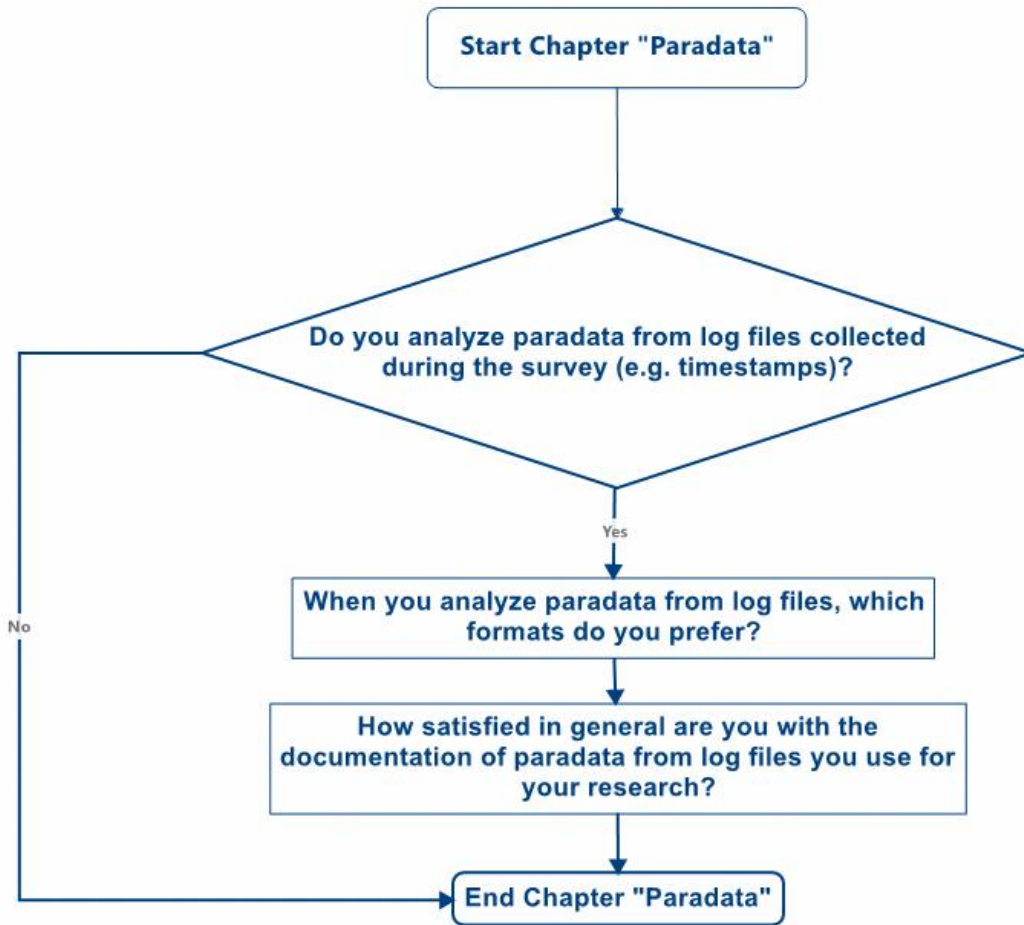
- Chapter 5 – Main Study - “Data Collection Instruments”



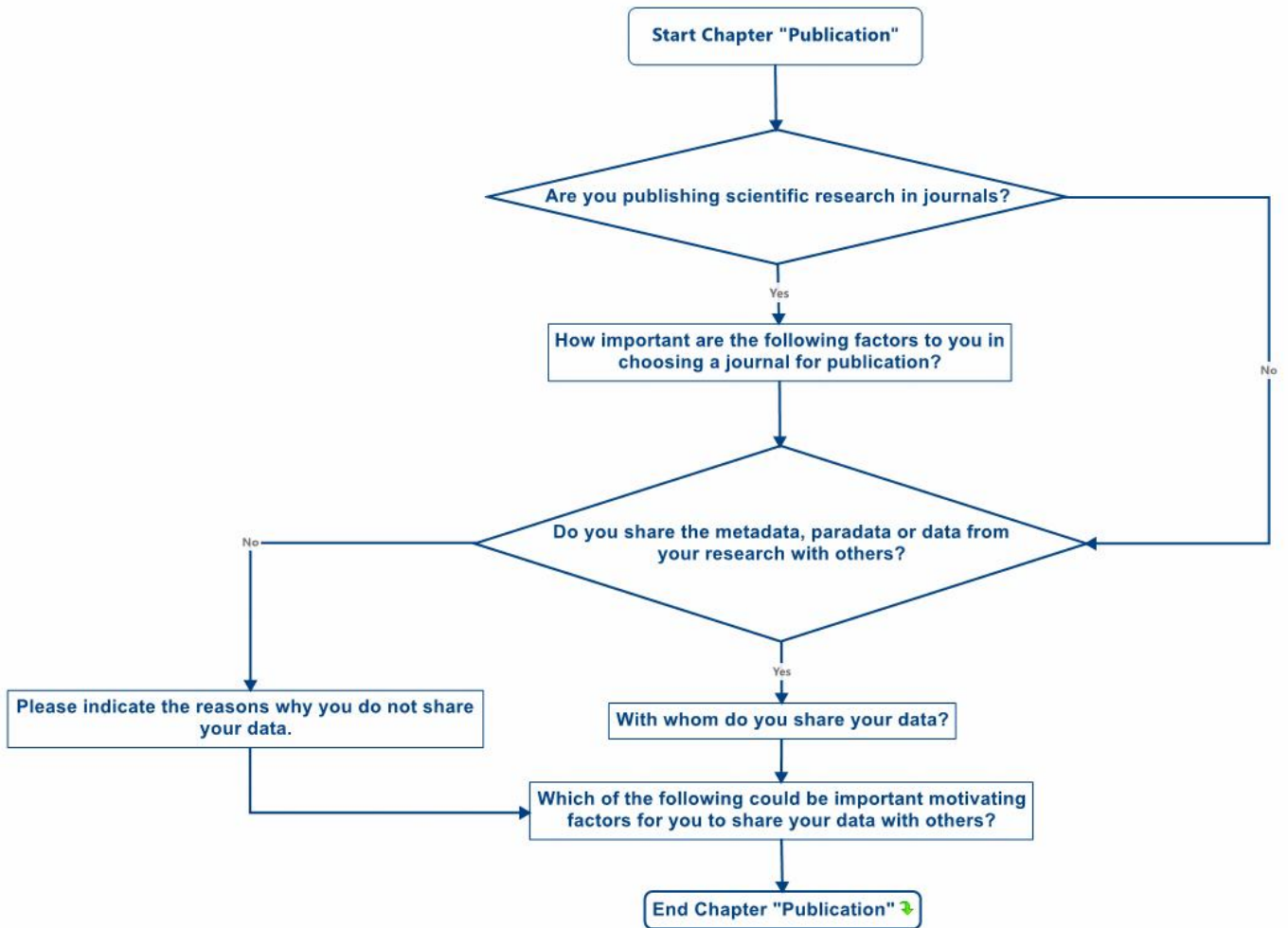
- Chapter 6 – Main Study - “Data Analysis”



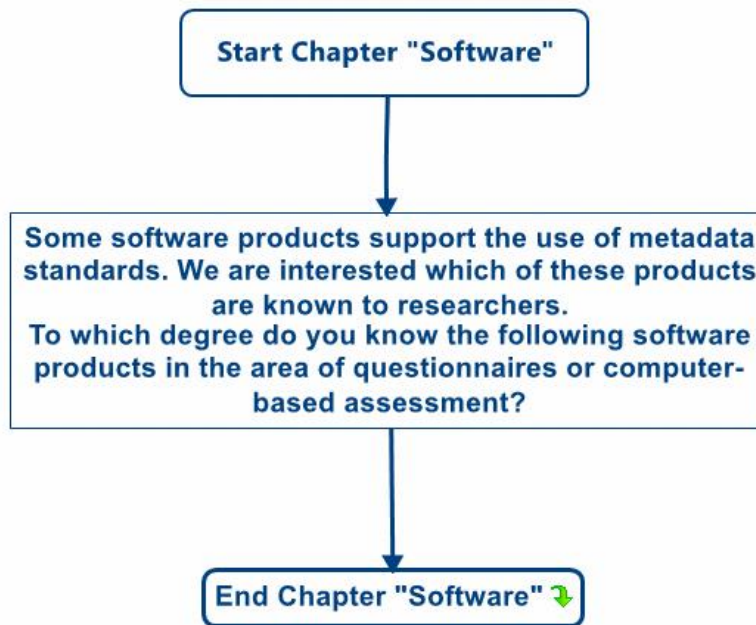
- Chapter 7 – Main Study - “Paradata”



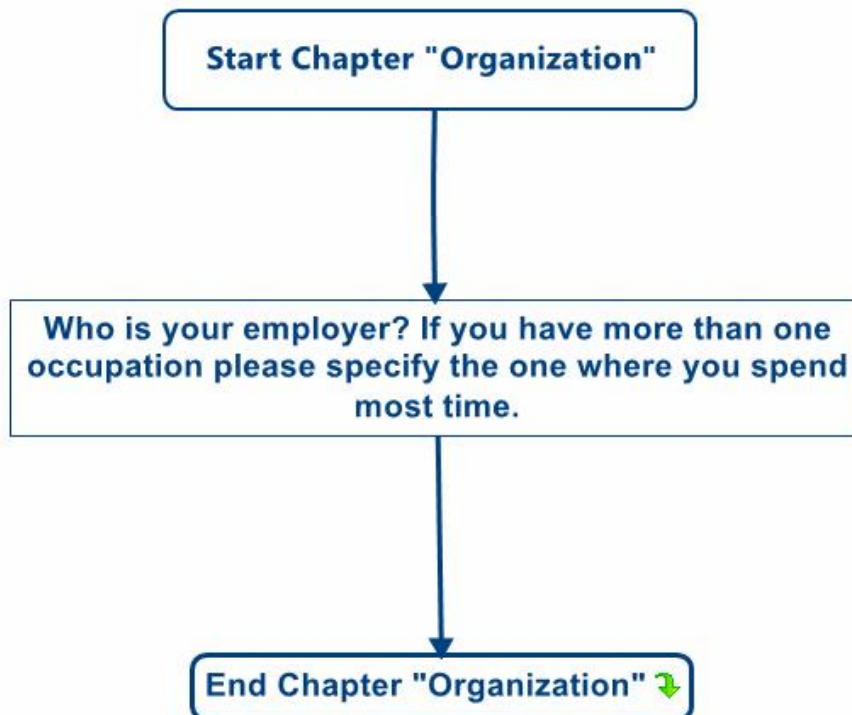
- Chapter 8 – Main Study - “Publication”



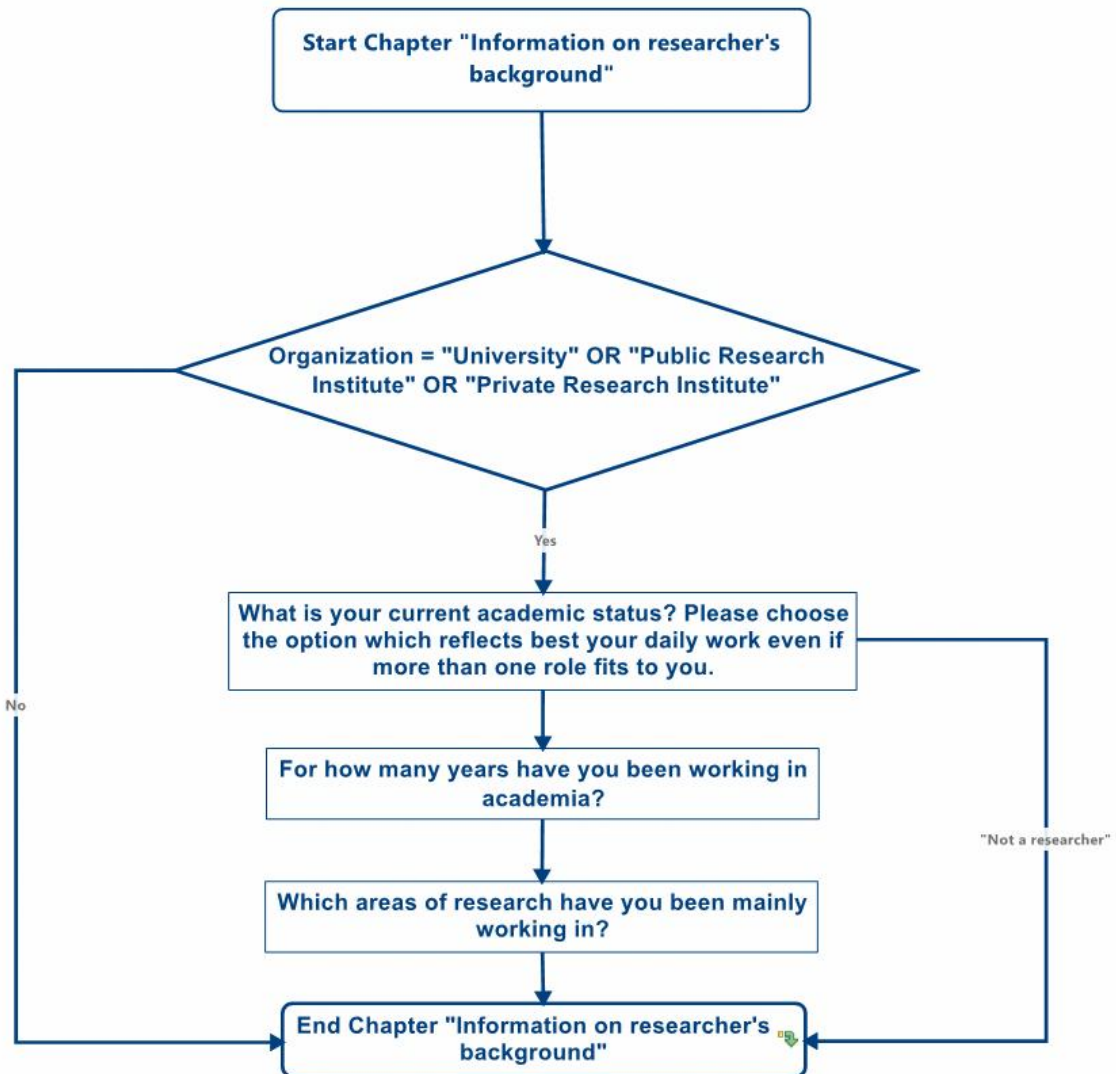
- Chapter 9 – Main Study - “Software”



- Chapter 10 – Main Study - “Organization”

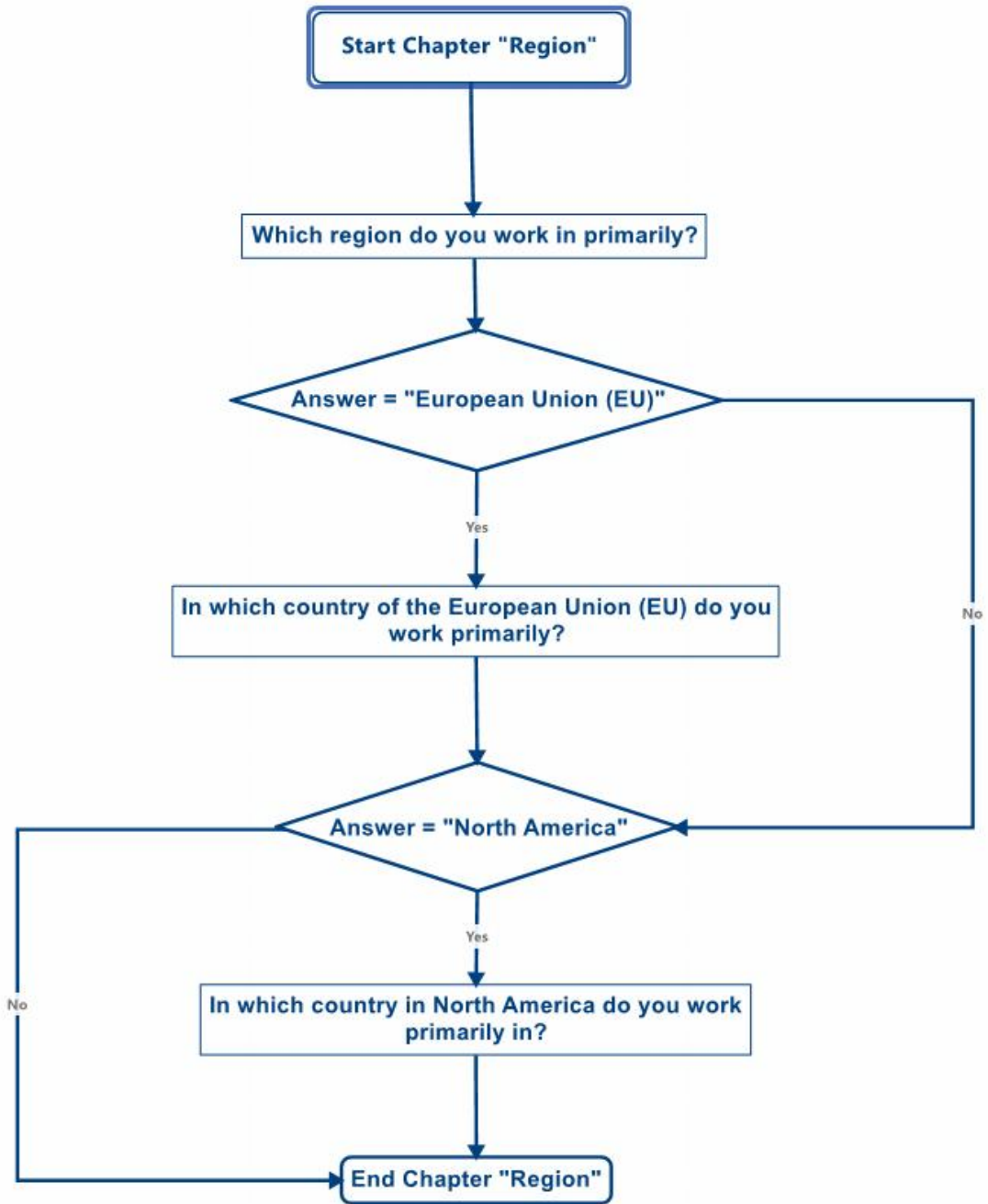


- Chapter 11 – Main Study - “Information on Researcher’s background”

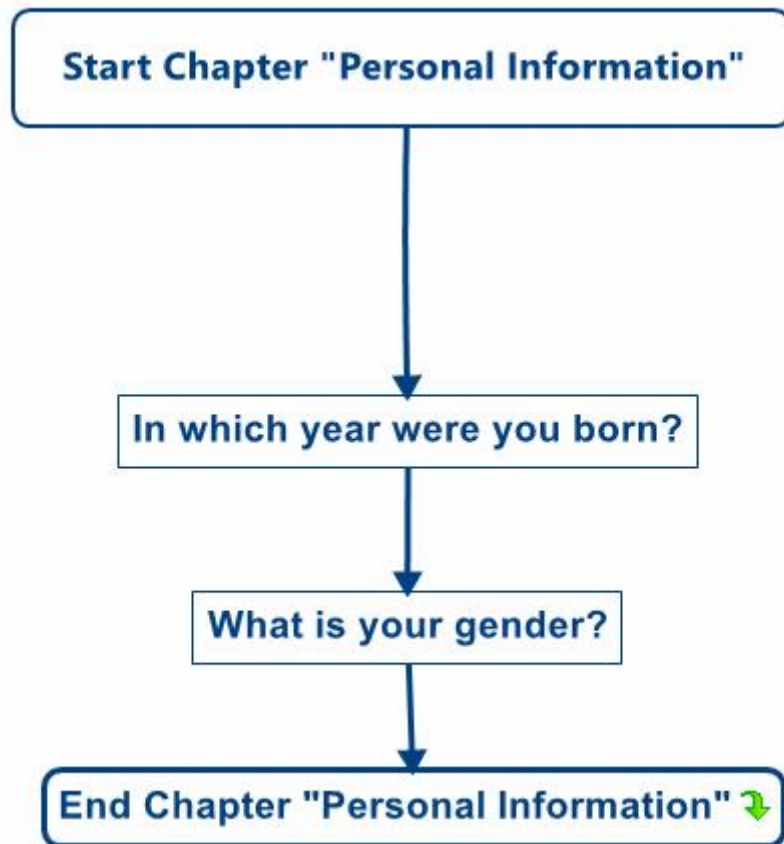




- Chapter 12 – Main Study - “Region”



- Chapter 13 – Main Study - “Personal Information”



### III. Invitation for participation in the study

*“Dear colleagues,*

*\*\*\* Apologies for cross-posting if you received this mail already though another mailing list \*\*\**

*I would like to invite you to participate in a web survey about metadata management in the social and educational sciences and – if possible – promote it in your network. In this study the expectations of researchers who design surveys are compared with the ones of researchers who do secondary analyses of published datasets. The results will be used to improve current metadata standards for both disciplines and we want also to use the findings later as requirements for the further development of open source software products.*

*Target groups of the survey are researchers in the field of social sciences and educational sciences, personnel at data collection or market research organizations, software developers of computer-based questionnaire or assessment software and researchers working at research data centres or data archives.*

*This research is part of my PhD dissertation at the University of Szeged where I currently reside for a research semester and I would personally appreciate it very much if people from your organization take their time to participate in the survey. The average time for filling in the questionnaire is about 15-20 min. The number of questions you will be asked varies between 30-40 questions depending on your answers. You can pause the questionnaire and continue answering it at a later period.*

***The participation in this study is completely voluntarily and your answers will be anonymized.***

*Here is the link to the web survey:*

<http://dmsurvey.limequery.org/index.php/346646>

*The survey is available until December 19th, 2014. For any questions regarding the survey please contact me at [ingo.barkow@gmail.com](mailto:ingo.barkow@gmail.com)*

*Thank you very much in advance!*

*Best regards,  
Ingo*

....

*Ingo Barkow, M.A., M.Sc.  
Doctoral School for the Educational Sciences  
University of Szeged, Hungary”*

#### IV. Mailing lists used for the data collections

Organization	Country	Field	Coverage
Doctoral School for Educational Sciences at the University of Szeged	Hungary	Educational Sciences	30-40 persons
GESIS - the Leibniz Institute for Social Sciences	Germany	Social Sciences	400-500 persons
DIPF – German Institute for International Educational Research	Germany	Educational Sciences	300-400 persons
Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)	Germany	Educational Sciences	Unknown
National Educational Panel Study (NEPS)	Germany	Educational Sciences, Social Sciences, Economics	200-300 persons
DDI Alliance	International	Social Sciences, Educational Sciences, Health Sciences, Statistics	Unknown
International Association for Social Sciences Information Services and Technology (IASSIST)	International	Social Sciences	Unknown
European Survey Research Association (ESRA)	International (mainly Europe)	Social Sciences	Unknown
University of Heidelberg, Institute for Education	Germany	Educational Sciences	Unknown
University of Luxembourg, Institute for Education	Luxembourg	Educational Sciences	Unknown
CentERdata	Netherlands	Software vendor for data management products	Unknown
TNS Infratest	Germany	Data Collection Agency	Unknown
Infas	Germany	Data Collection Agency	Unknown
Institute for Adult Learning (IAL)	Singapore	Educational Sciences	Unknown
Media Research Consultants (MRC)	Singapore	Data Collection Agency	Unknown
Institute for Employment Research (IAB), Nürnberg	Germany	Labour Market Research	Unknown
Institute for Educational Progress, Berlin	Germany	Educational Research	Unknown
Cornell University	United States	Social Sciences	Unknown

Minneapolis Population Center	United States	Social Sciences	Unknown
University of Santiago	Chile	Educational Sciences	Unknown
University of Wisconsin	United States	Social Sciences	Unknown
Swedish National Data Archive (SND)	Sweden	Social Sciences	Unknown
Rand Corporation	United States	Social Sciences	Unknown
Institute of Adult Education, Ljubljana	Slovenia	Educational Sciences	Unknown
University of Duisburg-Essen, Institute for Education	German	Educational Sciences	Unknown

## V. Field Test Questionnaire

To document the changes between the Field Test and the Main Study here the questions used in the pilot version. The order has been modified as according to Couper (2000) demographic information should be asked at the end. In the Field Test version demographic questions came at the beginning. The idea behind this was to use those variables for more filtering during the questionnaire.

### - Organisation

Code	OR01
Filter	None
Question Text	Who is your employer? If you have more than one employer please specify the one where you spend the most of your time.
Response Domain	Option List
Options	<ul style="list-style-type: none"> <li>“University (including Colleges and Technical Universities)”</li> <li>“Public Research Institute”</li> <li>“Private Research Institute”</li> <li>“Data Collection Agency”</li> <li>“Market Research Agency”</li> <li>“Software Development Company”</li> <li>“Other private sector organization”</li> </ul>
Rationale	To identify the work place of the respondent.
Changes	Demographic questions are in the Main Study at the end. “Data Collection Agency” and “Market Research Agency” merged.

Code	OR02
Filter	None
Question Text	Which of the following tasks describes best your day-to-day work within your organization?
Response Domain	Multiple Choice
Options	<ul style="list-style-type: none"> <li>“I am involved in survey design, experimental design or item development.”</li> <li>“I am involved in data collection.”</li> <li>“I am involved in data analysis of own datasets.”</li> <li>“I am involved in data analysis of third-party datasets.”</li> <li>“I am involved in software development.”</li> <li>“I am involved in preparing datasets for publication or archival.”</li> <li>“Other:”</li> </ul>
Rationale	To identify the work area of the respondent.

Changes	Omitted. Replaced by several filter questions.
---------	--

- Information about the researcher's background

Code	RB01
Filter	OR01.NAOK = "OR101" OR OR01.NAOK = "OR102" OR OR01.NAOK = "OR103"  "Who is your employer?" = "University (including Colleges and Technical Universities)" OR "Who is your employer?" = "Public Research Institute" OR "Who is your employer?" = "Private Research Institute"
Question Text	What is your current academic status? Please choose the option which reflects best your daily work even if more than one role fits to you.
Response Domain	Option List
Options	"Student" "Doctoral student / PhD Student" "Researcher (without PhD)" "Postdoctoral Researcher (with PhD)" "Assistant / Associate Professor" "Full Professor" "Not a researcher / not working in academia"
Rationale	To identify the academic status of the respondent. Code RB108 is necessary as there might be personnel working in academic institutions as non-scientists (e.g. administrative personnel).
Changes	Student divided into Bachelor/Master. "Not working in academia" omitted.

Code	RB02
Filter	RB01.NAOK IS NOT "RB108" AND (OR01.NAOK = "OR101" OR OR01.NAOK = "OR102" OR OR01.NAOK = "OR103")  "What is your current academic status?" IS NOT "Not a researcher" AND ("Who is your employer?" = "University (including Colleges and Technical Universities)" OR "Who is your employer?" = "Public Research Institute" OR "Who is your employer?" = "Private Research Institute")
Question Text	For how many years are you working in academia?
Response Domain	Number (Integer)
Rationale	To identify academic working experience.
Changes	Wording.

Code	RB03
Filter	RB01.NAOK IS NOT "RB108" AND (OR01.NAOK = "OR101" OR OR01.NAOK = "OR102" OR OR01.NAOK = "OR103")  "What is your current academic status?" IS NOT "Not a researcher" AND

	(“Who is your employer?” = “University (including Colleges and Technical Universities)” OR “Who is your employer?” = “Public Research Institute” OR “Who is your employer?” = “Private Research Institute”)
Question Text	In which areas of research are you mainly working in?
Response Domain	Multiple Choice
Options	“Agricultural Sciences” “Arts” “Computer Science” “Educational Sciences” “Engineering” “Human medicine, Health Science or Veterinary Sciences” “Law and Economics” “Linguistics and cultural studies” “Medicine” “Social Sciences” “Sports”
Rationale	To identify the academic field. The question was originally an option field so researchers had to choose their main topic, but the field test respondents found this too limiting.
Changes	Wording.

- Region

Code	RE01
Filter	None
Question Text	Which region do you work in primarily?
Response Domain	Option List
Options	“Africa” “Asia” “Australia / Oceania” “Central and South America” “European Union (EU)” “Europe (Non-EU)” “North America”
Rationale	To identify the region of the respondent
Changes	None.

Code	RE02
Filter	RE01.NAOK = “RE105” “Which region do you work in primarily?” = “European Union (EU)”
Question Text	In which country of the European Union (EU) do you work?
Response Domain	Option List
Options	“Austria” “Belgium” “Bulgaria” “Croatia” “Cyprus” “Czech Republic” “Denmark” “Estonia” “Finland”

	“France” “Germany” “Greece” “Hungary” “Ireland” “Italy” “Latvia” “Lithuania” “Luxembourg” “Malta” “Netherlands” “Poland” “Portugal” “Romania” “Slovakia” “Slovenia” “Spain” “Sweden” “United Kingdom”
Rationale	To identify the EU country where the respondent works in.
Changes	Addition of “primarily” in question text.

Code	RE03
Filter	RE01.NAOK = “RE107” “Which region do you work in primarily?” = “North America”
Question Text	In which country in North America do you work?
Response Domain	Option List
Options	“Canada” “United States of America (USA)”
Rationale	To identify the North American country where the respondent works in.
Changes	Addition of “primarily” in question text.

- Research Organization

Code	RO01
Filter	None
Question Text	Does your organization produce research data or conduct surveys?
Response Domain	Yes / No
Rationale	Separation of organizations involved in survey design from others thus shortening the questionnaire
Changes	Dimension changed from organization to individual researcher.

Code	RO02
Filter	RO01.NAOK = “Yes” “Are you involved in survey design, experimental design or instrument development?” = “Yes”
Question Text	Does your organization to your knowledge use data management plans when conducting research?
Response Domain	Option List
Options	Scale from “Always” to “Never” plus “Do not know”
Rationale	How familiar are researchers with data management plans



Changes	Dimension changed from organization to individual researcher.
---------	---

Code	RO03
Filter	RO01.NAOK = "Yes" "Are you involved in survey design, experimental design or instrument development?" = "Yes"
Question Text	How do you or how does your employer handle research data from research projects?
Response Domain	Option List
Choices	"Research data is stored in own research data facility or archive." "Research data is stored in third party research data facility or archive." "Research data stays with the Prime Investigator (PI) or project team and is available on request." "Research data stays with the Prime Investigator (PI) or project team, but cannot be forwarded to third parties." "Research data stays on the IT infrastructure of the project, but the handling afterwards is unknown." "Do not know"
Rationale	How often do researchers use research data centres
Changes	Option list was modified to multiple choice. Dimensions "Organization" and "Researcher" were reduced only to the researcher. Answers options "Research data stays with the Prime Investigator (PI) or project team and is available on request." and "Research data stays with the Prime Investigator (PI) or project team, but cannot be forwarded to third parties." were simplified.

Code	RO04
Filter	RO01.NAOK = "Yes" "Are you involved in survey design, experimental design or instrument development?" = "Yes"
Question Text	How many months lasts a typical research project in your organization?
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey
Changes	Wording and additional explanation.

Code	RO05
Filter	RO01.NAOK = "Yes" "Are you involved in survey design, experimental design or instrument development?" = "Yes"
Question Text	How much time in percent do you or does your organization use for documenting data in an average research project?
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey
Changes	Dimensions "Organization" and "Researcher" reduced to the researcher perspective.

Code	RO06
Filter	RO01.NAOK = "Yes" "Are you involved in survey design, experimental design or instrument development?" = "Yes"

Question Text	How many persons in average are involved in a typical research project in your organization?
Response Domain	Number (Integer)
Rationale	Used for calculation of time for data management in a whole survey
Changes	Order of questions (used to be RO05). Dimension of "Organization" was removed. In hindsight this proved as a wrong decision as some participants in the Main Study started calculating the respondents of the survey into the answer. In further rounds of the questionnaire the Main Study version has to be fixed.

- Metadata Standards

Code	MS01
Filter	None
Question Text	Did you ever use metadata standards in your research or does your organization work with metadata standards?
Response Domain	Yes / No
Rationale	Separation of organizations using metadata standards from others thus shortening the questionnaire
Changes	Omitted. Participants criticized they have no idea if they know metadata standards without seeing the list from the next question.

Code	MS02
Filter	MS01.NAOK = "Yes" "Did you ever use metadata standards in your research or does your organization work with metadata standards?" = "Yes"
Question Text	Which metadata standards do you know?
Response Domain	Matrix
Options	"I use this metadata standard very often." "I use this metadata standard often." "I use this metadata standard rarely." "I have never used this metadata standard." "I do not know this metadata standard."
Sub Questions	"Data Documentation Initiative (DDI)" "Dublin Core" "Learning Object Model (LOM)" "Learning Tools Interoperability (LTI)" "Machine Readable Catalogueing (MARC)" "Metadata Encoding and Transmission Standard (METS)" "QueDex" "Question and Test Interoperability (QTI)" "Sharable Content Object Reference Model (SCORM)" "Statistical Data and Metadata Exchange (SDMX)" "Text Encoding Initiative (TEI)"
Rationale	Which metadata standards are known to the respondents?
Changes	Filter removed. Wording of question improved. "I use this metadata standard very often" omitted.

Code	MS03
Filter	MS01.NAOK = "Yes" "Did you ever use metadata standards in your research or does your

	organization work with metadata standards?" = "Yes"
Question Text	Which of these metadata standards have been or are to your knowledge used in your organization?
Response Domain	Matrix
Options	"My organization uses this standard" "My organization used to use this standard but replaced it" "My organization tested this standard but never used it" "My organization does not use this standard" "Do not know"
Sub Questions	"Data Documentation Initiative (DDI)" "Dublin Core" "Learning Object Model (LOM)" "Learning Tools Interoperability (LTI)" "Machine Readable Catalogueing (MARC)" "Metadata Encoding and Transmission Standard (METS)" "QueDex" "Question and Test Interoperability (QTI)" "Sharable Content Object Reference Model (SCORM)" "Statistical Data and Metadata Exchange (SDMX)" "Text Encoding Initiative (TEI)"
Rationale	Which metadata standards are used by organizations?
Changes	Fully omitted as the dimension of "Organization" was removed.

Code	MS04
Filter	None
Question Text	How useful do you think is the usage of standardized metadata for research projects?
Response Domain	Option List
Options	Scale from "Very Useful" to "Not useful" plus "Do not know"
Rationale	Measuring satisfaction about metadata standards in general.
Changes	Omitted. Answer was embedded into Matrix question MS06.

Code	MS05
Filter	MS01.NAOK = "No" "Did you ever use metadata standards in your research or does your organization work with metadata standards?" = "No"
Question Text	Which are the reasons you or your organization do not use metadata standards?
Response Domain	Matrix
Options	Scale from "Strongly agree" to "Strongly disagree" plus "Do not Know"
Sub Questions	"We were not able to identify the fitting standards for our needs." "The current metadata standards in our discipline do not cover our needs." "We did not have time to identify a fitting metadata standard." "There are to our knowledge no metadata standards in our scientific area." "Our studies are too complex to use metadata standards." "There is no benefit in using metadata standards." "We are bound to use an internal standard and thus cannot use an external standard." "I am not involved in data documentation"
Rationale	Identify reasons for not using metadata standards
Changes	Omitted. Question has been replaced by MS06 in the Main Study which mixed positive and negative statements which are asked to all participants.

- Data

Code	DA01
Filter	None
Question Text	How often do you use the following data sources for your research?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Structured cognitive tests (e.g. computer-based assessments)” “Structured questionnaires (e.g. multiple choice)” “Questionnaires with mostly open questions (e.g. narrative interviews, oral history)” “Experimental designs” “Observational data (e.g. recordings, transcripts)”
Rationale	Which kinds of data sources are the respondents using for their research?
Changes	Wording

- Study Unit

Code	SU01
Filter	None
Question Text	Please identify the five categories which you think are most important if a study, project or survey has to be described and rank them on importance. All your answers must be different. Please select at most 5 answers
Response Domain	Ranking
Options	Ranks from 1 to 5
Sub Questions	“Concepts of the study” “Infrastructure” “Funding Sources” “Previous publications” “Research team” “Sampling methods” “Target group / Universe” “Analysis Plan” “Data cleaning methods” “Organizations involved” “Quality control mechanisms” “Structure of the study”
Rationale	These are general terms derived from metadata standards (e.g. DDI) or models (e.g. GLBPM) describing the first phase of the survey cycle. If researchers have resource issues (time or money) and thus cannot invest much time into data documentation it would be important to know what they consider to be valuable information.
Changes	More explanations in the answers. Question was changed from ranking to matrix due to complaints in the Field Test as the question was complicated and time-consuming.

- Items

Code	IT01
Filter	None
Question Text	Are you involved in the creation of items for questionnaire or assessment

	instruments in your institution?
Response Domain	Yes / No
Rationale	Separation of item designers from the rest of the respondents.
Changes	Wording of question text.

Code	IT02
Filter	IT01.NAOK = “Yes” “Are you involved in the creation of items for questionnaires or cognitive tests in your institution?” = “Yes”
Question Text	What metadata do you or your organization usually provide with a questionnaire or assessment item?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Analytic scripts” “Answer categories” “Concept” “Flow logic (e.g. routing, filtering, dynamic text)” “Fragments of software (e.g. code which makes the item reusable)” “Interviewer instructions” “Layout definitions” “References to research literature” “Scoring rules” “Statistical values from previous studies” “Stimuli in reusable form” “Variable definitions”
Rationale	These are general terms derived from metadata standards (e.g. DDI or QTI). As the filter questions before only lets respondents through who design items the question is which elements they deliver with an item if they would be considered to be data producers. This can be compared with the expectations of data users in question IT12. The question was asked the same way as matrix question in the field test, but it contained less examples.
Changes	Wording of question text.

Code	IT11
Filter	None
Question Text	Have you been re-using or are planning to re-use items or questions developed by other institutions?
Response Domain	Yes / No
Rationale	Separation of secondary item users from the rest of the respondents.
Changes	Wording of question text – “persons” added.

Code	IT12
Filter	IT11.NAOK = “Yes” “Have you been re-using or are planning to re-use items or questions developed by other institutions?” = “Yes”
Question Text	Please rank the five most important metadata for an item in a questionnaire or assessment which are most relevant for preservation or re-use in later studies?
Response Domain	Ranking
Options	Ranks from 1 to 5
Sub Questions	“Analytic scripts” “Answer categories”

	<p>“Concept”</p> <p>“Flow logic (e.g. routing, filtering, dynamic text)”</p> <p>“Fragments of software (e.g. code which makes the item reusable)”</p> <p>“Interviewer instructions”</p> <p>“Layout definitions”</p> <p>“References to research literature”</p> <p>“Scoring rules”</p> <p>“Statistical values from previous studies”</p> <p>“Stimuli in reusable form”</p> <p>“Variable definitions”</p>
Rationale	<p>These are general terms derived from metadata standards (e.g. DDI or QTI). As the filter questions before only lets respondents through who design items the question is which elements they deliver with an item if they would be considered to be data producers. This can be compared with the expectations of data users in question IT12. The question was asked the same way as matrix question in the field test, but it contained less examples.</p>
Changes	<p>Wording of question text. Question was changed from ranking to matrix due to complaints in the Field Test as the question was complicated and time-consuming. The original idea was to offer a similar question in two different response domains if a respondent has to go through both branches to make the study less boring. Nevertheless the ranking versions proved to be unpopular.</p>

- Data Collection

Code	DC11
Filter	None
Question Text	Do you analyze data which is provided by an internal or external data collection agency or do secondary analysis of datasets which were provided by a research data center or archive?
Response Domain	Yes / No
Rationale	Separation of secondary data users from the rest of the respondents.
Changes	Wording of the question text.

Code	DC12
Filter	DC11.NAOK = “Yes” “Do you analyze data which is provided by an internal or external data collection agency or do secondary analysis of datasets which were provided by a research data center or archive?” = “Yes”
Question Text	Which metadata or paradata (i.e. metadata derived from data collection processes) information does the data collection agency or archive normally deliver to you?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	<p>“Codebook with variable definitions”</p> <p>“Log files from each collected case”</p> <p>“Processed and cleaned data files”</p> <p>“Quality control reports”</p> <p>“Raw data files”</p> <p>“Report about happenings or incidents during the data collection process”</p> <p>“Textual descriptions about data cleaning processes”</p>

Rationale	These are files which are produced by data collection agencies during field work. The question is which of them are forwarded to the researcher to support the data analysis process.
Changes	Wording of the question text. More explanations in the answers.

Code	DC13
Filter	DC11.NAOK = “Yes” “Do you analyze data which is provided by an internal or external data collection agency or do secondary analysis of datasets which were provided by a research data center or archive?” = “Yes”
Question Text	Are you satisfied with the metadata and paradata information you receive from data collection agencies or data archives??
Response Domain	Option List
Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the metadata delivered by data collection agencies.
Changes	Wording of the question text.

Code	DC21
Filter	None
Question Text	Do you conduct data collection in your organization and prepare the results for internal or external parties?
Response Domain	Yes / No
Rationale	Separation of data producers from the rest of the respondents.
Changes	Wording of the question text.

Code	DC22
Filter	DC21.NAOK = “Yes” “Do you conduct data collection in your organization and prepare the results for internal or external parties?” = “Yes”
Question Text	Are your customers mostly internal or external?
Response Domain	Option List
Options	Scale from “Only external” to “Only internal”
Rationale	To distinguish if the data collection happens for research teams in the own facility or rather for third parties (the latter would be a sign for a commercial data collection or market research agency)
Changes	Wording of the question text.

Code	DC23
Filter	DC21.NAOK = “Yes” “Do you conduct data collection in your organization and prepare the results for internal or external parties?” = “Yes”
Question Text	Which metadata or paradata (i.e. metadata derived from data collection processes) information do you or your organization usually deliver to your clients?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Codebook with variable definitions” “Log files from each collected case” “Processed and cleaned data files”

	“Quality control reports” “Raw data files” “Report about happenings or incidents during the data collection process” “Textual descriptions about data cleaning processes”
Rationale	This question mirrors DC14 for data collection agencies by asking which additional files are delivered to the research teams for data analysis.
Changes	Wording of the question text. More explanations in the sub questions.

- Dissemination

Code	DI01
Filter	None
Question Text	Are you involved in disseminating datasets within your organization?
Response Domain	Yes / No
Rationale	Separation of data analysts from the rest of the respondents.
Changes	Omitted due to focus on log file analysis.

Code	DI02
Filter	DI01.NAOK = “Yes” “Are you involved in disseminating datasets within your organization?” = “Yes”
Question Text	Are your customers mostly internal or external?
Response Domain	Option List
Options	Scale from “Only external” to “Only internal”
Rationale	To distinguish if the data analysis happens for research teams in the own facility or rather for third parties.
Changes	Omitted due to focus on log files analysis.

Code	DI03
Filter	None
Question Text	Do you analyze paradata (data derived from the data collection process) from log files collected during the survey (e.g. timestamps)?
Response Domain	Yes / No
Rationale	Identification of researchers who disseminate paradata
Changes	Wording of question text.

Code	DI04
Filter	DI03.NAOK = “Yes” “Do you analyze paradata (data derived from the data collection process) from log files collected during the survey (e.g. timestamps)?” = “Yes”
Question Text	When you analyze paradata from log files, which formats do you prefer?
Response Domain	Multiple Choice
Options	“CSV file” “Excel file” “Mplus file” “R file” “Relational Database” “SAS file” “SPSS file” “Stata file”



	“TXT file” “Word document” “XML file”
Rationale	Log files come in different recording formats. This question tries to identify the preference of researchers.
Changes	None.

Code	DI05
Filter	DI03.NAOK = “Yes” “Do you analyze paradata (data derived from the data collection process) from log files collected during the survey (e.g. timestamps)?” = “Yes”
Question Text	How satisfied in general are you with the documentation of paradata from log files you use for your research?
Response Domain	Option List
Options	Scale from “Very satisfied” to “Very dissatisfied” plus “Do not know”
Rationale	Measuring satisfaction about the paradata documentation.
Changes	Wording of the question text.

Code	DI06
Filter	DI01.NAOK = “Yes” “Are you involved in disseminating datasets within your organization?” = “Yes”
Question Text	Which of the following dissemination processes do you use?
Response Domain	Matrix
Options	Scale from “Always” to “Never” plus “Do not know”
Sub Questions	“Analysis of Public Use Files (highly anonymized datasets)” “Analysis of Scientific Use Files (moderately anonymized datasets)” “Analysis of raw datasets (non-anonymized datasets)” “Analysis of datasets stored within a Virtual Research Environment” “Analysis of datasets stored within a data enclave or secure data service” “Analysis of data stored in a relational database” “Analysis of data stored in a data warehouse or analytical database” “Analysis of data using remote calculation services (e.g. job submission)” “Analysis of personal data extracts created by portal websites (e.g. variable shopping basket)” “Analysis of log files created during the data collection process”
Rationale	Research data centres offer different ways of accessing the data for dissemination.
Changes	Question was moved to other chapter. Dimension of “Analysis” was removed. Question was too complex for target group.

- Publication

Code	PU01
Filter	None
Question Text	Are you publishing results from your own or third party research?
Response Domain	Yes / No
Rationale	Separation of researchers from the rest of the respondents.
Changes	Wording of question text.

Code	PU02
Filter	PU01.NAOK = “Yes” “Are you publishing results from your own or third party research?” = “Yes”
Question Text	Please rank the three most important motivations for you when you publish an article.
Response Domain	Ranking
Options	Ranks from 1 to 3
Sub Questions	“Costs” “Fast publishing process” “Impact factor” “Linkage to results / data” “Open access” “Reputation of the journal”
Rationale	From a data documentation point of view this question tries to analyse how important data linkage is in comparison to typical quality factors of journals.
Changes	Wording of question text. Changed from ranking to matrix.

Code	PU03
Filter	None
Question Text	Do you share the metadata, paradata or data from your research?
Response Domain	Yes / No
Rationale	Separation between researchers who share their data and researchers who do not share their data.
Changes	Wording of question text

Code	PU04
Filter	PU03.NAOK = “Yes” “Do you share the metadata, paradata or data from your research?” = “Yes”
Question Text	With whom do you share your data?
Response Domain	Multiple Choice
Options	“I deliver the data to a research data centre or archive.” “I hand over the data to a designated person at my organization.” “I share my data only with researchers I know.” “I share my data with other researchers from the same field.” “I share my data with other researchers from other fields.” “I share my data only with researchers from my organization.” “I share my data with other persons who are interested in my research.”
Rationale	These choices indicate how far researchers are willing to spread their data to third parties. If somebody says “I share my data only with researchers I know.” forwarding the data to a RDC and providing enhanced publications is not possible though the person in PU03 claims they are sharing their data.
Changes	None.

Code	PU05
Filter	PU03.NAOK = “No” “Do you share the metadata, paradata or data from your research?” = “No”
Question Text	Please rank the top three reasons why you do not share your data.
Response Domain	Ranking
Options	Ranks from 1 to 3
Sub Questions	“I do not know where to hand in my data for archiving.” “I have no resources or support from my employer to hand in the data.”

	<p>“Preparation of the data for archiving needs too much effort.”</p> <p>“My organization or I have security concerns.”</p> <p>“Somebody could misinterpret my results.”</p> <p>“Somebody could use my data before me.”</p>
Rationale	This question is for researchers who indicated in PU03 they are not willing to share their data and gives them reasons to evaluate why they do not do it.
Changes	Wording of question text. Changed from ranking to matrix.

Code	PU06
Filter	PU03.NAOK = “Yes” “Do you share the metadata, paradata or data from your research? = “Yes”
Question Text	Which of the following do you think could be the three most motivating factors for you or others to publish the data according to their research?
Response Domain	Ranking
Options	Ranks from 1 to 3
Sub Questions	<p>“Archive informs data producer which users accessed the data.”</p> <p>“Archive publically provides a list of publications based on the data.”</p> <p>“Co-authorship in secondary publications based on the data”</p> <p>“Data is cited in secondary publications”</p> <p>“Support by the employer (e.g. financial resources, time, staff)”</p> <p>“The data producer receives a financial compensation”</p>
Rationale	Identify motivating factors for data sharing
Changes	Wording of question text. Changed from ranking to matrix. Filter changes to include all participants.

- Software

Code	SO01
Filter	None
Question Text	Do you use survey or assessment software to deliver questionnaires or assessments?
Response Domain	Yes / No
Rationale	Separation of organizations using assessment software from others thus shortening the questionnaire
Changes	Omitted. Participants criticized they have no idea if they know the software products without seeing the list from the next question.

Code	SO02
Filter	SO01.NAOK = “Yes” “Do you use survey or assessment software to deliver questionnaires or assessments?” = “Yes”
Question Text	Which software products in the area of questionnaires or computer-based assessment do you know?
Response Domain	Matrix
Options	<p>“I know this software very well.”</p> <p>“I know this software well.”</p> <p>“I have a basic understanding of what this software does.”</p> <p>“I have only heard of this software, but have never seen it.”</p> <p>“I do not know this software.”</p>
Sub Questions	<p>“Blaise”</p> <p>“CASES”</p>

	"CBA Itembuilder" "eDia" "Globalpark / Unipark" "Limesurvey" "MyItemwriter" "Nipo / Odin / Compass" "Questify" "Questionmark" "Rogatus Survey / Repository" "Surveymonkey" "Testdrive" "Testing Assistee par Ordinateur (TAO)" "Voxco"
Rationale	This question contains a mix of different software products from survey work and it tries to identify which products are known. Survey modes (CAPI, CATI, P&P) as well as research area (educational sciences, social sciences) are mixed here as this software products are in reality also used in multiple domains (e.g. TAO for questionnaires, Blaise for simple assessments) therefore the line cannot be easily drawn.
Changes	Wording of question text. Option "I know this software very well" omitted.

Code	SO03
Filter	SO01.NAOK = "Yes" "Do you use survey or assessment software to deliver questionnaires or assessments?" = "Yes"
Question Text	Which of those software tools to your knowledge have been or are used in your organization?
Response Domain	Matrix
Options	"My organization uses this software" "My organization used to use this software but replaced it" "My organization tested this software but never used it" "My organization does not use this software" "Do not know"
Sub Questions	"Blaise" "CASES" "CBA Itembuilder" "eDia" "Globalpark / Unipark" "Limesurvey" "MyItemwriter" "Nipo / Odin / Compass" "Questify" "Questionmark" "Rogatus Survey / Repository" "Surveymonkey" "Testdrive" "Testing Assistee par Ordinateur (TAO)" "Voxco"
Rationale	Identify the usage of different software products within an organization.
Changes	Omitted due to removal of the "Organization" dimension.

- Personal Information

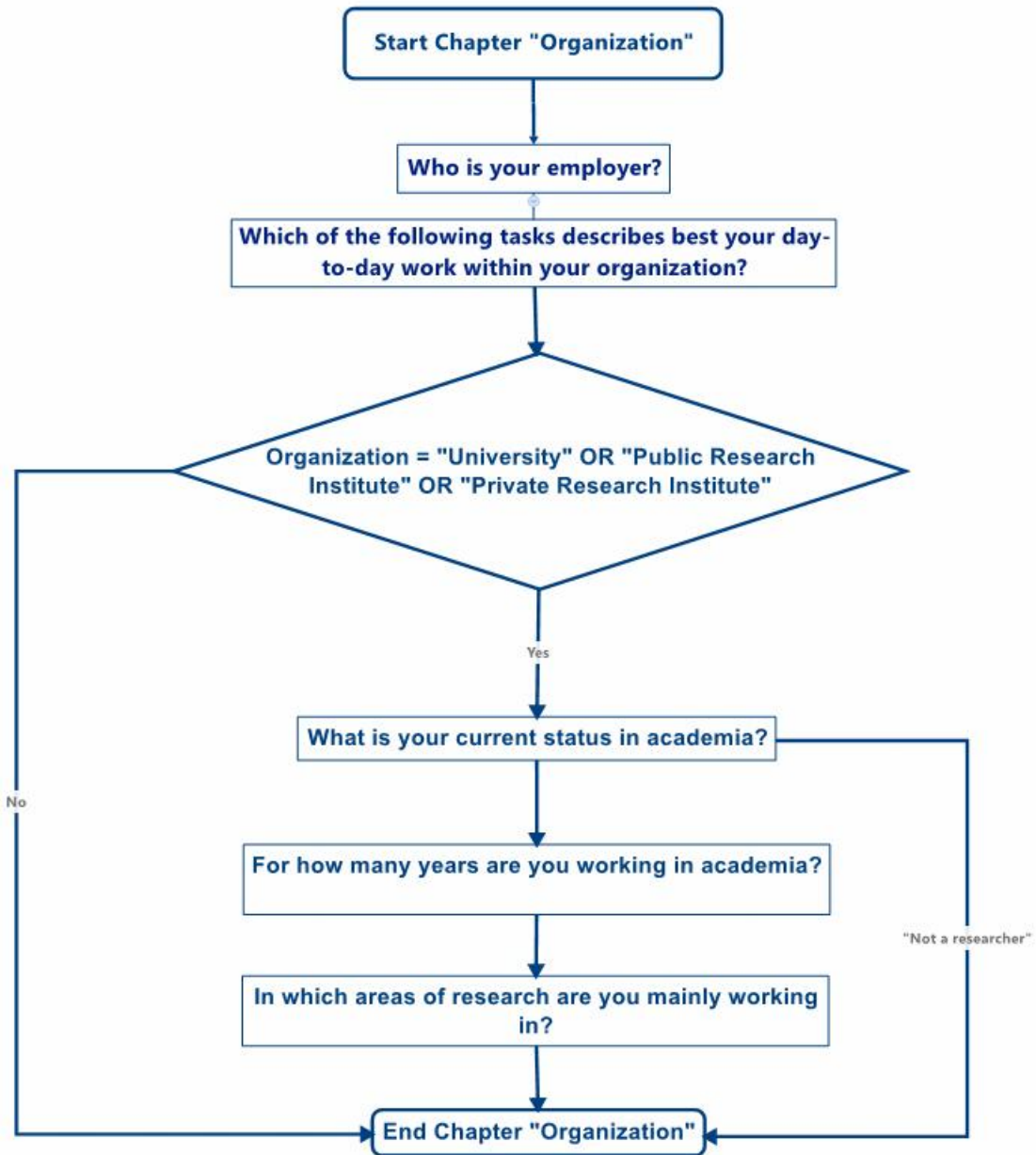
Code	PI01
Filter	None
Question Text	Please specify the year you were born
Response Domain	Number (Integer)
Rationale	To identify the respondents' age.
Changes	Wording of question text

Code	PI02
Filter	None
Question Text	Please specify your gender
Response Domain	Gender
Rationale	To identify the respondents' gender. The survey uses the standard scale of Limesurvey which uses "male" and "female" as these can be considered sufficient for this sample size (the use of "transgender" and other options have no analysis potential add due to very small numbers to be expected). As all questions are optional respondents can skip this question so there should not be a moral dilemma for persons who do not feel themselves comfortable with only two options.
Changes	Wording of question text.

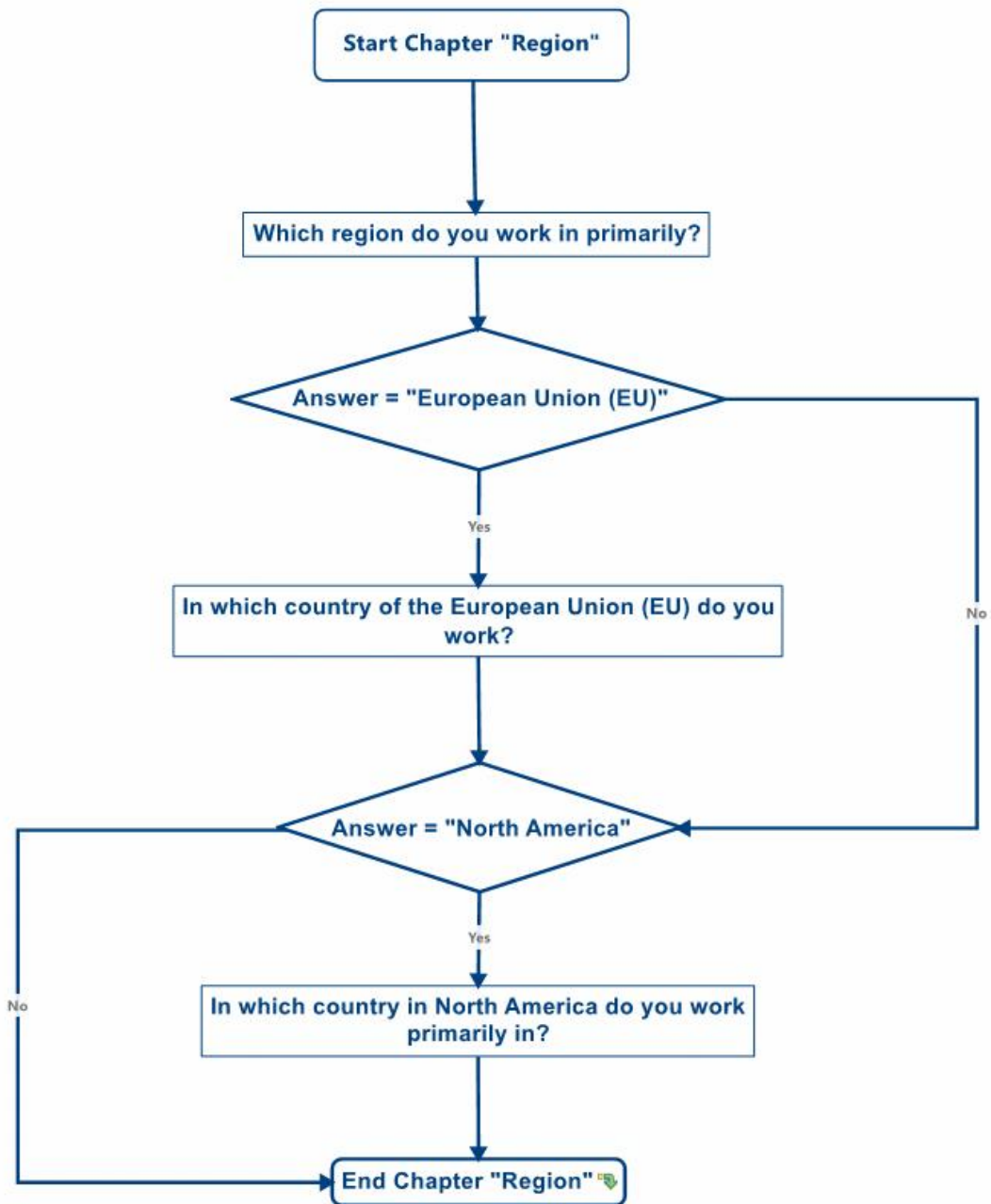
**VI. Field Test Questionnaire Flow Logic**

Similar to Appendix I. where the flow logic of the Main Study questionnaire was shown here the same approach for the Field Test questionnaire which differs in several respects.

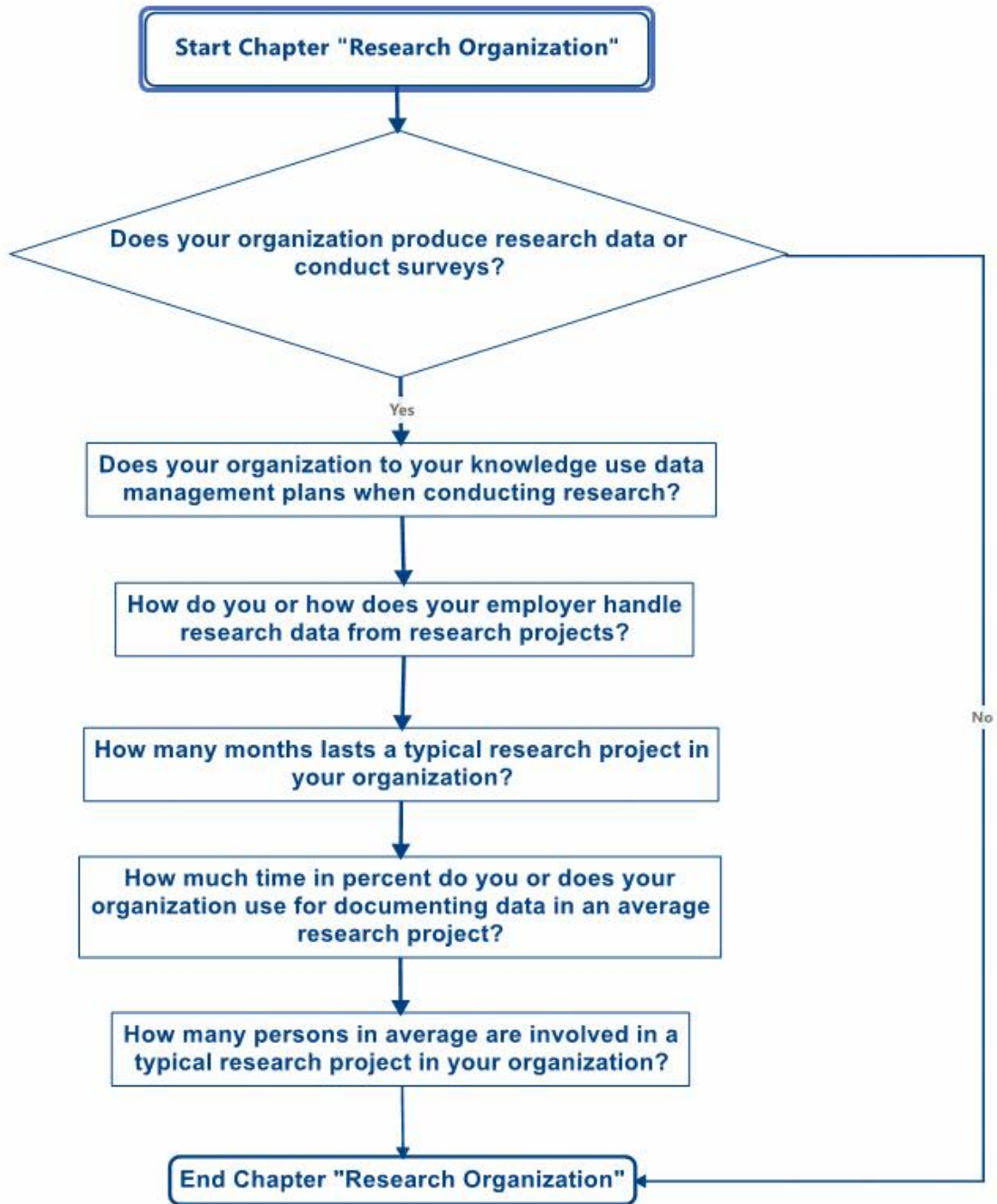
- Chapter 1 – Field Test - "Organization"



- Chapter 2 – Field Test - “Region”

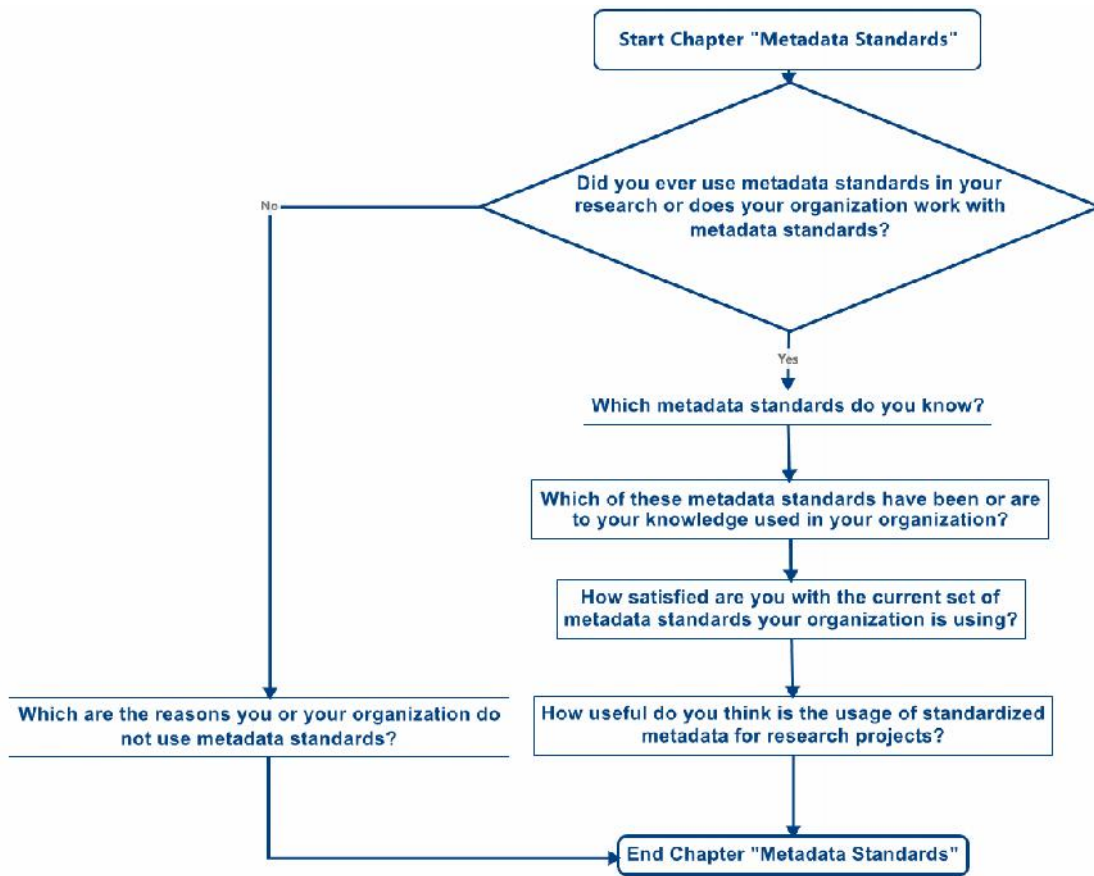


- Chapter 3 – Field Test - “Research Organization”

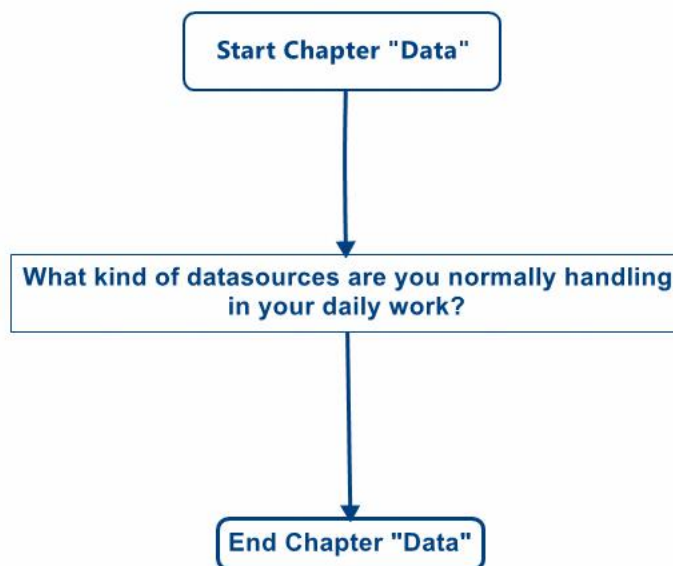


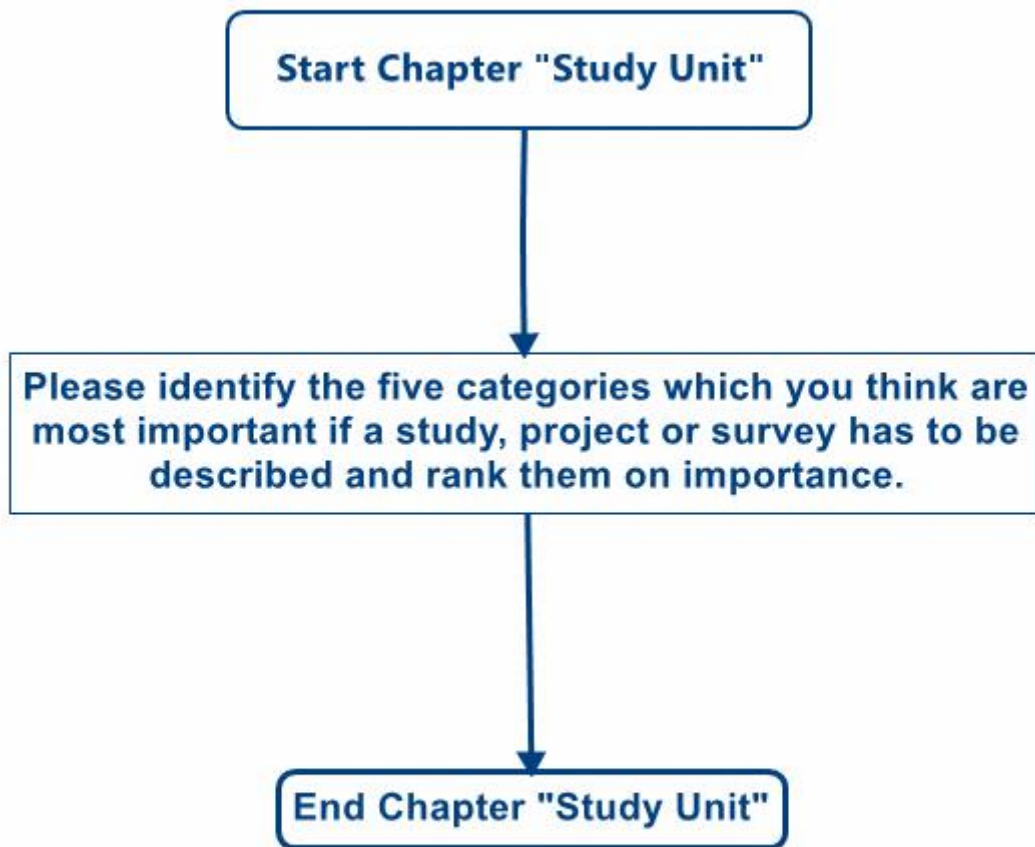


- Chapter 4 – Field Test – Metadata Standards

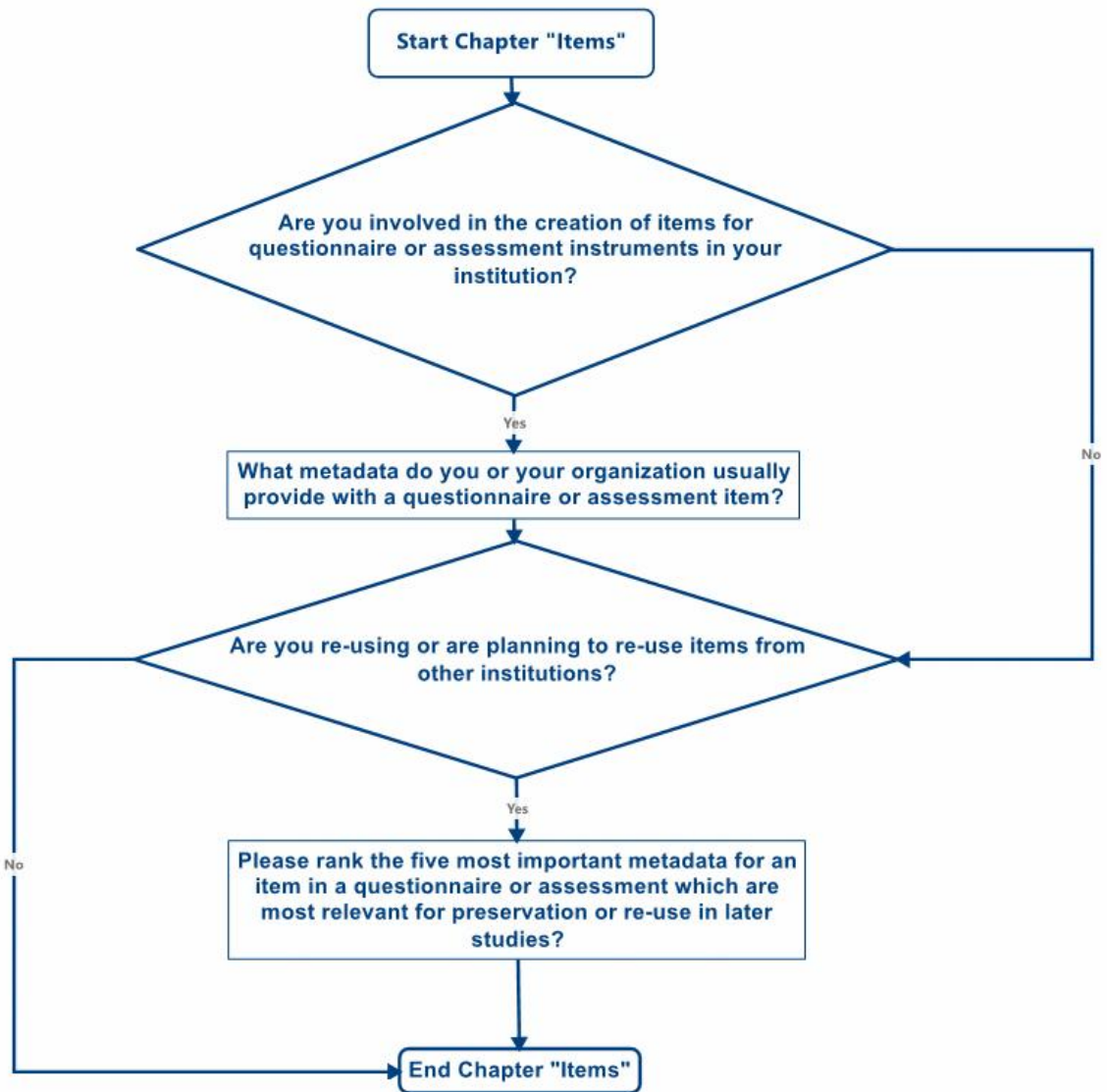


- Chapter 5 – Field Test – Data

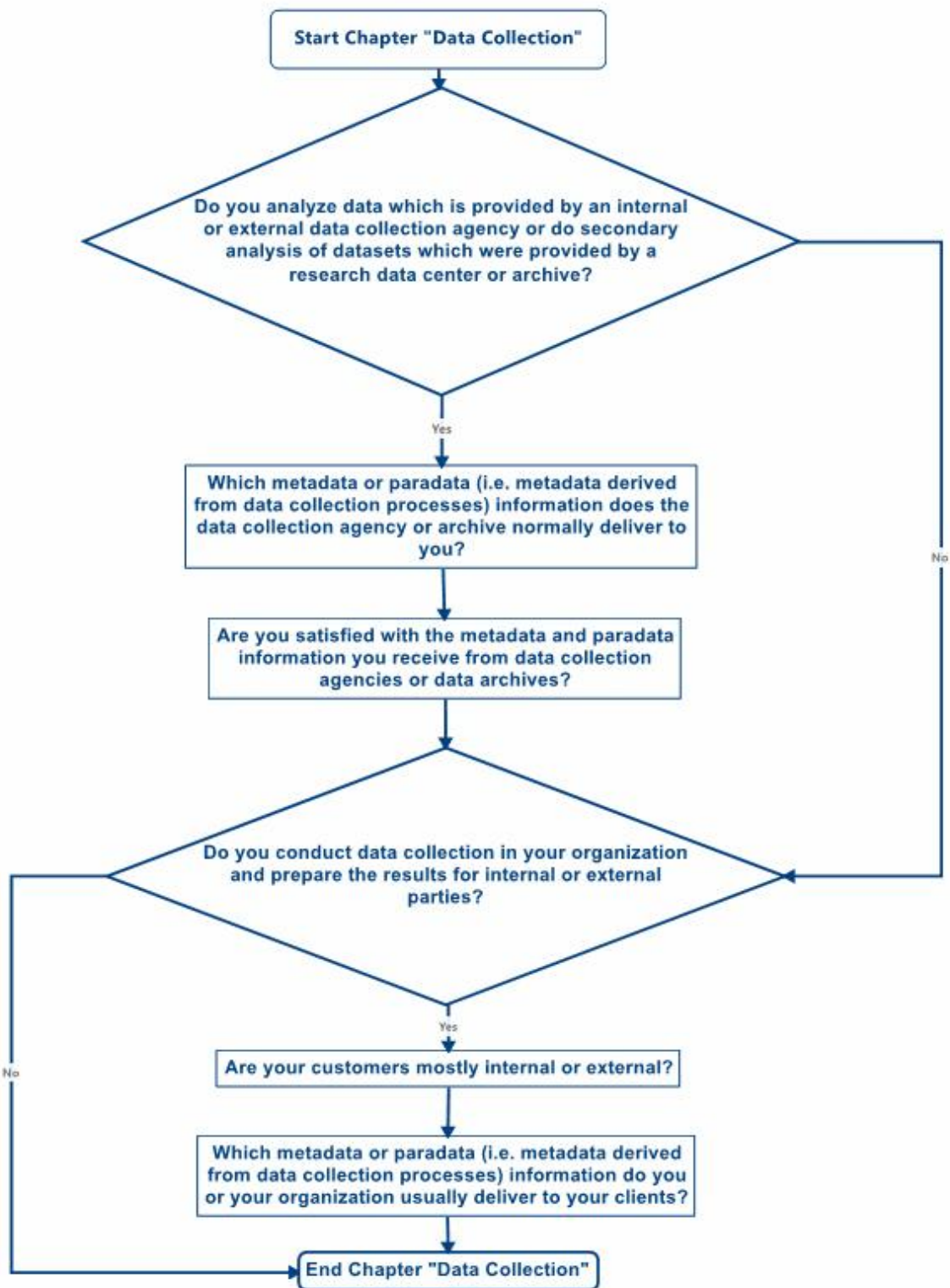




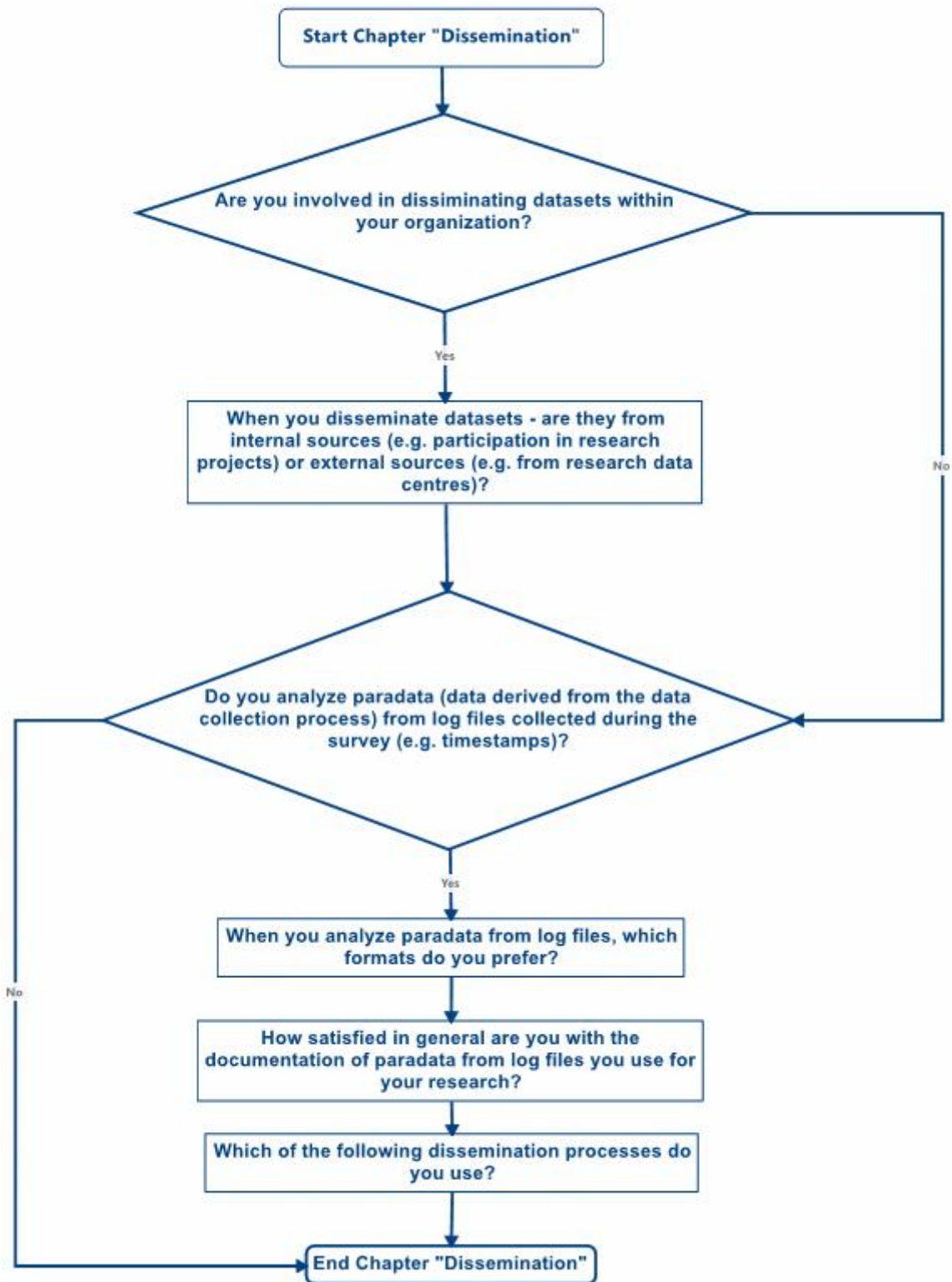
- Chapter 7 – Field Test – Items

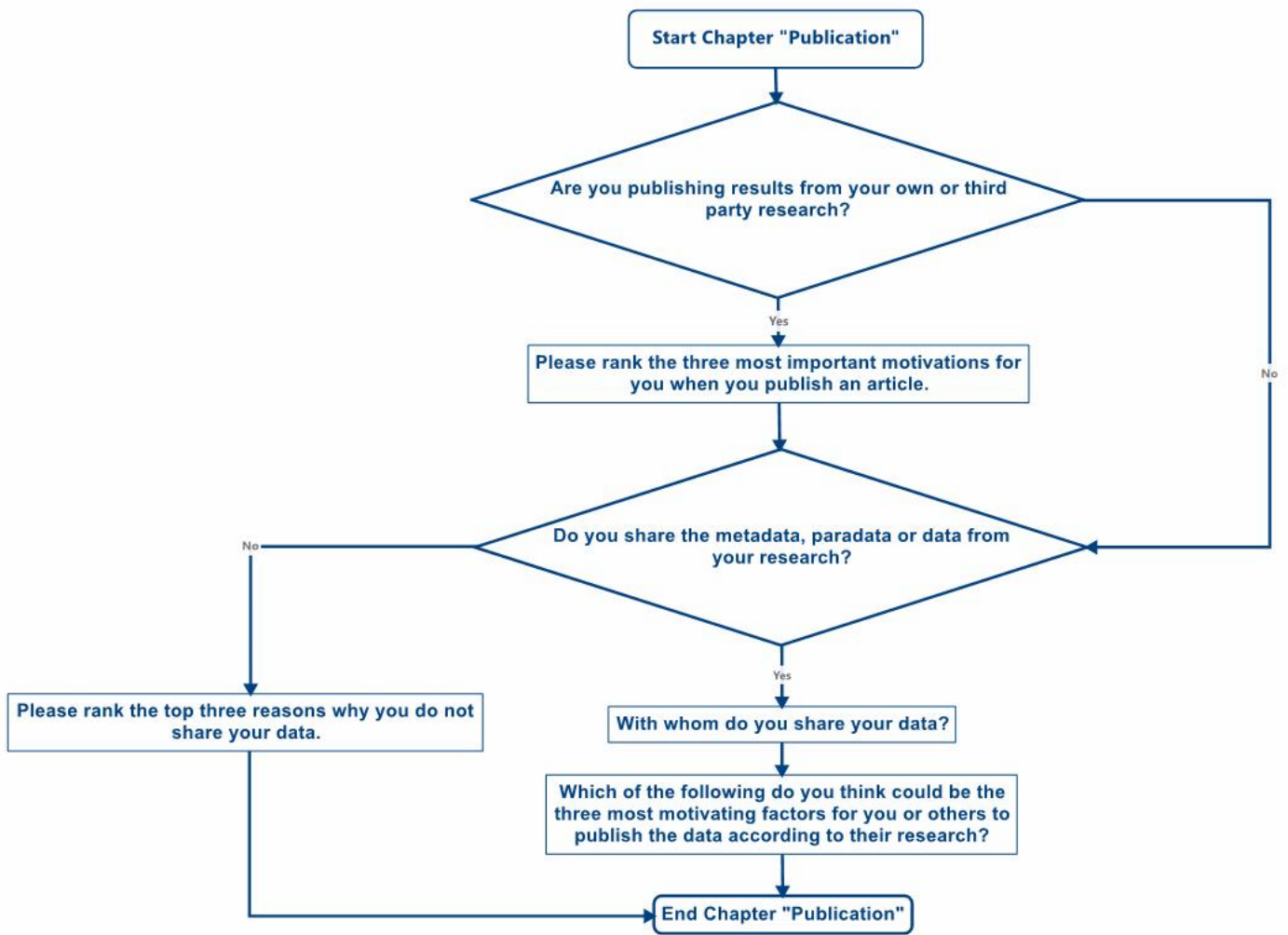


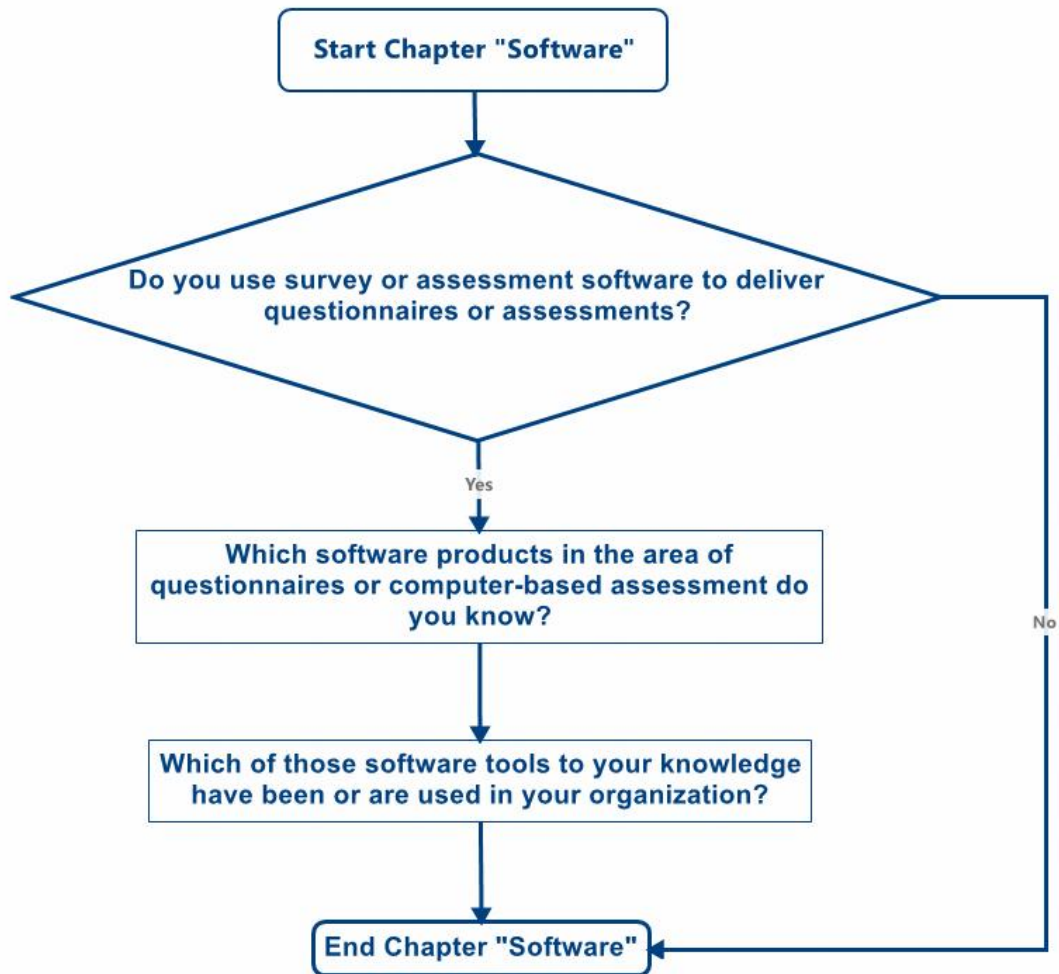
- Chapter 8 – Field Test – Data Collection



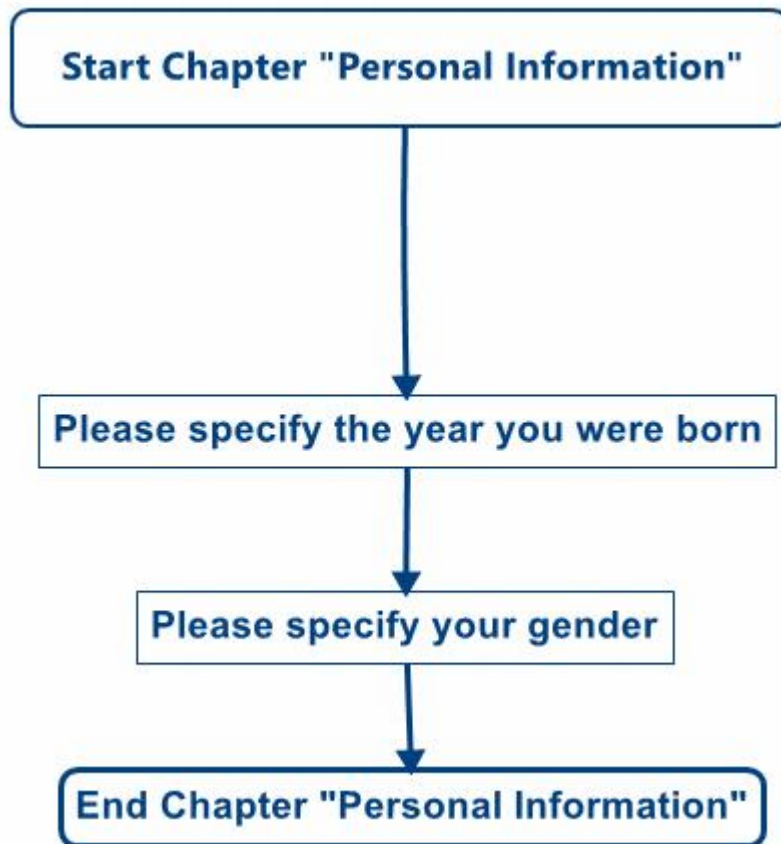
- Chapter 9 – Field Test – Dissemination







- Chapter 12 – Personal Information





## VII. Member list of the DDI Alliance

- *University of Alberta, Canada*
- *Australian Bureau of Statistics (ABS)*
- *Australian Data Archive (ADA)*
- *University of California, Berkeley -- Computer-Assisted Survey Methods Program and UCDATA*
- *Centre for Longitudinal Studies, Institute of Education, University of London (Associate Member)*
- *Centro De Investigaciones Sociologicas (CIS), Spain*
- *Colectica (Associate Member)*
- *Cornell Institute for Social and Economic Research (CISER)*
- *Danish Data Archive*
- *Data Archiving and Networked Services (DANS), The Netherlands*
- *Epidemiology France, Aviesan - ITMO Santé Publique*
- *Eurostat*
- *Finnish Social Science Data Archive*
- *Food and Agriculture Organization (FAO)*
- *German Institute for International Educational Research (DIPF)*
- *German Socio-Economic Panel Study (SOEP)*
- *GESIS - Leibniz Institute for the Social Sciences*
- *University of Guelph*
- *Institute for Social and Economic Research (ISER)*
- *Institute for the Study of Labor (IZA)*
- *Inter-university Consortium for Political and Social Research (ICPSR)*
- *University of Kansas, Institute for Policy and Social Research (IPSR) (Associate Member)*
- *Massachusetts Institute of Technology (MIT)*
- *University of Minnesota, Minnesota Population Center*
- *National Institute of Statistics and Economic Studies (INSEE)*
- *Norwegian Social Science Data Service (NSD)*
- *Open Data Foundation (Associate Member)*
- *Research Data Centre of the German Federal Employment Agency, Institute for Employment Research (IAB)*
- *Stanford University*
- *Statistics New Zealand*
- *Survey Research Operations, University of Michigan*
- *Swedish National Data Service (SND)*
- *Swiss Foundation for Research in Social Sciences (FORS)*
- *United Kingdom Data Archive*
- *University of Toronto Libraries*
- *University of Toronto Scholars Portal*
- *University of Washington, Center for Studies in Demography & Ecology (CSDE)*
- *U.S. Bureau of Labor Statistics (Associate Member)*
- *World Bank, Development Data Group (DECDG)*

(Retrieved from <http://www.ddialliance.org/alliance/structure#members>)

## VIII. Main Test Questionnaire in re-usable metadata standards and questionnaire formats

Last but not least the dissertation closes its appendices with examples of the questionnaires in different formats from DDI Lifecycle v3.2 to Triple-S. The conversions were not easy and prove the point of the whole dissertation. The survey was conducted in Limesurvey V2.05 which can handle for saving the questionnaire its own format Limesurvey Survey Structure (LSS) files and Limesurvey Survey Archive (LSA) files. These files cannot be processed with any other software than Limesurvey itself. Furthermore there is an export functionality in queXML v1.4 with the disadvantage all survey logic gets lost (only the questions themselves can be imported or exported).

For creating the flow charts in the previous chapters and the output in this chapter the questionnaires had to be re-developed in different tools. One of them in Canard which was developed by a colleague of the author Sam Spencer formerly from the Australian Bureau of Statistics (ABS) in Canberra, Australia. Nevertheless to provide proper data documentation several re-development, migration and metadata enrichment processes were necessary and it can be doubted a scientist would undergo any of them as they were quite time-consuming. The author believes it would be hypocritical to publish more than 200 pages on metadata management and the flaws of it while offering no adequate metadata documentation.

Therefore the metadata documentation and the datasets will be published in the research data centre of the German Institute for International Educational Research (DIPF) at

<http://www.forschungsdaten-bildung.de>

Furthermore all metadata and data can be requested from me directly by mail:

[barkow@dipf.de](mailto:barkow@dipf.de)

Metadata and datasets are available under Creative Commons (CC) license at no costs.

The questionnaire is available in the following formats:

- Data Documentation Initiative (DDI) Codebook V2.5
- Data Documentation Initiative (DDI) Lifecycle V3.1
- Data Documentation Initiative (DDI) Lifecycle V3.2
- queXML v1.4
- Structured Questionnaire Building Language V0.1
- Triple-S V2.0

The datasets are available in the following formats:

- SPSS file
- Stata file
- Comma-Separated Value (CSV) file
- Excel Spreadsheet (XLS) file

Further proprietary formats for questionnaires like Blaise 4.8/5.0, Nipo or CASES are possible as converters for these tools from DDI exist. Nevertheless they are normally connected to manual post-processing so the author refrained from a default conversion from this commercial formats.

Conversions of questionnaires are possible with a lot of effort, unfortunately this cannot be said about cognitive instruments for computer-based assessment as the dissertations proves. Therefore the goal should be following the example of the social sciences and the metadata standards established there.

Here some snippets how some of the formats look like. Only one question of the questionnaire was converted as an example. The question on data sources was used as it at the same time consisted of a whole chapter. The snippets also show different levels of data loss during conversion.

### a. Snippet from Data Documentation Initiative (DDI) Lifecycle V3.1

Chapter “Data Analysis” in DDI V3.1:

```
<?xml version="1.0"?>

<ddi:DDIInstance xmlns:ddi="ddi:instance:3_1" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:sqbl="sqbl:1" xmlns:a="ddi:archive:3_1" xmlns:r="ddi:reusable:3_1"
xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns:d="ddi:datacollection:3_1"
xmlns:l="ddi:logicalproduct:3_1" xmlns:c="ddi:conceptualcomponent:3_1"
xmlns:ds="ddi:dataset:3_1" xmlns:s="ddi:studyunit:3_1" xmlns:g="ddi:group:3_1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" id="x0" version="0.0.1"
agency="com.kidstrythisathome.ddirepo.legostormtroopr"><g:ResourcePackage
id="x1"><g:Purpose id="x2"><r:Content/></g:Purpose><g>DataCollection><d>DataCollection
id="qImport_346646_DataCollection"><d:QuestionScheme
id="qImport_346646_QuestionScheme"><d:QuestionItem id="DA01"
objectSource="www.kidstrythisathome.com#DA01"><d:QuestionText
xml:lang="en"><d:LiteralText><d:Text>How often do you use the following data sources for your
research?
</d:Text></d:LiteralText></d:QuestionText><d:CodeDomain><r:CodeSchemeReference><r:ID>
DA01_CodeList</r:ID></r:CodeSchemeReference></d:CodeDomain></d:QuestionItem></d:Quest
ionScheme><d:ControlConstructScheme
id="qImport_346646_QuestionConstructs"><d:QuestionConstruct
id="DA01_QC"><d:QuestionReference><r:ID>DA01</r:ID></d:QuestionReference></d:Questio
nConstruct></d:ControlConstructScheme><d:ControlConstructScheme
id="qImport_346646_ModuleLogic"><d:Sequence
id="qImport_346646_ModuleLogic"><d:ControlConstructReference><r:ID>sec_title_1</r:ID></d
:ControlConstructReference><d:ControlConstructReference><r:ID>DA01</r:ID></d:ControlCons
tructReference></d:Sequence></d:ControlConstructScheme></g>DataCollecti
on><g:LogicalProduct><l:LogicalProduct
id="qImport_346646_LogicalProduct"><l:CategoryScheme
id="DA01_CategoryScheme"><l:Category id="DA01_DA11"><r:Label
xml:lang="en">Always</r:Label></l:Category><l:Category id="DA01_DA12"><r:Label
xml:lang="en">Often</r:Label></l:Category><l:Category id="DA01_DA13"><r:Label
xml:lang="en">Sometimes</r:Label></l:Category><l:Category id="DA01_DA14"><r:Label
xml:lang="en">Rarely</r:Label></l:Category><l:Category id="DA01_DA15"><r:Label
xml:lang="en">Never</r:Label></l:Category></l:CategoryScheme><l:CodeScheme
id="DA01_CodeList"><l:Code><l:CategoryReference><r:ID>DA01_DA11</r:ID></l:CategoryRe
ference><l:Value>DA11</l:Value></l:Code><l:Code><l:CategoryReference><r:ID>DA01_DA12
```

```

</r:ID></l:CategoryReference><l:Value>DA12</l:Value></l:Code></l:Code><l:CategoryReferen
ce><r:ID>DA01_DA13</r:ID></l:CategoryReference><l:Value>DA13</l:Value></l:Code><l:Co
de><l:CategoryReference><r:ID>DA01_DA14</r:ID></l:CategoryReference><l:Value>DA14</l:
Value></l:Code></l:Code><l:CategoryReference><r:ID>DA01_DA15</r:ID></l:CategoryReferen
ce><l:Value>DA15</l:Value></l:Code></l:CodeScheme></l:LogicalProduct></g:LogicalProduct
></g:ResourcePackage></ddi:DDIInstance>

```

## b. Snippet from queXML

Chapter “Data Analysis” in queXML:

```

<section id="31">
  <sectionInfo>
    <position>title</position>
    <text>Data Sources</text>
    <administration>self</administration>
  </sectionInfo>
  <question>
    <text>How often do you use the following data sources for your research?</text>
    <subQuestion varName="DA01_DA101">
      <text>Structured cognitive tests (e.g. computer-based assessments)</text>
    </subQuestion>
    <subQuestion varName="DA01_DA102">
      <text>Structured questionnaires (e.g. multiple choice)</text>
    </subQuestion>
    <subQuestion varName="DA01_DA103">
      <text>Qualitative text data (e.g. narrative interviews, oral history)</text>
    </subQuestion>
    <subQuestion varName="DA01_DA104">
      <text>Experimental designs</text>
    </subQuestion>
    <subQuestion varName="DA01_DA105">
      <text>Observational data (e.g. recordings, transcripts)</text>
    </subQuestion>
    <response varName="DA01">
      <fixed>
        <category>
          <label>Always</label>
          <value>DA11</value>
        </category>
        <category>
          <label>Often</label>
          <value>DA12</value>
        </category>
        <category>
          <label>Sometimes</label>
          <value>DA13</value>
        </category>
        <category>
          <label>Rarely</label>
          <value>DA14</value>

```

```

</category>
<category>
  <label>Never</label>
  <value>DA15</value>
</category>
</fixed>
</response>
</question>
</section>

```

### c. Snippet from Structured Questionnaire Building Language (SQBL)

Chapter “Data Analysis” in SQBL:

```

<sqbl:QuestionModule xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:sqbl="sqbl:1" version="1" name="qImport_346646" xsi:schemaLocation="sqbl:1
https://raw.githubusercontent.com/LegoStormtroopr/sqbl-schema/master/Schemas/sqbl.xsd">
  <sqbl:WordSubstitutions/><sqbl:TextComponents>
    <sqbl:TextComponent xml:lang="en">
      <sqbl:LongName>Data Management Survey - Main Test</sqbl:LongName>
      <sqbl:Title>Data Management Survey - Main Test</sqbl:Title>
    </sqbl:TextComponent>
  </sqbl:TextComponents>
  <sqbl:ModuleLogic>
    <sqbl:Statement name="sec_title_1">
      <sqbl:TextComponent xml:lang="de"/><sqbl:TextComponent xml:lang="en">
        <sqbl:StatementText>Data Sources</sqbl:StatementText>
      </sqbl:TextComponent>
    </sqbl:Statement>
    <sqbl:Question name="DA01">
      <sqbl:TextComponent xml:lang="de"/><sqbl:TextComponent xml:lang="en">
        <sqbl:QuestionText>How often do you use the following data sources for your research?
      </sqbl:QuestionText>
    </sqbl:TextComponent>
  <sqbl:ResponseType>
    <sqbl:CodeList>
      <sqbl:Codes>
        <sqbl:CodePair code="DA11">
          <sqbl:TextComponent xml:lang="en">Always</sqbl:TextComponent>
        </sqbl:CodePair>
        <sqbl:CodePair code="DA12">
          <sqbl:TextComponent xml:lang="en">Often</sqbl:TextComponent>
        </sqbl:CodePair>
        <sqbl:CodePair code="DA13">
          <sqbl:TextComponent xml:lang="en">Sometimes</sqbl:TextComponent>
        </sqbl:CodePair>
        <sqbl:CodePair code="DA14">
          <sqbl:TextComponent xml:lang="en">Rarely</sqbl:TextComponent>
        </sqbl:CodePair>
        <sqbl:CodePair code="DA15">
          <sqbl:TextComponent xml:lang="en">Never</sqbl:TextComponent>
        </sqbl:CodePair>
      </sqbl:Codes>
    </sqbl:CodeList>

```

```

</sqbl:ResponseType>
<sqbl:SubQuestions>
  <sqbl:SubQuestion>
    <sqbl:TextComponent xml:lang="en">
      Structured cognitive tests (e.g. computer-based assessments)
    </sqbl:TextComponent>
    <sqbl:TextComponent xml:lang="de"/></sqbl:SubQuestion>
    <sqbl:SubQuestion>
      <sqbl:TextComponent xml:lang="en">
        Structured questionnaires (e.g. multiple choice)
      </sqbl:TextComponent>
      <sqbl:TextComponent xml:lang="de"/></sqbl:SubQuestion>
      <sqbl:SubQuestion>
        <sqbl:TextComponent xml:lang="en">
          Qualitative text data (e.g. narrative interviews, oral history)
        </sqbl:TextComponent>
        <sqbl:TextComponent xml:lang="de"/></sqbl:SubQuestion>
        <sqbl:SubQuestion>
          <sqbl:TextComponent xml:lang="en">
            Experimental designs
          </sqbl:TextComponent>
          <sqbl:TextComponent xml:lang="de"/></sqbl:SubQuestion>
          <sqbl:SubQuestion>
            <sqbl:TextComponent xml:lang="en">
              Observational data (e.g. recordings, transcripts)
            </sqbl:TextComponent>
            <sqbl:TextComponent xml:lang="de"/></sqbl:SubQuestion>
          </sqbl:SubQuestions>
        </sqbl:Question>
      </sqbl:ModuleLogic>
    </sqbl:QuestionModule>

```

#### d. Snippet from Triple-S 2.0

Chapter “Data Analysis” in Triple-S:

```

<?xml version="1.0"?>
<!DOCTYPE sss PUBLIC "-//triple-s//DTD Survey Interchange v2.0//EN"
"http://www.triple-s.org/dtd/sss_v20.dtd">
    <sss version="2.0" modes="interview analysis"><date/><time/><origin>SQBL
Output</origin><user>Canard</user><survey><record ident="V"><variable ident="1"
type="single"><name>DA01</name><label><text mode="interview" xml:lang="en">How often
do you use the following data sources for your research? </text></label><position
start="1"/></variable></record></survey></sss>

```