The Data Quality Index:
Improving Data Quality in Irish Healthcare Records

David Hickey[1][a], Rita O Connor[1][b], Pauline McCormack[1][c], Peter Kearney[1][d], Roosa Rosti[1][e],
and Rob Brennan[1, 2][f]

[1]*School of Computing, Dublin City University, Dublin, Ireland*
[2]*ADAPT Centre, Dublin City University, Dublin, Ireland*
{david.hickey26, rita.oconnor83, pauline.mccormack28, peter.kearney8, roosa.rosti2}@mail.dcu.ie, rob.brennan@dcu.ie

Keywords: Data Governance, Data Quality, Healthcare, Data Management

Abstract: This paper describes the Data Quality Index (DQI), a new data quality governance method to improve data quality in both paper and electronic healthcare records. This is an important use case as digital transformation is a slow process in healthcare and hybrid systems exist in many countries such as Ireland. First a baseline study of the nature and extent of data quality issues in Irish healthcare records was conducted. The DQI model and tools were then developed, based on established data quality and data governance principles. Evaluation of the model and tools showed a significant improvement in data quality was achieved in a healthcare setting. This initial evaluation of the model was against paper healthcare records, but the model can also be used as part of an electronic healthcare record system.

1 MOTIVATION

In the Irish healthcare system, the patient Healthcare Record (HCR) is, for the most part, a paper file that contains all aspects of the patient journey, in as structured a form as paper will allow. However, by its very nature, this paper record is hard to manage and control, varies in the quality of the data therein, and it is often difficult to follow and to easily understand (Health Service Executive, 2011). This lack of governance in paper records can give rise to problems in providing the best level of care for the patient. It has been shown that the data quality of healthcare records has a direct impact on the quality of care administered (Urquhart et al., 2018) and the time spent with patients (Fennelly et al., 2020).

Improving patient outcomes is a key driver of digital transformation of healthcare and these are based on the introduction of Electronic Healthcare Records (EHR). These are a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting (HIMSS). Advantages of electronic health records are the closing of gaps present in paper HCRs, and utilizing standardized data entry controls. A number of studies on patient medical records (Greenhalgh et al., 2010) and informatics in patient care over the past decade suggest a need to ensure that measures are put in place to safeguard the quality of data in electronic health records, and include a warning of the potential for a 'landslide of poor quality data' to negatively impact clinical outcomes (Hussey & Kennedy, 2016). Others on this journey of digital transformation have made mistakes which are well reported (Bowman, 2013), and the promise of improvements in data quality are often not realized (Charnock, 2019). The most common mistake in digitisation seems to be the digitalization of the existing analogue process, leading to embedding of the same data quality problems into the electronic HCR (EHR). This a key component of successful EHR deployment is a robust design for data quality from the architecture phase

[a] https://orcid.org/0000-0002-9999-5583
[b] https://orcid.org/0000-0001-6222-4702
[c] https://orcid.org/0000-0001-6502-9996
[d] https://orcid.org/0000-0002-9243-3297
[e] https://orcid.org/0000-0002-8341-6153
[f] https://orcid.org/0000-0001-8236-362X

that takes into account paper record quality as well as electronic record quality. There is currently a lack of common frameworks to support the healthcare industry to achieve this.

The research question studied in this paper is: *To what extent can the introduction of a data quality governance process for both paper and electronic healthcare records improve data quality in healthcare organisations undergoing digital transformation?* Here data quality refers to the quality dimensions of completeness, validity, accuracy, consistency, availability and timeliness of the data (ISO/IEC 25012). Data governance is defined as 'the exercise of authority and control (planning, monitoring and enforcement) over the management of data assets' a collection of practices and processes which help to ensure the formal management of data assets within an organization (DAMA DMBOK).

The technical approach has been to develop a set of requirements for joint paper-HCR quality governance; to validate the nature and extent of the data quality issues of healthcare records in a typical Irish healthcare setting; to develop a model for use with the Healthcare Record that will improve data quality based on sound data quality governance methodologies; to build and evaluate tools based on the model; to iterate the design, based on real-world testing using a design science approach.

The contribution of this paper is a new framework for data quality governance in both paper-based and electronic healthcare records that is suitable for organisations undergoing digital transformation and that supports building in quality processes from the start into the EHR-based system.

The rest of this paper is organised as follows: section 2 discusses related work, section 3 specifies our research methodology, section 4 derives system requirements and discusses a baseline data quality study, section 5 describes the design of the data quality index method, section 6 evaluates the system and section 7 provides conclusions.

2 RELATED WORK

This section discusses electronic healthcare records, data governance and data quality in the context of digital transformation of healthcare systems.

2.1 The Electronic Healthcare Record

As technology has become more prevalent people and organisations have inevitably moved from paper-based forms of data collection to electronic means.

This is no different in the healthcare sector where a big push towards the development of electronic healthcare records has been seen throughout the world (Stone, 2014). An Electronic Healthcare Record (EHR) is a digital version of a patient's paper file and can be best described as "a longitudinal collection of electronic health information about individual patients and populations" (Gunter & Terry, 2005).

Following in the footsteps of many other countries, Ireland has already begun to digitise the paper patient healthcare into electronic form (Grogan, 2020). This is recognised in the 2013 Department of Health & HSE's eHealth Strategy (Health Service Executive & Dept of Health, 2013) along with the 2017 Committee on the Future of Healthcare Sláintecare Report (Houses of the Oireachtas, 2017), which note the potential of eHealth to be the biggest and most effective driver of change and improvement of patient outcomes across the health system. The ability for healthcare professionals and patients to have access to up-to-date clinical records at the point of care, can lead to improved decision making, reduction in errors and higher quality of patient care.

While there has been much research and associated literature on the topic of the Electronic Healthcare Record, most of the focus has been on the technology challenges, data sharing, security concerns, privacy, decision support, and improvements in clinical outcomes. Data governance and data management are themes that were rarely discussed in relation to the electronic healthcare record, and data quality is even less frequently discussed in the literature. Below we provide highlights of relevant work to date.

2.2 Data Governance

As organisations hold more and more data, particularly the huge amounts held in electronic healthcare systems (Fenton et al., 2017), data governance is an important consideration in deriving value from the data and in assuring the quality of data. The data governance quality domain is central in Khatri and Brown's data governance model (Khatri & Brown, 2010). Having reviewed literature related to designing data governance within an organisation the Nagle, Sammon and Cleary model of the "information supply chain (ISC)" (Nagle et al., 2019) was found to be the most relevant for Irish healthcare records. This model produced an easy-to-use artefact (the ISC card) and drove data governance best practice throughout the information supply chain.

The importance of a data governance framework for electronic healthcare records cannot be understated.

The OECD recommends that governments should establish a "national health data governance framework" (*Recommendation of OECD Council on Health Data Governance*, 2016) and that organisations processing personal health data demonstrate that they meet national expectations for health data governance. The OECD also set out a number of recommendations for nations for establishing this.

A "Guidance for Information Governance" report produced by the Health Information and Quality Authority of Ireland (HIQA) (*Guidance on Information Governance*, 2012) for healthcare providers in Ireland highlights the need for data governance within the healthcare environment. Interestingly no revisions have been made to the document since 2012, which again strengthens the need for further activity in this area and the Irish need to develop a model in line with our research question that considers the current state of art, legislation and practices, which inevitably have developed and changed in the last eight years.

2.3 Data Quality

The OECD recommendations mention data quality as an important consideration in implementing electronic healthcare records and data governance in relation to this. HIQA in their guidance document also recognise the importance of data quality and the effect it has on quality of care and improved outcomes. This impact on patient outcomes highlights the importance of data quality to enable the electronic health record to achieve its aim in improving quality of care. For example, true system interoperability in an electronic health record require data to be of high quality to ensure reliability of the inter-system communications.

The importance of data quality is also stressed in an earlier report by Gordon and Greene (Gordon & Greene, 2014) which argues that healthcare organisations may be facing "faster incorrect information" generation through the implementation of different systems and applications using inconsistent metrics and generating data in different formats. This has the potential to create poor quality data (lack of consistency, understandability, portability) and data errors where data is shared. If data is poor quality it has been shown to lead to inadequate care being provided and therefore worse patient outcomes. (*Guidance on Information Governance*, 2012).

Data quality as presented in the electronic healthcare record literature has identified specifically the data quality dimensions of, 'accuracy, completeness, consistency, credibility and timeliness of data' (Parsons et al., 2012). The importance of accuracy and completeness show how HCR can deliver clear advantages through the implementation of EHRs as forms or schemata with standardised data entry controls. The traditional paper-based healthcare record where data is entered manually has been seen to contribute to data quality issues (Charnock, 2019). This also contributes to the issue of lower quality patient care as identified throughout the literature. The importance of high-quality data in EHRs can also have legal, social and care impacts as patients' lives are at risk where a lack of data accuracy or completeness presents itself.

Thus it can be seen that there is a deep need for data quality models that can be used for both the paper-based and electronic healthcare records. As digital transformation of healthcare proceeds these needs are growing due to the increased dependence on algorithmic IT supports.

3 METHODOLOGY

We took an Agile approach to project management in order to allow the project team the flexibility to work within a structured framework, while providing the flexibility to iterate as the model requirements, design, testing and evaluation evolved.
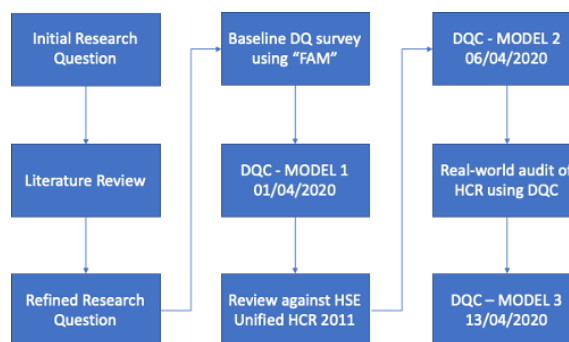

Figure 1: Project management approach.

The main steps were (Figure 1):
- Review of previous work and literature;
- Identification of Governance Model identified– 'Decision Domains for Data Governance' (Khatri and Brown, 2010);
- Project focus on Data Quality within the five domains in that model;

- 'Information Supply Chain' (Nagel et al., 2019) used to as the stimulus for our own model development;
- Baseline review of paper HCR using 'Friday Afternoon Measurement' methodology; (Redman, 2016)
- Initial development of Data Quality Card (DQC) and Data Quality Index (DQI) Model;
- Alignment of entities and attributes with HSE standard led to revised DQI Model (v2);
- Simulated Healthcare Record evaluation and revision of attribute definitions;
- Finalisation of DQI Model (v3).

Stakeholder involvement involved a hospital Data Protection Officer and an Electronic Healthcare Record Privacy Lead with clinical expertise, a Project Manager, a Legal expert and Innovation and Design-Thinking expert. The evaluation methodology used concepts from the Design Science/Design Thinking discipline. The Experimental design evaluation method (Hevner et al., 2004), was applied so that we could analyse the artefact (our DQI model) in a controlled environment for utility with regard to data quality levels.

4 REQUIREMENTS

The target of our study was a robust design for a data quality governance framework from the architecture phase of the Irish healthcare HCR that takes into account paper record quality as well as electronic record quality. From our evaluation of the state of the art in data quality for paper-electronic healthcare records (section 2), we determined that none of this met the requirements of the current digital transformation of the Irish healthcare system, and so we set about developing a new model to address these issues. The requirements for this data quality governance model are:
- An easy-to-use artefact for non-IT staff;
- Applicable to paper and electronic records;
- Can be applied historically to provide a quality score;
- Can be used on new records to ensure high quality;
- Capable of being developed into a software solution for the electronic record;
- Ideally applicable to other domains;
- Support the Irish Unified Healthcare Record standard (Table 1);
- Addresses the most frequent errors identified in Irish Healthcare records (Table 2).

Table 1: Unified Healthcare Record (HSE Ireland).

CATEGORY	
1. Registration	**7. Consent**
Patient name	Consent Forms
DOB	**8. Clinical**
MRN	Cardiovascular
Contact Details	Vascular
Gender	Neurophysiology
Marital Status	Pulmonary
GP Name	GIT
Medical Card	Urology
Next of Kin	Audiology
Admission details	EEG
Discharge details	
2. Administrative	**9. Laboratory**
Patient labels	Biochemistry
Front sheets	Haematology
Relevant Billing	Microbiology
Private Insurance Forms	Histopathology
3. Correspondence	Immunology
Referral Letters	Molecular Diagnosis
Discharge Communications	
Ambulance Transfer Sheets	
Other Correspondence	
4. Clinical Notes	**10. Radiology**
All clinical notes	Diagnostic Imaging Reports
Integrated Care	X-Ray
Intensive Care	MRI
Emergency Dept	PET
Treatment Notes	Ultrasound
5. Nursing	**11. Medicines**
Temp and Obs Sheets	Drug Prescribing Sheets
Nursing Care plans	Prescriptions
Intensive Care Nursing	Nutritional Supplements
Evaluations	Blood Transfusion
6. Procedures	**12. Social Care**
Procedure Forms	Assessment Forms
Anaesthetic Records	Care Cards
Epidural Records	Food Diaries
Implant Records	Dietary
Blood Loss Sheets	
Instrument Count Sheets	
Theatre Checklists	
Theatre Nursing	

4.1 Baseline Data Quality Study

To quantify the extent of data quality problems to solve, and to measure its extent, we applied an approach known as the Friday Afternoon Measurement (FAM) (Redman, 2016) methodology. This approach involves reviewing the last 100 data records created and looking at no more than 10-15 data attributes for each record, and marking obvious errors. Then counting up the number of error-free records gives a number from 0 – 100 called the data quality or "DQ" score.

Our baseline FAM review was carried out based on a single reference hospital's HealthCare Record (HCR). In our baseline review, we got a DQ score of 11 (Figure 2) indicating that there would be only 11 perfect records out of the 100 reviewed. This confirmed the extent of data quality issues in the paper healthcare record. This is a low score as 33-67 would be typical scores in baseline studies according to Redman.

Attributes Correct	% of Paper Records
0	2%
1	4%
2	6%
3	4%
4	3%
5	9%
6	11%
7	21%
8	14%
9	7%
10	8%
11	11%

DQ 11
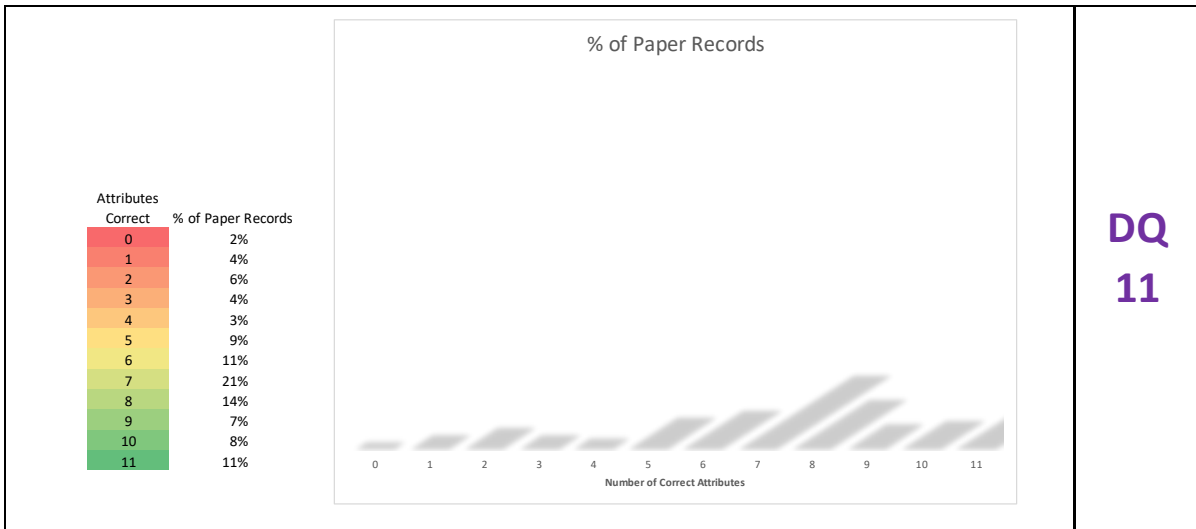
Figure 2: Baseline measurement of data quality in HCR.

We extracted the typical errors from our baseline, and then compared these with errors that had previously been identified by the HSE in their review of Standards for healthcare records management (Health Service Executive, 2011). This allowed us to arrive at a "Top 20" of most common data quality issues in the healthcare record (Table 2). A key requirement for any data quality method would be to mitigate the frequency of these types of errors.

5 DATA QUALITY INDEX METHOD DESIGN

The starting point for our quality method development was the ISC Card developed by Nagle, Sammon and Cleary (Nagle et al., 2019). Their concept provided a Data Governance framework for people, process and technology covering all stages in the Information Supply chain.

The basic building block of our model is the "Data Quality Card" or DQC (Figure 4). The DQC measures the quality of one entity with many attributes. Each entity may have a number of specific attributes which relate only to that entity, and attributes in common with other entities. Each attribute on a data quality card is evaluated to see if it is present and accurate, which gives a "true" result. Our model then provides a "Data Quality Index" or DQI for each entity. The DQI is the percentage of attributes with perfect quality (in keeping with the FAM methodology).

Table 2: Top 20 errors in Irish HCR.

TOP DATA QUALITY ISSUES IN PATIENT RECORD		
		PAPER HCR
1	Name	o
2	DOB	o
3	MRN	o
4	Missing patient label	X
5	Address (physical or email)	X
6	Phone No	X
7	Dates & format	X
8	Times & format	X
9	Chronology	X
10	Stale data	X
11	Clinical data missing	X
12	Inacessible at Point of Care	X
13	Lost record	X
14	Forms missing	X
15	Transcription error	X
16	Illegibility error	X
17	Source not identified	X
18	Signature missing	X
19	Missing "media"	X
20	Fragmented record error	X
X	= Likely to occur	
o	= Unlikely to occur	

Our scenario relates to the Healthcare Record. As we have seen earlier (Table 1), each healthcare record has a number of discrete sections or entities. In Ireland, the Unified Healthcare Record (Health Service Executive, 2011) has 12 standard sections or entities defined. Applying our DQI Model, one Data Quality Card (DQC) is thus created for each entity (or section) in the healthcare record (Figure 3).

The overall DQI model is then made up of a number of Data Quality Cards, with each DQC representing one entity in the scenario being examined – giving 12 cards in our Healthcare Records scenario.

Each DQC has its own DQI score, which allows easy identification of quality issues at entity and attribute level. Combining all Data Quality Cards then provides the overall model with a single Data Quality Index.

$$DQI = \frac{sum(DQCattributes=true)}{sum(DQCattributes)} \qquad (1)$$

In the DQI model, the number of cards, the number of entities, the entity-specific attributes and the common attributes are all flexible, which potentially allows the model to be applied to any data quality scenario for any sector.

It is intended that a DQC can be used as an artefact or tool (either printed or electronic) to calculate the DQI of any individual entity. We have also developed a simple data entry form for the model, which can be printed or completed electronically (Figure 6).
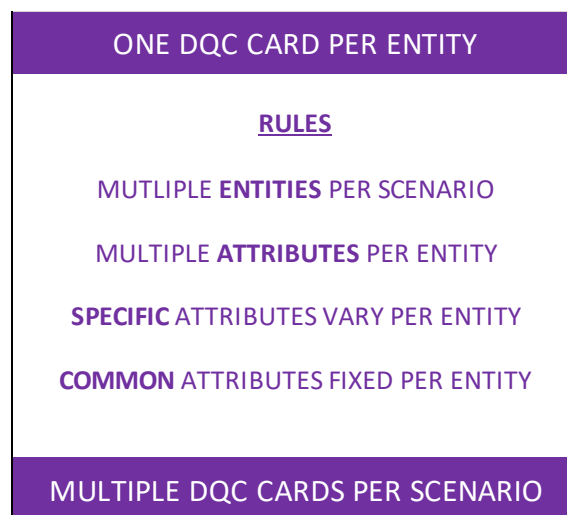
Figure 3: DQI model rules.
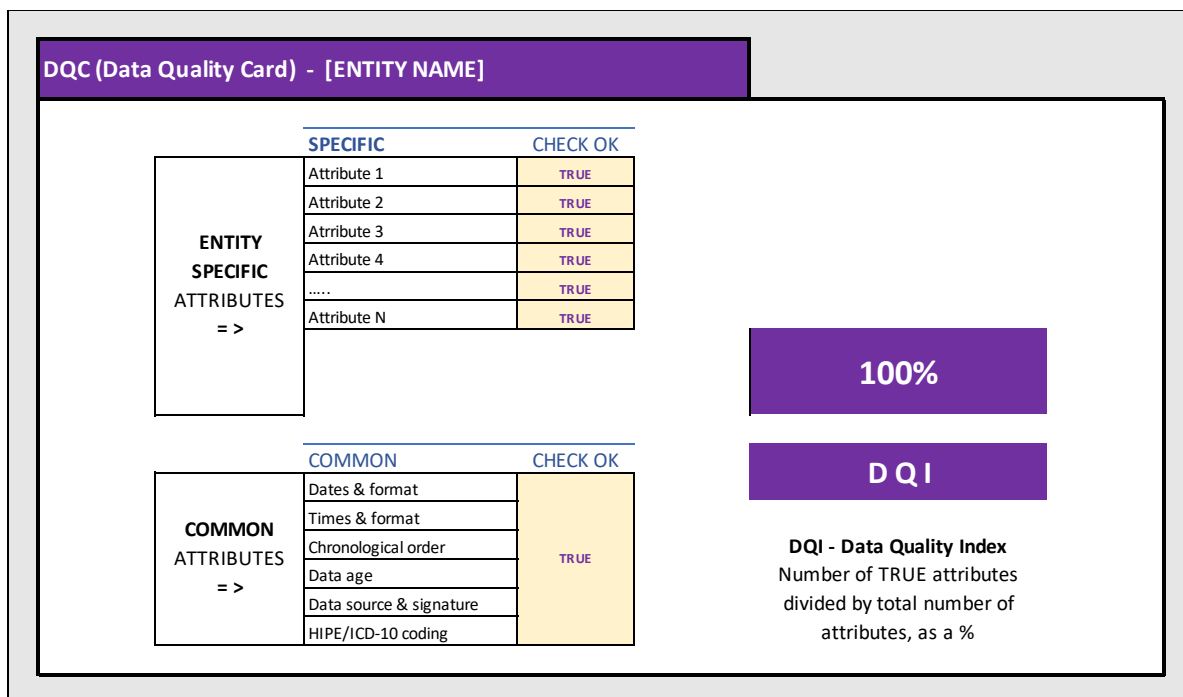
Figure 4: The Data Quality Card (DQC).

We envisage this simple data entry form being used in different ways: firstly, it can be printed out on a single sheet and used as an audit tool on paper records; secondly, it can be included as a checklist for new records; finally, it can be used as an interactive (electronic) checklist to be completed and directly drive the model giving individual DQI scores for each entity as well as an overall DQI score for the record.

Finally, the complete DQI Model rolls up the data from all the constituent Data Quality Cards, into a single aggregated view (see Figure 7 below). The complete Model is linked to each underlying DQC and to the Data Entry Form.

A simple software model was developed for the project, that allowed the user to see the DQI of each DQC change in real time as data quality was evaluated using the data entry form. At the same time, an overall DQI score was displayed.

This allowed the model to be evaluated, but also gives users an immediate visualisation of where data quality issues exist in the underlying record, to an entity and attribute level.

5.1 DQI Design for the Irish Unified Healthcare Record

We developed a first version of the DQI Model based on the individual hospital HCR studied in our data quality baseline, and it had 11 entities and 80 attributes. We then reviewed the model to include all entities in the Unified Healthcare Record. This resulted in Version 2 of the DQI model which has 12 entities corresponding to the sections in the Unified HCR, and 100 attributes specified by the HSE. This now makes the model usable in any healthcare setting in Ireland. During evaluations (Section 6), we found a number of form attributes were ambiguous or lacked clarity and so we redefined these and created a new (Version 3) of the DQI model.
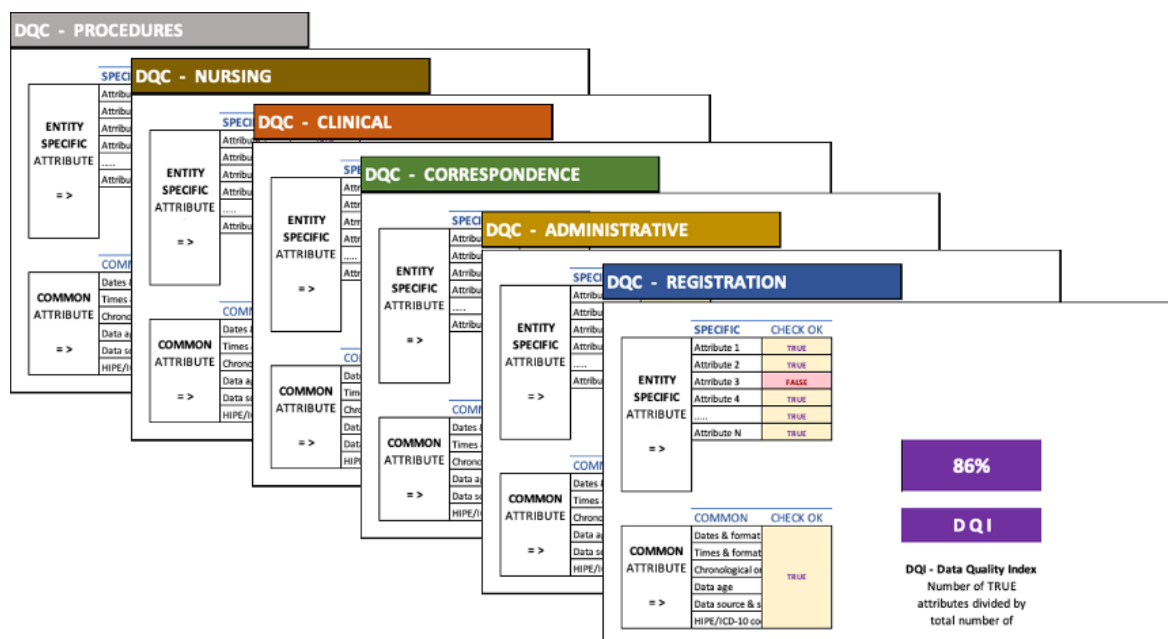
Figure 5: multiple Data Quality Cards make up the DQI model.

DATA QUALITY CARD (DQC) - CHECKLIST FOR PATIENT HEALTHCARE

Please check the following (MASTER) Patient Record Details and if accurate or not required

MASTER PATIENT RECORD

- ☐ PATIENT NAME
- ☐ DOB
- ☐ MRN
- ☐ CLINICIAN

- ☐ Gender
- ☐ Email address
- ☐ Home address
- ☐ Previous address

- ☐ Marital status
- ☐ Phone - mobile
- ☐ Phone - home
- ☐ Phone - other

- ☐ Next of kin name
- ☐ Next of kin address
- ☐ Next of kin phone
- ☐ Next of kin mobile

- ☐ Mot
- ☐ Dat
- ☐ Ref
- ☐ Mo

Please check that the following are present (or not required) and accurate in the patient record

ADMINISTRATIVE SECTION

- ☐ Patient Label
- ☐ Front Sheet(s)
- ☐ Billing Information
- ☐ Private Insurance

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

CORRESPONDENCE SECTION

- ☐ Referral Letter(s)
- ☐ Admission Notes
- ☐ Discharge Communications
- ☐ Ambulance Trar

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

CLINICAL NOTES SECTION

- ☐ All Clinical Notes including Integrated Care Pathways and ICU notes
- ☐ Emergency Department

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

NURSING NOTES SECTION

- ☐ Vital Signs Observations
- ☐ Fluid Balance
- ☐ Nursing Care Plans
- ☐ Intensive Care

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

PROCEDURES SECTION

- ☐ Procedure Forms
- ☐ Anaesthetic Forms
- ☐ Epidural Infusion
- ☐ Implant Records
- ☐ Blood L

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

CONSENT SECTION

- ☐ Consent form - Adult
- ☐ Consent form - child
- ☐ Consent form - research
- ☐ Disc

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

CLINICAL MEASUREMENTS SECTION

- ☐ Cardiovascular
- ☐ Haemodynamic
- ☐ Neurophysiology
- ☐ Pulmonary Function
- ☐ GIT R

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

LABORATORY RESULTS SECTION

- ☐ Biochemistry
- ☐ Haematology
- ☐ Microbiology
- ☐ Histopathology
- ☐ Immunology

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

RADIOLOGY AND DIAGNOSTIC IMAGING RESULTS SECTION

- ☐ X- Ray
- ☐ CT Scan
- ☐ MRI
- ☐ Ultrasound
- ☐ PET Scan

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

PRESCRIBED MEDICINES SECTION

- ☐ Drug Prescribing
- ☐ Prescriptions
- ☐ Nutritional Supplements
- ☐ Blood Transfusions

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

HEALTH & SOCIAL CARE PROFESSIONALS SECTION

- ☐ Assessment
- ☐ Care Cards
- ☐ Dietary Form
- ☐ Psychology Report

☐ Common attributes (source, retention date, date & time formats, chronological order, signature

Figure 6: Data entry checklist for the DQI model.

DATA QUALITY MODEL FOR PAPER RECORDS - DQC (Data Quality Card)
CALCULATION OF DQI (Data Quality Index)

DATA QUALITY CARD (DQC) - 12X100

OVERALL DATA QUALITY INDEX — 0%

DQC Model (V1.0) based on single Reference-Hospital Healthcare Record
DQC Model (V2.0) added 21 attributes from HSE Unified Healthcare Record (Ireland)
DQC Model (V3.0) based on field evaluation of V2.0 and clarification of attributes

REGISTRATION (CHECKED)

Attribute	CHECKED
Patient Name	FALSE
DOB	FALSE
MRN	FALSE
Gender	FALSE
Marital Status	FALSE
Home Address	FALSE
Previous Address	FALSE
Phone - Mobile	FALSE
Phone - Home	FALSE
Phone - Other	FALSE
Email Address	FALSE
Next of Kin Name	FALSE
Next of Kin Address	FALSE
Next of Kin Phone	FALSE
Next of Kin Mobile	FALSE
Date of Admission	FALSE
Referral source	FALSE

SECTION SPECIFIC ATTRIBUTES ⇒ / COMMON ATTRIBUTES ⇒

Section	Attribute	CHECKED	Attribute	CHECKED
ADMINISTRATIVE	Patient Label	FALSE		
	Front Sheet	FALSE		
	Relevant Billing	FALSE		
	Private Insurance	FALSE		
	Dates & format	FALSE		
	Times & format	FALSE		
	Chronological order	FALSE		
	Data age	FALSE		
	Data source/signed	FALSE		
	HIPE/ICD-10 coding	FALSE		
CORRESPONDENCE	Referral letters	FALSE		
	Admission Notes	FALSE		
	Discharge Notes	FALSE		
	Ambulance Transfer	FALSE		
	Living Wills	FALSE		
	Dates & format			
	Times & format			
	Chronological order	FALSE		
	Data source/signed			
	HIPE/ICD-10 coding			
CLINICAL	All Clinical Notes	FALSE		
	Emergency Dept	FALSE		
	Health & Social Care	FALSE		
	Nurse Specialist	FALSE		
	Chronological order	FALSE		
NURSING	Vital Signs Obs	FALSE		
	Fluid Balance	FALSE		
	Nursing Care Plan	FALSE		
	ICU Notes	FALSE		
	Evaluations			
	Early Warning Score			
	Chronological order	FALSE		
PROCEDURES	Procedure Forms	FALSE		
	Anaesthetic Forms	FALSE		
	Epidural Infusion	FALSE		
	Implant Records	FALSE		
	Blood Loss Sheets	FALSE		
	Swab Count Sheets	FALSE		
	Instrument Count			
	Theatre Checklist			
	Chronological order	FALSE		
CONSENT	Consent - Adult	FALSE		
	Consent - Child	FALSE		
	Consent - Research	FALSE		
	Unadvised discharge	FALSE		
	Chronological order	FALSE		
MEASUREMENT	Cardiovascular	FALSE		
	Haemodynamic	FALSE		
	Neurophysiology	FALSE		
	Pulmonary Function	FALSE		
	GIT Reports	FALSE		
	Urology Reports	FALSE		
	Audiology Reports	FALSE		
	EEG Reports	FALSE		
	Chronological order	FALSE		
LAB RESULTS	Biochemistry	FALSE		
	Haematology	FALSE		
	Microbiology	FALSE		
	Histopathology	FALSE		
	Immunology	FALSE		
	Molecular/Diagnostic	FALSE		
	Chronological order	FALSE		
IMAGING	X-Ray	FALSE		
	CT Scan	FALSE		
	MRI	FALSE		
	Ultrasound	FALSE		
	PET Scan	FALSE		
	Chronological order	FALSE		
MEDICINES	Drug Prescribing	FALSE		
	Prescriptions	FALSE		
	Nutritional Sup'ment	FALSE		
	Blood Transfusion	FALSE		
	Chronological order	FALSE		
SOCIAL CARE	Assessment Forms	FALSE		
	Care Cards	FALSE		
	Dietary	FALSE		
	Psychology	FALSE		
	Chronological order	FALSE		

Common attributes (Dates & format, Times & format, Chronological order, Data age, Data source/signed, HIPE/ICD-10 coding) are listed for each section.

Figure 7: The complete DQI model dashboard for the Irish Unified Healthcare Record (aggregated view).

6   EVALUATION

6.1 Deployment Case Study

First, we consider the impact of deploying data quality cards (DQC) and the data quality index (DQI) into manual workflows in the Irish healthcare system. Manual (paper) Healthcare records are created on an ad-hoc basis over the duration of a patient's treatment lifecycle by different healthcare professionals. These records are created with varying attention to detail, regarding the data quality recorded and the format used. The resultant anomalies and poor data quality in the healthcare records negatively impacts the decision-making capabilities of medical staff when treating patients.

The DQC and DQI Data Entry Form allow (a) a health care auditor who is assigned the task of reviewing the data quality of paper-based healthcare records, to easily generate a Data Quality Index (DQI) score or measurement for the records on file in an automated manner; and (b) a health care user to ensure data quality at the point of entry of the data by generating a DQI score using either a paper checklist or an electronic data entry form.

The benefits are:

- Quick identification of the entity / attribute(s) within the healthcare record where data quality is a problem;
- Automation of the count of erroneous records which can be cleansed;
- Identification of trends in poor data quality can be easily identified (and recommendations can be made with regard to how they are readily fixed);
- Creators of new paper healthcare records may take greater care when completing records, due to regular data quality audits.

6.2 Domain Expert Validation

In order to quantify the value of the DQC/DQI, a clinical expert user assessed the utility of them with regard to the data quality of all components of the patient record i.e. Registration, Administrative, Correspondence, Clinical, Nursing, Procedures, Consent, Measurement etc).

The data quality was measured by using the Conceptual Framework of Data Quality (Wang & Strong, 1996) with quality dimensions as follows:

- Intrinsic Data Quality refers to the accuracy or believability of the data;
- Contextual Data Quality refers to the data quality associated with the user's task and must be assessed in relation to the purpose of the work, thereby ensuring that whatever the user needs is readily available when it is needed;
- Representational Data Quality includes aspects related to the format of the data (concise and consistent representation) and the meaning of data (interpretability and ease of understanding);
- Accessibility Data Quality relates to the fact that a user has access to, and knows how to retrieve what he/she needs when required to complete a specific task.

The evaluation methodology was inspired by concepts from the Design Science/Design Thinking discipline. The Experimental design evaluation method (Hevner et al., 2004), was applied so that we could analyse the artefact in a controlled environment for usability with regard to data qualities. We evaluated the data quality of manual health care records regarding the "Top 20" data quality errors as outlined in Table 2.

The first evaluation of our DQI Model was conducted by creating new HealthCare Records and looking for an improvement in data quality. In creating these records, we used the DQCs to focus on the "Top 20" errors identified in our FAM review (See Table 2 above). Using DQI Model V2 at the point of data entry provided a DQI score of 37%.

We had expected a greater improvement in data quality, and so we revisited the language and definitions of each of the (100) attributes used to measure the DQI index on the form used by the assessor. When we applied this final version of the DQI Model, again focussing on checking for the "Top 20" errors, we now found we achieved a significant improvement in data quality with a DQI of 81%.

The data quality results are outlined in Table 2 below. In transitioning from Iteration 1 (DQI Model V2) to Iteration 2 of the evaluation (DQI Model V3), the data quality score improved from 37% of records being error free to 81% of the records evaluated being error free. This was achieved by improving the labelling of fields on the assessment form. One attribute of contextual data quality, 'Missing data' was relabelled to become more accurate and relevant as 'Clinical data missing'. From a representational data quality perspective, three attributes were made more meaningful and interpretable. These are: 'Address' edited to become 'Address - physical or email', 'Fragmented' edited to 'Forms missing' and 'Linked Record Error' edited to become 'Fragmented record error'. Finally, one attribute was refined to highlight the point at which accessibility became an issue

regarding the patient's treatment, i.e. from 'record inaccessible' to 'inaccessible at point of care'.

Table 3: Summary of evaluation results.

Evaluation DQ Type	Iteration 1	Iteration 2
Intrinsic	DQI = 31% from Model v2	DQI = 81% from Model v3
Contextual	1 x DQ attribute incomplete	Rename item 11
Representational	3 x DQ attributes vague for user	Identify item 14 forms missing. Rename items 5 & 20
Accessibility	1 x DQ attribute incomplete	Identify item 12 record unavailable

These findings are consistent with our understanding that high-quality data should be intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the user of Healthcare records. It has also shown that the structured formal approach of Wang et al. has allowed us to achieve our goal of measuring data quality in addition to transforming the data quality facilitated by the use of DQC/DQI in auditing health care records.

6.3   Data Protection / GDPR Benefits

While the focus of the project was not on data protection and privacy per-se, we do note some GDPR-related benefits that accrue due to the introduction of the model. These include accuracy (Art 5.1(d)), integrity and confidentiality (Art 5.1.(f)), accountability (Art 5.2) and tracking consent (Art 7 and Art 8).

7   CONCLUSIONS

We set out to determine to what extent can the introduction of a data quality governance process for both paper and electronic healthcare records improve data quality in healthcare organisations undergoing digital transformation. This led us to develop the DQC/DQI model and easy-to-use checklist (Figure 6) and tool, which can be applied to both the paper and electronic records. We applied the DQC/DQI model to the Irish Unified Healthcare Record and developed a simple dashboard (Figure 7). The DQC/DQI model forms the basis for validation rules or prompts for embedding into an electronic healthcare record (EHR) application.

Application of the DQI Model showed a significant improvement in data quality scores during implementation and testing against paper healthcare records. We consistently achieved DQI scores in excess of 80% when only focusing on the most common quality errors in healthcare records. This suggests the model could also provide improved data quality within EHR's, and form the basis for integrated data quality governance in the roll-out of EHRs in Ireland.

The DQI Model requires further testing in a hospital and healthcare environment to demonstrate the effectiveness of the approach in a wider variety of applications and with a larger number of users.

It is hoped that the DQI model can effectively support the transition of paper to electronic healthcare record solutions in Ireland and in other jurisdictions.

REFERENCES

Bowman, S. (2013). Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspectives in Health Information Management*, *10*, 1c.

Charnock, V. (2019). Electronic healthcare records and data quality. *Health Information & Libraries Journal*, *36*(1), 91–95. https://doi.org/10.1111/hir.12249

Fennelly, O., Cunningham, C., Grogan, L., Cronin, H., O'Shea, C., Roche, M., Lawlor, F., & O'Hare, N. (2020). Successfully implementing a national electronic health record: A rapid umbrella review. *International Journal of Medical Informatics*, *144*, 104281. https://doi.org/10.1016/j.ijmedinf.2020.104281

Fenton, S., Low, S., Abrams, K., & Butler-Henderson, K. (2017). Health Information Management: Changing with Time. *Yearbook of Medical Informatics*, *26*, 72–77. https://doi.org/10.15265/IY-2017-021

Gordon, L., & Greene, J. (2014). Better data through IT governance. *Hospital & Health Network*, *88*, 38–39.

Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Russell, J., & Potts, H. W. W. (2010). Adoption and non-adoption of a shared electronic summary record in England: A mixed-method case study. *BMJ*, *340*(jun16 4), c3111–c3111. https://doi.org/10.1136/bmj.c3111

Grogan, L. (2020). Paperless Healthcare. *World of Irish Nursing & Midwifery*, *28*(3), 44–45.

Guidance on Information Governance (p. 68). (2012). Health Information and Quality Authority.

Gunter, T. D., & Terry, N. P. (2005). The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions. *Journal of Medical Internet*

Research, *7*(1), e3. https://doi.org/10.2196/jmir.7.1.e3

Health Service Executive. (2011). *Standards and Recommended Practices for Healthcare Records Management*. HSE Ireland.

Health Service Executive, & Dept of Health. (2013). *EHealth Strategy for Ireland* (p. 80).

Hevner, March, Park, & Ram. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75. https://doi.org/10.2307/25148625

Houses of the Oireachtas. (2017). *Slaintecare Report* (p. 192). Committee on the Future of Healthcare. https://www.oireachtas.ie/en/committees/32/future-of-healthcare/

Hussey, P. A., & Kennedy, M. A. (2016). Instantiating informatics in nursing practice for integrated patient centred holistic models of care: A discussion paper. *Journal of Advanced Nursing*, *72*(5), 1030–1041. https://doi.org/10.1111/jan.12927

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Nagle, T., Sammon, D., & Cleary, W. (2019). *A New Approach to Business Value Driven Planning for Data Projects* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/jg4pr

Parsons, A., McCullough, C., Wang, J., & Shih, S. (2012). Validity of electronic health record-derived quality measurement for performance monitoring. *Journal of the American Medical Informatics Association*, *19*(4), 604–609. https://doi.org/10.1136/amiajnl-2011-000557

Recommendation of OECD Council on Health Data Governance. (2016).

Redman, T. C. (2016, July 28). Assess Whether You Have a Data Quality Problem. *Harvard Business Review*. https://hbr.org/2016/07/assess-whether-you-have-a-data-quality-problem

Stone, C. P. (2014). A Glimpse at EHR Implementation Around the World: The Lessons the US Can Learn. *Health Institute for E-Health Policy*, 12.

Urquhart, C., Currell, R., Grant, M., & Hardiker, N. (2018). Nursing record systems: Effects on nursing practice & healthcare outcomes. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD002099.pub3

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*. https://doi.org/10.1080/07421222.1996.11518099