Data Quality Trust: A Provenance-based Data Quality Assessment Method and its Integration with Interoperable Electronic Medical Record Systems

by

Jean-Philippe Stoldt

Master of Science, University of Derby, 2011

Diplom, Nordakademie University of Applied Sciences, 2004

A Dissertation Submitted in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

© Jean-Philippe Stoldt, 2022

University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Data Quality Trust: A Provenance-based Data Quality Assessment Method and its Integration with Interoperable Electronic Medical Record Systems

by

Jean-Philippe Stoldt

Master of Science, University of Derby, 2011

Diplom, Nordakademie University of Applied Sciences, 2004

Supervisory Committee

Dr. Jens Weber, Supervisor

(Department of Computer Science)

Dr. Neil Ernst, Member

(Department of Computer Science)

Dr. Morgan Price, Outside Member

(Division of Medical Sciences)

Supervisory Committee

Dr. Jens Weber, Supervisor

(Department of Computer Science)

Dr. Neil Ernst, Member

(Department of Computer Science)

Dr. Morgan Price, Outside Member

(Division of Medical Sciences)

ABSTRACT

Data quality is a critical requirement for data-driven clinical decision making in modern healthcare. It is a key prerequisite to many clinical analytics applications, yet much of the research to date focuses on assessing data quality of electronic health records for secondary use. New approaches to data quality assessment are needed that allow clinical data users to quickly assess whether the quality of diagnostic data can be trusted for a clinical decision or not. This dissertation proposes a data quality trust model and provenance-based assessment method for considering contextual data quality during clinical decision making at the point of care.

Our software engineering approach uses fuzzy logic to infer relative data quality trust from a data user's trust preferences with respect to agents, data production

methods, verification activities, and the certification of agents and data production methods. Extensions to the FHIR interface standard for data quality trust allow for platform interoperability across system contexts. We consider dual process theories to propose user interface extensions that visualize data quality trust for clinical users in heuristic and systematic cognitive processing modes. A visual prototype with an existing SMART on FHIR app for a pediatric hypertension clinical example demonstrates the feasibility of the assessment method with electronic medical record systems and clinical workflows. We show how application of data quality trust with more complex clinical examples, such as diagnosis of food or medication allergies, may enable predictive and prescriptive system functionality to guide diagnostic workflows. Trust model and data quality assessment method may be adapted for other application domains that rely on data quality for data-driven decision making.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Prof. Jens Weber, for his knowledge and support, and for his guidance, encouragement and patience throughout my research journey. Without him this dissertation would not have been possible. I am thankful to Prof. Morgan Price who has generously offered his expertise in applying research methods, explaining medical concepts, and making my research more relevant with complex clinical examples. I also owe my gratitude to Prof. Neil Ernst who provided extensive feedback that strengthened the software engineering research aspects of my project. Thank you to Prof. Liam Peyton who acted as the External Examiner for this dissertation.

Special thanks are owed to my family, my wife Jennifer, and my children Adrian and Melissa, for their support throughout the years of my doctoral program. I owe thanks to my parents who from an early age have encouraged intellectual curiosity, an appreciation for knowledge, and the pursuit of higher education.

Glossary

API Application Programming Interface. 61, 62, 66, 78, 80, 103, 111

CDA Clinical Document Architecture. 61

DQbC Data Quality by Contract. 54

EHR Electronic Health Record. 1, 4, 13–15, 17, 25, 62, 80

EMR Electronic Medical Record. 2, 3, 5–7, 9–11, 14, 25, 55, 59–62, 66–68, 70, 72, 73, 75, 80–83, 90, 92, 103, 110, 111, 117, 119, 121, 136, 138, 141, 146, 147, 149–151, 153, 155–158, 160, 161

FDA United States Food and Drug Administration. 37, 110

FHIR Fast Healthcare Interoperability Resources. 3, 60–68, 72, 75, 76, 80, 90, 92, 94, 102, 103, 107, 109, 111, 150

HIS Health Information System. 2, 3, 55, 61, 119, 157

HL7 Health Level Seven. 61

HTML HyperText Markup Language. 62, 71–73, 102

HTTP Hyper Text Transfer Protocol. 61, 111

OAuth2 Open Authorization 2.0. 62

OIDC OpenID Connect. 62

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses. 19

REST Representational State Transfer. 61, 62, 66, 80

RIM Reference Information Model. 61

SMART Substitutable Medical Applications and Reusable Technologies. 60–62, 67, 72, 73, 75, 92, 94, 102

SQL Structured Query Language. 72

W3C World Wide Web Consortium. 27, 28

Chapter 1

Introduction

Data quality is a critical requirement for data-driven clinical decision making in modern healthcare. Poor data quality leads to poor clinical decisions which in turn lead to poor patient outcomes. This chapter introduces and motivates a dissertation that seeks to provide a software engineering solution to problems with data quality assessment during clinical decision making. A clinical example provides the research context for practical illustration. We define several research objectives that drive our approach and methodology. A dissertation outline provides the road map for the following chapters.

1.1 Motivation for Data Quality Research

A growing number of devices and systems are generating large volumes of health data used by clinical analytics applications with increasing maturity [3, 4]. Interoperability initiatives are opening previously closed application ecosystems to enable exchange of Electronic Health Record (EHR) data for new data use cases [5]. Data quality considerations are particularly important with patient observations that are used in the diagnosis and treatment of some conditions. Data quality is also paramount

to ensure the safety of patients by avoiding harmful exposure from misdiagnosis or treatment errors [6]. Data quality concerns with Health Information Systems (HIS), such as Electronic Medical Record (EMR) systems, have a long history in healthcare research due to the multi-faceted data capture and management challenges inherent with the socio-technical systems of modern healthcare [7, 8, 9, 10, 11, 12].

A recent survey of healthcare executives [13] found that only 20% fully trusted their data to make timely clinical decisions, and 51% reported that data integration and interoperability were the most significant barriers to achieving their strategic priorities related to data analytics. More than half of respondents said that bad data quality had serious consequences, including ineffective or slow decision making and the inability to identify gaps in care. There are several well documented patient safety incidents that were at least in part due to underlying data quality issues [14]. Software engineering research can help address data quality assessment challenges through innovative software solutions.

1.1.1 Data Quality Assessment and Result Interpretation

Many of the data quality assessment methods proposed to date assess data quality of data item collections for secondary use in retrospect, or support the post-hoc analysis subsequent to a patient encounter [10, 11, 12]. However, those summative measures are often not suitable for clinical decision making at the point of care if they do not consider the data quality of individual data items for a decision. For example, a physician may make a patient diagnosis based on low quality measurement data or data of unknown provenance, potentially leading to misdiagnosis and incorrect treatment. New approaches to data quality assessment are needed that allow clinical data users to quickly assess whether the quality of a data item can be trusted for a clinical decision or not. Help in interpreting numerical assessment results not only for

what the numbers are but with labels of what they mean in the decision context on a good/bad scale, may increase use of information and better assist decision makers in making high quality choices [15, 16].

1.1.2 Electronic Medical Record Interoperability

Healthcare systems around the world are under continuous pressure to deliver innovative healthcare solutions while containing rising costs. Information technologies promise to improve healthcare delivery with technological advances by reducing costs through intelligent automation of labor-intensive workflows and the creation of reusable HIS components. EMR systems form the backbone of patient-centered clinical workflow environments. Many healthcare organizations limit physicians in their choice of EMR platforms due to the high cost and effort required to integrate those EMR systems with other HIS. Resulting market entry barriers prevent much needed innovation from third party applications, also referred to as "apps".

Mobile technology platforms have seen widespread consumer adoption over the last decade. Consumer app stores have developed a thriving economy for apps that extend capabilities of mobile phone platforms. Enterprise software vendors have followed suit to create similar app store experiences for many industry verticals. The healthcare software industry, and in particular EMR software vendors, have been lagging in these developments. Until the advent of Fast Healthcare Interoperability Resources (FHIR), missing data interoperability standards limited app integration points in many EMR systems. Improved interfaces, such as those implementing the FHIR standard, could help facilitate the development of apps-based information economies in healthcare to lower technology costs and spur innovation [5].

1.1.3 Electronic Medical Record User Interfaces

Cognitive aspects should be considered when visualizing system functionality that addresses data quality issues. Several recent studies have investigated causes of professional burnout among physicians [17, 18]. In one study by the Mayo Clinic [19], the usability of current EHR systems received a failing grade by physician users when evaluated using a standardized metric of technology usability. The study reported a strong dose-response relationship between system usability and the odds of burnout. These study findings highlight the importance of designing clinical user interfaces that consider usability and cognitive requirements of clinical data users at the point of care. With growing complexity of EHR-based systems, many clinical users also suffer from clinical reminder "alert fatigue" and cognitive overload. The American Medical Association identified reducing cognitive load as one of the priorities for improving usability of EHRs [19]. Focusing a data user's attention with user interfaces that address a data user's personal beliefs and contextual data quality information needs may reduce cognitive load and improve system usability for those clinical users.

1.1.4 Clinical Example

It can be difficult for software engineering researchers without medical training to grasp complex medical problems. Simple and relatable real-world clinical examples can help with the development of technical solutions to clinical data quality challenges without mounting unnecessary domain knowledge barriers. Once applied to a simple example, those solutions can then be applied to more complex medical problems. A prototype applied to a simple clinical example context may illustrate novel software engineering solutions for a broader audience. We propose pediatric hypertension as such a simple and relatable real-world example to illustrate our research project, before we move to more complex clinical examples with allergies. Different facets

of the diagnosis and treatment of pediatric hypertension can be supported through EMRs. Chapter 5 describes this clinical example in detail.

1.2 Research Objectives

The previous sections have shown that there is compelling motivation to address research challenges in data quality assessment methods, EMR platform interoperability, and cognitive user interfaces that can help improve the diagnosis and treatment of some conditions. The hypothesis of this dissertation is that it is possible to extend EMR systems with algorithms and mechanisms to reason about data quality during clinical decision making. We set three main research objectives with several sub-objectives that the dissertation seeks to address in validating our hypothesis, namely:

1. Development of a novel data quality assessment method for individual data items;

    (a) The method should provide a contextual data quality assessment result that represents the relative degree of data quality trustworthiness for individual data items;

    (b) Assessment results should be provided so they are available for reasoning about data quality during clinical decision making at the point of care;

    (c) Assessment results should have a simplified contextual result representation that adds meaning with labels that can be mapped onto a good/bad scale;

2. Developing a design pattern for the integration of assessment method with interoperable EMR systems;

(a) The method should transmit data quality assessment results for individual data items using open interface standards, while also taking into account legacy applications that cannot be upgraded with those new interfaces yet;

(b) The method should support transmission of these data quality assessment results in low network bandwidth environments to allow for a wider range of possible application settings and mobile usage contexts;

(c) The method should support visualization of assessment results taking cognitive user interface design principles into account; and

3. Demonstrating feasibility of the new approach with a visual prototype that can be integrated with the clinical workflows of a clinical example.

1.3 Validation Framework

Based on the research objectives established in the previous section, we devise a validation framework that allows for evaluation of research results against those objectives (Chapter 6). It is designed to provide validation in four major research areas: literature search, contextual data quality assessment method, integration with EMR systems, and application to a clinical example. Threats to validity for construct, internal, and external validity must be identified and mitigated to assure trustworthiness of our research results within this framework.

1.3.1 Literature Search

The first validation is met if we can demonstrate that there are no existing solutions that meet our research objectives.

1.3.2 Contextual Data Quality Assessment Method

The second validation is met if the new assessment method can generate a data quality label that provides contextual meaning to a numerical assessment result for individual data items. For validation purposes, the method must be able to generate contextual assessment results with a minimum set of three labels for data quality (for example, labels of *high*, *medium*, *low*). The system response time must be within current usability engineering guidelines.

1.3.3 Integration with EMR Systems

The third validation is met if we can demonstrate how the new data quality assessment method can be integrated with both interoperable EMR systems and legacy EMR systems. Interoperable EMR system integration must support low network bandwidth environments. Visualization of assessment results and user interface integration must adhere to cognitive user interface design principles.

1.3.4 Application to Clinical Example

The fourth validation is met if we can demonstrate that the new data quality assessment method can be integrated with clinical workflow and visual user interface of a clinical example application. This should be shown with a visual prototype and a detailed data example that provides descriptive validation of how the novel data quality assessment approach can be effectively applied to the clinical example context.

1.4 Research Approach and Methodology

Software engineering research provides the methodological foundation for this dissertation project. An overview of typical research questions addressed by software

engineering research includes questions such as *How can I evaluate the quality/correctness of X?* leading to the development of a method for analysis or evaluation, or *Does X even exist, and if so what is it like? Is it possible to accomplish X at all?* which can be answered with a feasibility study or exploration [20]. While the research paradigm is not always clearly communicated in software engineering research papers [20], Greg et al. [1] describe software engineering research with a socio-technologist / developmentalist paradigm which creates knowledge by *conceptualizing, designing, building of prototypes (as proof-of-concept, proof-by-demonstration) and/or by formal (mathematical, logical) proofs and descriptions.* Figure 1.1 illustrates the three phases of this software engineering research methodology: the conceptual, the formal, and the developmental phase. Software engineering research must address issues in at least two of the three phases to qualify as rigorous research.

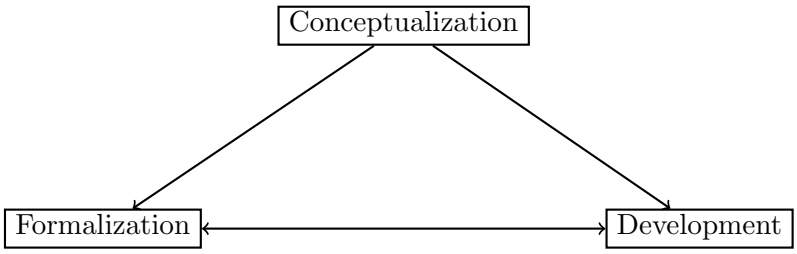


Figure 1.1: Software Engineering Research Framework [1]

1.4.1 Software Engineering Research Guidelines

Montesi et al. [21] classified and described research article types published in leading software engineering journals and conferences to provide guidance on what constituted good research in software engineering. While empirically evaluated software engineering research was growing and generally preferred, there was a perceived lack of clear and comprehensive guidelines on how to write a good empirical software engineering research paper. Relevance for practice, novelty, and originality were the most

important requirements for publication. Shaw found that in software engineering research, a slice of life example was likely to be convincing research evidence, especially if accompanied by an explanation of why the simplified example retained the essence of the problem being solved [20]. The highest quality software engineering research creates novel conceptualizations which are formally defined or prototyped [1].

1.4.2 Validation Research Approach

Wieringa [22, 23] proposed that the validation of artifacts through software engineering research should be contrasted from post-implementation evaluation research. Evaluation research is described as field research of the properties of implemented artifacts. In evaluation research, the benefit of hindsight and the experience of external stakeholders with the implemented artifact allows for the subsequent improvement of design theories about the artifact. In contrast, the goal of validation research is to predict how an artifact would interact with its context, without actually observing the implemented artifact in a real-world context. Validation research is experimental and usually done in the laboratory. It involves exposure of an artifact prototype to various scenarios presented by a model of the context, to demonstrate how the prototype would respond. Frequently used research methods in validation research are modeling, simulation, and testing. Wieringa describes "single-case mechanism experiments" as experimental methods for validation research [22, 23].

The approach to the research objectives adopted by this dissertation is to conceptualize a model and assessment method, to integrate this conceptualization with an existing domain knowledge base (EMR systems), and to demonstrate feasibility of the conceptualization through a prototype that is applied to a clinical example context.

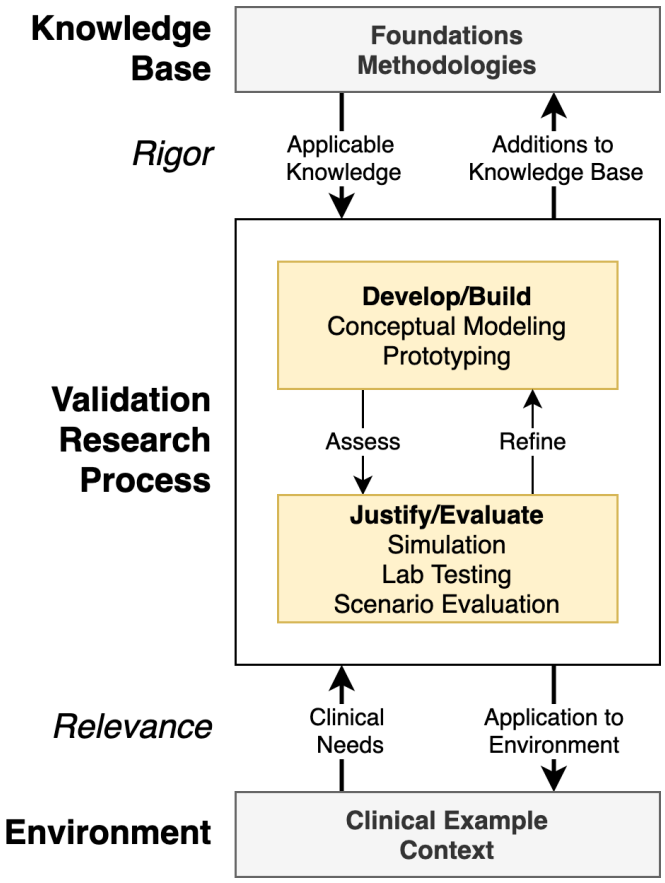


Figure 1.2: Research Methodology and Iterative Research Process

1.4.3 Research Methodology and Iterative Research Process

We have adapted an existing Design Science research framework [24] to create our mixed methods methodology that will guide iterative development and evaluation of artifacts, while assuring relevance and rigor in conducting the research. A search for applicable knowledge in foundations and methodologies from an existing knowledge base assures research rigor in Chapter 2. The core develop/build and justify/evaluate validation research cycle generates a novel model and assessment method based on existing knowledge in Chapter 3. Chapter 4 draws applicable knowledge about EMR systems from the knowledge base to develop/build and justify/evaluate a design pat-

tern that is iteratively assessed and refined. Chapter 5 assures relevance of the novel design pattern by applying it to the environment of a clinical example context. Chapter 6 evaluates the overall research results of this dissertation before we summarize research contributions to be added back to the knowledge base in Chapter 7. Figure 1.2 illustrates our research methodology and iterative research process.

1.5 Dissertation Outline

Our research approach and methodology drive the organisation of this dissertation. It is organized in seven chapters as follows.

Chapter 1 introduced and motivated this research project. A research approach and methodology was selected to address several research objectives.

Chapter 2 presents a literature review and search for existing solutions which confirms that a novel approach is required to meet the research objectives.

Chapter 3 proposes a novel data quality assessment approach through a data quality trust core model and assessment method. Two extensions to core model and assessment method provide additional functionality to address the research objectives.

Chapter 4 transfers model and method to the EMR system domain with a design pattern. It shows data integration with both interoperable and several legacy EMR systems are possible, and it develops visualization guidance for extending EMR system user interfaces with data quality trust.

Chapter 5 applies data quality trust to a clinical example context with a detailed computation example and visual prototype. It demonstrates feasibility of integrating the data quality trust approach with an existing clinical workflow and

several usage scenarios. We show how data quality trust could also be applied to more complex clinical examples.

Chapter 6 evaluates our research against several validation criteria. It discusses how threats to construct, internal, and external validity have been addressed. Several research challenges and limitations are noted.

Chapter 7 summarizes our research contributions. The dissertation concludes with a research outlook and suggestions for future work.

The next chapter will present a literature review and search for existing solutions which confirms that a novel approach is required to meet the research objectives.

Chapter 2

Literature Review

This chapter describes a literature review of data quality assessment methods. Before a solution to the research objectives can be developed, a review of existing concepts and solutions is required to place the research project in context. After a narrative review of data quality in the context of clinical decision making with EHRs, we introduce the concepts of data plausibility, believability and trustworthiness, and review prior work with data trustworthiness in other domains. Then, we report on a scoping review of existing data quality assessment solutions and data quality research in the field of EHRs. The review confirms the need for a novel approach to contextual data quality assessment during clinical decision making. As part of the rigor cycle of our research methodology and iterative research process (Figure 1.2), we draw applicable knowledge in foundations and methodologies from an existing knowledge base. Main research methods embedded in this chapter are narrative literature review, systematic literature search, and scoping literature review.

2.1 Data Quality of Electronic Medical Records

Data quality of EMR data can be characterized as the "fitness for use" [25] of a data item for a particular data use case. During clinical decision making, this fitness for use relates to how well an individual EMR data item can support a clinical decision made by a clinical data user. Data quality is particularly important for diagnostic data since this data is used throughout the health system and acts as a prompt for other interventions within individual patient consultations [25]. The terms EMR and EHR are often used interchangeably to describe digital health records in the literature. We use the term EMR to describe the computer-based patient record specific to a single clinical practice, such as a family health team or group practice. We use the broader term EHR to describe an integrated computer-based patient record collection of a person's encounters with the health care system with a comprehensive view of a patient's health history.

Data quality research focused on EHRs has seen a number of proposed models, quality dimensions and assessment methods over time [10]. Weiskopf and Weng proposed a harmonized data quality assessment terminology and framework for the secondary use of EHR data with five quality dimensions that were empirically derived from the literature (completeness, correctness, currency, plausibility, and concordance) [11]. Their definitions are shown in Table 2.1.

Table 2.1: Harmonized Data Quality Dimensions (Weiskopf and Weng, 2013)

Dimension	**Definition**
Completeness	Is a truth about a patient present in the EHR?
Correctness	Is an element that is present in the EHR true?
Concordance	Is there agreement between elements in the EHR, or between the EHR and another data source?
Plausibility	Does an element in the EHR make sense in light of other knowledge about what that element is measuring?
Currency	Is an element in the EHR a relevant representation of the patient state at a given point in time?

Kahn and colleagues further defined three data quality categories (conformance, completeness, plausibility) with two data quality assessment contexts (verification, validation) [10]. Many of the models, measures and assessment methods for EHR data quality have in common that they focus on the data quality of entire datasets for quality improvement activities related to patient care, and secondary purposes such as research and disease surveillance [26, 12].

2.2 Data Quality during Clinical Decision Making

Clinical decision making requires that a healthcare provider can rely on and trust the data quality of individual data items used in the decision context. Different data users may assess the data quality of the same data item differently, subject to their personal beliefs and contextual data use case requirements. Data believability can be defined as "the extent to which data are accepted or regarded as true, real and credible"

[27]. There are several aspects that make a contextual data quality assessment more suitable for decision making at the point of care, compared to data quality dimensions which seek to assess data based on intrinsic or absolute measures alone. The concept of data trustworthiness, or trust in the data quality of a data item, can help explain and quantify a data user's beliefs and perceptions associated with contextual data quality. Weiskopf and Weng [11] harmonized the literature terms "Believability" and "Trustworthiness" as a part of the plausibility data quality dimension (Table 2.1).

2.3 Data Believability and Trustworthiness

Provenance has been used to verify trustworthiness or data quality in several research domains. Golbeck combined provenance with trust in social networks [28]. In her research, the semantic web of trust requires that users describe their beliefs about others. Based on a person's list of who they know and how much they trust them, social network information can be automatically compiled and processed. Bertino and Lim proposed a provenance trust model that evaluates trustworthiness of data and data providers [29]. Their work represents a comprehensive framework for assuring information trustworthiness, based on the concepts of trust scores and confidence policies. Sun et al. proposed a method to assess trustworthiness of medical data based on provenance [30]. Rajbhandari et al.[31] examined how provenance information is associated with a workflow in a bio-diversity application. Their definition of trust is concerned with trustworthiness of an outcome. This stands in contrast with other approaches that see trust as subjective information or objective opinions formed on the basis of factual evidence or recommendation by arbitrating authorities. In separate work, Rajbhandari et al. [32] proposed a fuzzy model for calculating workflow trust using provenance data. Richardson and colleagues developed a belief combination

function consistent with fuzzy logic that aggregates trust in a web of trust, in which each user trusts a small number of other users based on personal beliefs [33]. Prat and Matnick [34] propose to measure data believability based on trustworthiness of the source, the reasonableness, and the temporality of data. Trustworthiness in an agent for a domain is measured by a trustworthiness value, normalized between 0 and 1. They suggest a more subjective and user-centered assessment of believability as a natural extension to their research.

2.4 Scoping Review of Automated Data Quality Assessment Methods

Retrospective analysis of EHR data quality has been much of the focus of data quality assessment research to date. In contrast, real-time data quality assessment of individual data items using automated methods has only received limited attention. To validate this hypothesis and to assure novelty of a new approach, we conducted a literature search for existing models and methods that allow for automated data quality assessment of individual EHR data items.

2.4.1 Search Objective and Research Questions

The objective of this scoping literature review was to find evidence of automated data quality assessment methods applied to EHR systems. Our literature review questions were:

1. Does the existing literature provide evidence for the application of automated methods for data quality assessment of EHR system data?
    a. What are the data quality dimensions assessed?

b. Do the assessment methods provide real-time, retrospective, or prospective assessments?

c. Are the assessment methods suitable to provide contextual data quality assessment of individual data items during clinical decision making?

d. How have contextual data quality assessment results been visualized?

2.4.2 Scoping Review Method

The Scoping Literature Review [35] provides a flexible methodology to conduct exploratory research into a topic of interest to identify gaps in the literature and to aid the planning of future research. We also incorporated guidelines for performing systematic literature reviews in software engineering by Kitchenham and Charters [36] and the literature review tool Parsifal. Carrying out our scoping review consisted of five stages: 1) Identifying the research questions; 2) Identifying relevant studies; 3) Study selection; 4) Charting the data; and 5) Collating, summarizing, and reporting the results. Tables A.2, A.3 and A.4 describe the criteria used for this review. Table A.5 lists the articles selected for full-text review.

2.4.3 Search Strategy

Given the interdisciplinary nature of our research in the engineering and medical research domains, we selected MEDLINE, INSPEC and EI COMPENDEX databases to balance research effort with search coverage of the domains. MEDLINE was searched through PubMed, and both INSPEC and EI COMPENDEX were searched through the EngineeringVillage search interface. Both interfaces allow for the use of controlled search vocabularies and filtering of results by publication date. The initial search strategy focused on controlled vocabularies and medical subheadings (MeSH)

for the terms "Data Quality" and "Electronic Health Records" but these searches yielded incomplete results. In MEDLINE, there was no defined MeSH term for data quality and a MeSH term "Data Accuracy" was only introduced in 2016. Table A.1 summarizes search results of some of the exploratory database queries made with varying breadth and search term restrictions. We decided that a broader search query and a manual review of a larger number of articles would limit the risk of missing important research. Finally, we conducted our search in PubMed for articles with publication years 2013-2020 that were classified with MeSH term "medical record systems, computerized" and that had the term "data quality" in their title or abstract. The resulting PubMed query *("Data Quality" [TIAB]) AND (medical record systems, computerized [Mh])* returned 293 results in July 2020. In EngineeringVillage, we searched for articles that were classified with the controlled vocabulary term "electronic health records" and that had the term "data quality" in their title or abstract. Restricting the search to publication years 2013-2020 to mirror our Pubmed search did not reduce the number of articles found. The resulting EngineeringVillage query '*("Data Quality") WN KY AND (electronic health records WN CV)*' returned 66 articles in July 2020. All references were imported into the literature management program Zotero. After excluding 31 duplicate article references, 328 journal articles and conference papers with abstracts were selected for relevance screening of title and abstracts. 262 papers did not meet inclusion criteria for various reasons leaving 66 papers for full-text screening. 39 papers were selected for a detailed review of automated data quality assessment methods and their visualization. None of the papers reviewed described methods for automated data quality assessment in real-time based on believability or trustworthiness of electronic health record data. Figure 2.1 visualizes our literature review stages using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) format [37].

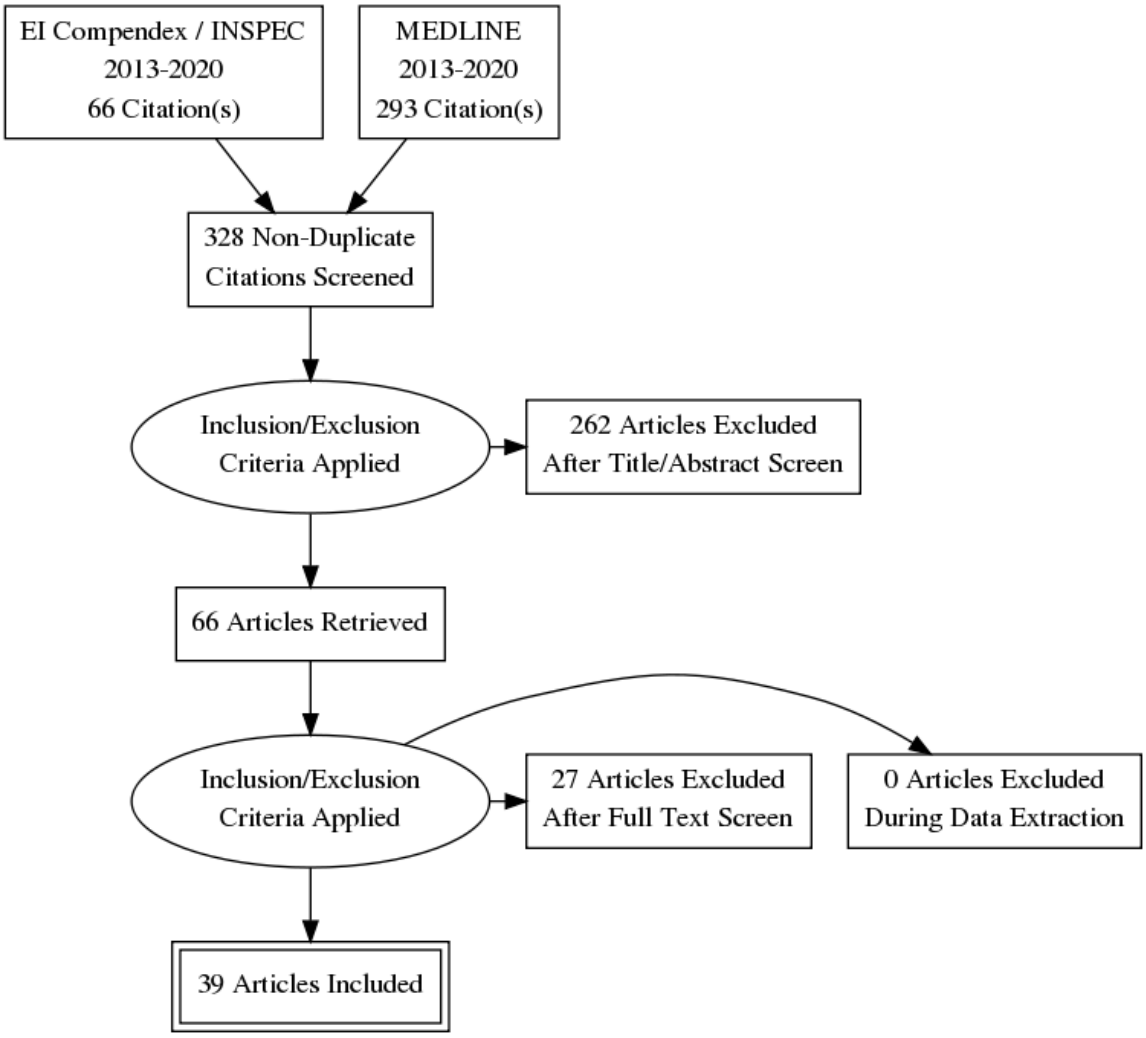


Figure 2.1: PRISMA Diagram for Systematic Literature Search

2.4.4 Search Results

As part of the article data characterization and extraction stage, we grouped the articles by harmonized data quality dimensions that they covered (Table 2.1).

Data Completeness and Correctness

Many of the articles reviewed during full-text screening dealt with data quality dimensions of completeness or correctness. A majority of these articles describe methods to assess record completeness and value conformance. The methods described data quality assessment of data collections or data quality assessment based on preset absolute measures which did not allow for the contextual data quality assessment of individual data items. Articles that only described dimensions of completeness or conformance were excluded from further data extraction during the full-text review stage. Instead, articles that described data quality assessments of plausibility, concordance, provenance or trustworthiness were deemed more suitable for contextual data quality assessment of individual data items in specific use case contexts. Table A.5 provides the articles selected for full-text review. The following sections briefly summarize articles that have contributed ideas to a novel data quality assessment method described in subsequent chapters.

Data Concordance

Three articles described automated methods for assessing the concordance of data. Brown et al [38] provided best practices for data quality checking in comparative effectiveness research in distributed research networks. They report that the FDA's Mini-Sentinel project conducts data quality checking on multiple levels. Advanced checks focus on patterns, trends and cross-variable relationships. Unusual patterns or cross-site variability can be identified with a graph of monthly pharmacy dispensations

or a table of the annual distribution of inpatient and outpatient encounters. Deng et al. [39] describe validity checks of temporal concordance between recorded hospital movements and EHR timestamps. Research by Horth et al. [40] provides an approach to record matching and concordance testing for structured and unstructured data between EHR and healthcare information exchange databases.

Data Plausibility

Nine papers described automated methods for assessing the plausibility of data. Daymont et al. [41] developed an algorithm to identify weight and height values that should be excluded from data analysis in pediatric EHR growth data. The method uses a standard deviation score to compare measurements with a weighted moving average of a child's other measurements to identify implausible values and values that were recorded repeatedly without separate measurement. Estiri and Murphy evaluated the utility of data encoding for outlier detection to determine plausibility of clinical observation data. Their research compared the outlier detection performance of multiple neural network-based encoding methods using semi-supervised encoding with silver-standard distributions as training data [42]. Estiri et al. [43] developed a unsupervised clustering-based anomaly and outlier detection approach for detecting implausible observations in EHR data as an alternative algorithmic solution to existing standard deviation and Mahalanobis distance approaches. Their evaluation with EHR laboratory test results showed high levels of specificity and sensitivity for the new method. Huser et al. [44] reported on the Achilles tool from the Observational Health Data Sciences and Informatics Consortium (OHDSI) that performs data characterization and rules-based data quality checks to test model conformance, data completeness and data plausibility. However, the article mentions plausibility checks but it does not describe how exactly plausibility is automatically assessed. Kapsner et

al. [45] described the MIRACUM data quality assessment approach for conformance, completeness and plausibility analysis based on OMOP and i2b2 databases. To assess plausibility, they relied on domain expert knowledge of hospital data managers and physicians to define relationships between data elements for automatically checking respective values against defined constraints. A semi-automated data quality assessment method for conformance, completeness and plausibility was reported by Khare et al. [46]. Their approach checked data plausibility through nine data quality checks that probed for unexpected facts, implausible events, numerical outliers, implausible dates, categorical outliers, temporal outliers, implausible distributions, unexpected most frequent values, and unexpected differences compared to previous data cycles. Pezoulas et al. [47] developed an automated framework for medical data curation that involves a three step data curation process. Data evaluation and data quality control stages have anomaly and similarity detection features to evaluate plausibility of values. The method is focused on curating datasets for secondary use and has limited use for assessing individual data items. Sirgo et al. [48] compared their automatic ICU-DaMa queries to assess conformance, completeness and plausibility. Upon further review of the article, it does not contain a detailed description of a plausibility assessment method. Stoldt and Weber [49] describe the application of FHIR-enabled data quality probes that check for statistical plausibility of data for safety assurance. Ta et al. [50] proposed automated data quality assessment of plausibility by normalizing and clustering concept frequencies with the K-means data mining algorithm.

Plausibility with Data Provenance and Trustworthiness

Three papers described automated methods for assessing data quality based on data provenance or trustworthiness. McCoy et al. developed and evaluated a clinician reputation metric to facilitate the identification of appropriate problem-medication

pairs. They found that the clinician reputation metric achieved a high specificity when used to identify problem-medication pairs generated through a crowd-sourcing methodology [51]. Stoldt and Weber [49] proposed data quality probes to assess intrinsic-meta data quality with FHIR resources. They write that provenance of a diabetes data item in the active problem list can be probed by checking where the data originated, the procedure how the data was obtained, or currency of the data. Sun et al [30] developed a provenance-based assessment method to calculate data trustworthiness from the data itself (assessed through measures of accuracy, completeness, validity, and conflict), trustworthiness of its data source, and trustworthiness of data transmission intermediaries.

Visualization of Data Quality Assessment Results

Several articles describe methods for visualizing the results of data quality assessments but none of the articles provided guidance on how contextual data quality assessment results could be visualized. However, several articles provided ideas for visualization elements that could be applied to contextual data quality assessment results. Alvarez Sanchez et al. [52] report on TAQIH, a visual tool for exploratory data analysis and data quality assessment, that employs a missing value heatmap, a density plot for correlation analysis, and outlier detection. TrialViz visualizes queries for completeness and correlation of variables in EMR databases [53]. Noselli et al. describe a visual tool to profile completeness and correctness data quality measures [54]. Process-mining of patient control-flows can be visualized to highlight data quality issues, as shown by Perimal-Lewis et al. [55].

2.4.5 Discussion of Literature Search Results

Most of the articles selected for the full-text screening stage covered completeness or conformance data quality dimensions for data collections in retrospective. From the articles with automated methods selected for full-text review, nine articles described methods for plausibility assessment, three articles described methods for assessing data concordance, and only three articles described methods to assess contextual data quality of individual data items based on provenance or trustworthiness. One method by Sun et al. [30] showed a promising approach to assess data quality based on trustworthiness of research data in retrospective. However, this method described secondary use of EHR data and required additional development and refinement to meet our objectives of real-time use during clinical decision making. Most data quality assessment methods reviewed did not describe visualization of data quality assessment results in sufficient detail. A key limitation of the review was that in the context of this dissertation project it was conducted by a single author only which may have introduced researcher bias. Additional authors could have improved the review process through independent verification during article selection and data extraction, as confirmed through measures of interrater reliability.

2.5 Summary

This chapter described a literature review of data quality assessment methods in the context of clinical decision making with EMRs. The concepts of data provenance and trustworthiness were introduced. A scoping review of existing data quality assessment solutions and existing data quality research in the field of EHRs confirmed the need for a novel approach to providing contextual data quality assessment results that are available during clinical decision making.

The next chapter will propose a novel data quality assessment approach through a data quality trust core model and assessment method. Two extensions to core model and assessment method will provide additional functionality to address the research objectives.

Chapter 3

Model and Assessment Method for Data Quality Trust

This chapter conceptualizes a novel approach to data quality assessment that estimates data quality of individual data items based on their provenance and a data user's trust in individuals and methods used in generating the data. It consists of a core model representation of provenance-based data quality trust, an automated method to assess trust, and two model and method extensions for verification and certification features. As part of the validation research process (Figure 1.2), this chapter represents the core develop/build and justify/evaluate iterative research cycle which generates a novel model and assessment method based on existing knowledge. Main research methods embedded in this chapter are conceptual modeling, prototyping, simulation, and lab testing.

3.1 Data Quality Trust Model

We begin with a description of our core model representation of provenance-based data quality trust. The data provenance standard introduced by the World Wide

Web Consortium (W3C) [2] forms the basis of our novel approach. W3C models the provenance of data based on the interrelationship of agents, activities and entities, as shown in Fig. 3.1. We align our model terminology with terms used by W3C.

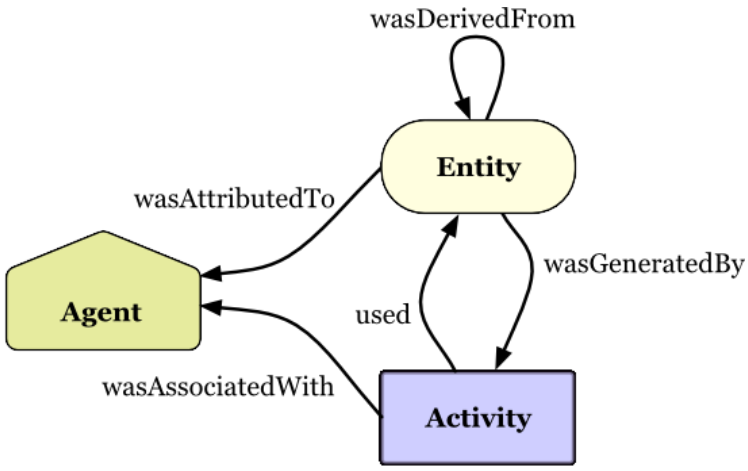


Figure 3.1: W3C Provenance Model [2]

A transformation activity occurs when an agent applies a data transformation method to generate a new data entity. The new data entity may be derived from any number of antecedent data entities. It may also be generated without antecedents. Figure 3.2 illustrates a basic transformation activity.

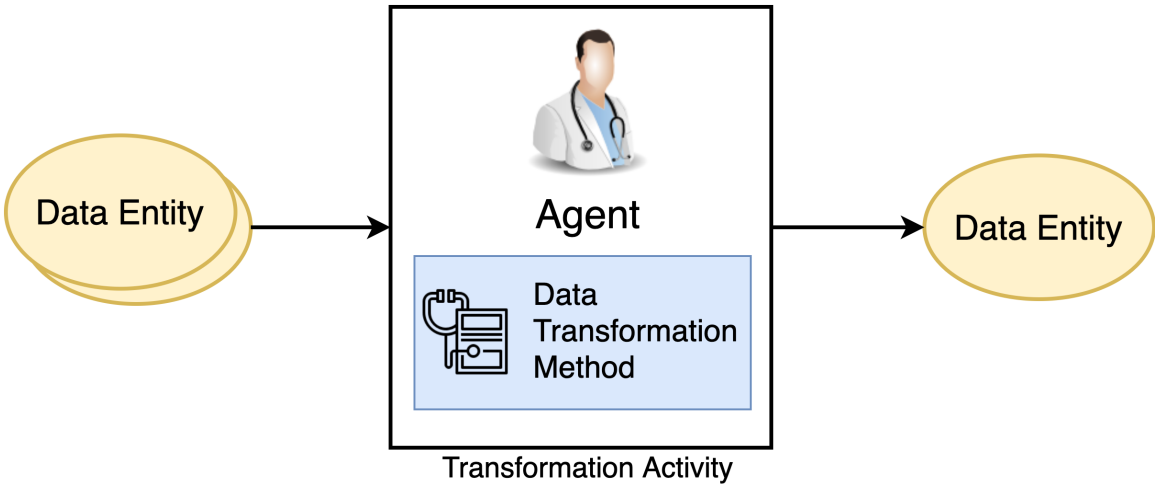


Figure 3.2: Illustration of Data Transformation Activity

3.1.1 Data Quality Trust

We define **Data Quality Trust** as **a proxy measure to capture the relative degree to which a data entity is able to satisfy a data user's contextual data quality requirements and beliefs**. In the context of this dissertation, it is based on who generated the data entity (agent), how it was generated (transformation method), and what data entities it was derived from (antecedents). Figure 3.3 illustrates how a data user's trust in the quality of a given data entity d can be inferred from agent trust in the individual that generated d, and reliability trust in the data production method that generated d. The inferred data quality trust is constrained by the antecedent data entity with the lowest data quality trust. Set inference rules define the association between agent trust, reliability trust, data quality trust of antecedents, and the resulting data quality trust for the data entity.

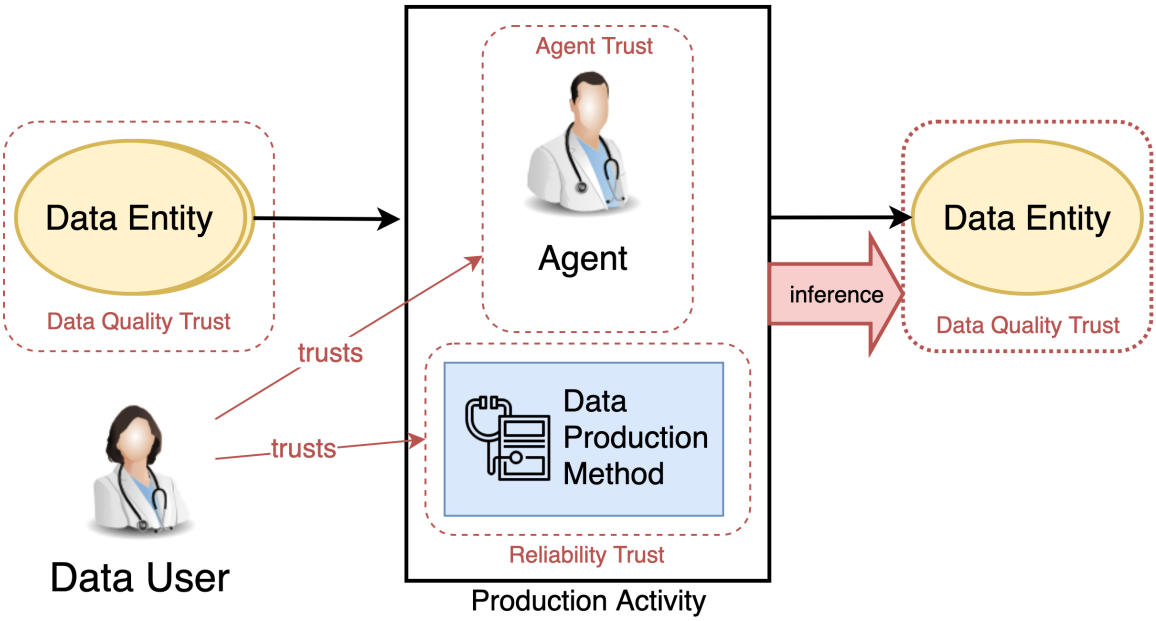


Figure 3.3: Illustration of Data Quality Trust Inference

3.1.2 Trust Representation with Fuzzy Sets

It is difficult to quantify levels of trust in traditional numerical value sets. Fuzzy theory can numerically quantify uncertain variables through fuzzification [56]. Given the uncertainty associated with how varying levels of trust are perceived by different data users, a representation of relative trust levels with linguistic variables and fuzzy sets is appropriate. Computing with linguistic variables can also be applied to inference rules that propagate the constraints from premises to conclusions in fuzzy logic [57]. For a given fuzzy set of the linguistic variable T for trust, the membership function μ_T quantifies the degree of membership for a base variable t to T within the interval $[0, 1]$. A trapezoidal membership function has the ability to approximate other functions, and to store both triangular and trapezoidal membership function types. The membership function $\mu_T(t)$ describes fuzzy membership of trust for the

labels *low*, *medium*, *high*. Eq. 3.1 shows how different values for parameters p_{1-4} characterize μ_T for every trust label in T.

$$\mu_T(t, p_{1-4}) = \begin{cases} 0, & \text{if } (t < p_1) \quad \text{or} \quad (t > p_4) \\ \dfrac{t - p_1}{p_2 - p_1}, & \text{if } p_1 \leq t \leq p_2 \\ 1, & \text{if } p_2 \leq t \leq p_3 \\ \dfrac{p_4 - t}{p_4 - p_3}, & \text{if } p_3 \leq t \leq p_4 \end{cases} \tag{3.1}$$

There are several techniques for generating membership functions for a particular set of trust labels and usage contexts which may involve heuristics, or a detailed user study and pattern analysis of study results [58]. A simple trust representation with three triangular or trapezoidal membership functions for labels *low*, *medium*, *high* (Figure 3.5) can help illustrate the concepts proposed in this dissertation.

3.1.3 Trust in the Capability of Agents

Agents vary in their capabilities. We define capability as the degree to which an agent is able to perform a data transformation method correctly to generate a new data entity. We express agent trust as a fuzzy set A with the labels *low*, *medium*, *high*. Agent trust in our model is a relative trust level assigned by a data user. A membership function μ_A maps the base agent trust variable a to overlapping trust labels in A.

3.1.4 Trust in the Reliability of Production Methods

Data production methods vary in their reliability. We define reliability as the degree to which a production method can be depended on for precision, and correctness or accuracy. Reliability trust in our model is a relative trust level assigned by a data user. We express reliability trust as a fuzzy set R with the labels *low*, *medium*, *high*. A membership function μ_R maps the base reliability trust variable r to overlapping trust labels in R.

3.1.5 Simplified Contextual Assessment Result

A trust computation with fuzzy logic enables contextual data quality assessment with simplified contextual results as represented by trust labels. One of the distinct features of fuzzy logic at the core of model and assessment method is that it allows for "Computing with Words" [59, 57]. Rather than providing a detailed numerical result, data quality trust can be aggregated to trust labels such as *low*, *medium*, *high*, hiding complexity from the user. Other visualizations, such as traffic light representations with labels *red*, *yellow*, *green*, may also be suitable for particular usage contexts.

3.1.6 Provenance Trees and Multi-level Data Provenance

Our data quality trust model is not limited to the single level data provenance discussed so far. A provenance tree organizes data entities and their trust levels in a multi-level network of data transformations that ultimately generate data entity d. The root node is the data entity d. Leaf nodes are antecedent data entities to d that were not derived from other data entities. Nodes are antecedent data entities to d that were derived from other data. They have one or more leafs or nodes as antecedents.

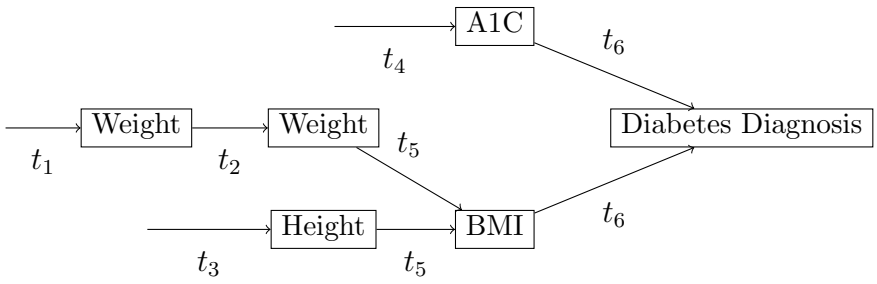


Figure 3.4: Data provenance tree for diabetes diagnosis

In a decision tree, each path from the tree root to a leaf corresponds to a conjunction of attribute tests, and the tree itself is a disjunction of these conjunctions. Diagnosis of clinical conditions can be modeled as such a decision tree where a disjunction of several conjunctions of constraints on decision attributes describes multiple decision options. In contrast, a data provenance tree in our model represents only a single path through the decision tree where all node attributes form part of the conjunction of antecedent data entities.

Fig. 3.4 shows a provenance tree example for a diabetes diagnosis data entity which was derived from several A1C, weight, height and BMI data entities and their transformations. Transformations from leaf nodes t_1, t_3 and t_4 are patient observations made without antecedent data entities. Node transformation t_2 is a data transformation that modified antecedent data entity properties to generate a new data entity. This could be a data transmission between different system contexts. Node transformation t_5 used antecedent weight and height data entities to calculate and generate a new BMI data entity. Root node transformation t_6 generates a physician diagnosis data entity using antecedent A1C and BMI data. Each transformation t_{1-6} consists of an agent and a data transformation method that produce a new data entity.

3.2 Data Quality Trust Assessment Method

Based on our core model representation of data quality trust, we now describe an automated method that infers provenance-based data quality trust for a data entity from trust in production methods and agents that generated the data entity.

3.2.1 Fuzzy Association Rules

Fuzzy association rules define the relation between production method reliability trust and agent trust, and the resulting data quality trust for a given data entity. Our model differentiates between three types of associations depending on the level of production method automation: fully automated, semi-automated, and manual data production. These three types are sufficient to represent data production by highly regulated automated medical devices, by medical devices supporting human reasoning, and by human reasoning that does not require a device to produce data. A manual data production method is most reliant on the capability of the performing agent, whereas a fully automated data production method is least reliant on the capability of the performing agent.

Associations in our model take the form of IF-THEN rules. For example, *IF (Reliability Trust is low) AND (Agent Trust is low) THEN (Data Quality Trust is low)*. The number of rules is the product of the number of labels in each input set, nine fuzzy association rules result in the data quality trust labels *high*, *medium*, *low*. On the surface, these rules may resemble associations found in crisp logic, but they represent associations of fuzzy sets and follow fuzzy logic.

Fully-Automated Data Production

A fully-automated data production method is less reliant on the capability of the performing agent. Table 3.1 defines nine fuzzy association rules.

- Rules 1-3 cover data production methods with high reliability trust. An agent with medium or lower capability can still generate high data quality trust.
- Rules 4-6 cover data production methods with medium reliability trust. An agent with lower capability can still generate medium data quality trust.
- Rules 7-9 cover data production methods with low reliability trust. Regardless of agent capability level, this data production activity generates low data quality trust.

Table 3.1: Fuzzy Association Rules for Fully-Automated Production Method

Rule	**Reliability Trust**	**Agent Trust**	**DQ Trust**
1	high	medium	high
2	high	high	high
3	high	low	high
4	medium	medium	medium
5	medium	high	medium
6	medium	low	medium
7	low	medium	low
8	low	high	low
9	low	low	low

Semi-Automated Data Production

A semi-automated data production method is equally reliant on the capability of the performing agent, and on the reliability of the data production method. Table 3.2 defines nine fuzzy association rules. Rules 1-9 generate data quality trust based on the minimum of reliability trust and agent trust.

Table 3.2: Fuzzy Association Rules for Semi-Automated Production Method

Rule	**Reliability Trust**	**Agent Trust**	**DQ Trust**
1	high	medium	medium
2	high	high	high
3	high	low	low
4	medium	medium	medium
5	medium	high	medium
6	medium	low	low
7	low	medium	low
8	low	high	low
9	low	low	low

Manual Data Production

A manual data production method is mostly reliant on the capability of the performing agent. Table 3.3 defines nine fuzzy association rules. In contrast to fully automated data production, data quality trust in manual data production reflects the trust level of the performing agent.

Table 3.3: Fuzzy Association Rules for Manual Production Method

Rule	**Reliability Trust**	**Agent Trust**	**DQ Trust**
1	high	medium	medium
2	high	high	high
3	high	low	low
4	medium	medium	medium
5	medium	high	high
6	medium	low	low
7	low	medium	medium
8	low	high	high
9	low	low	low

Measurement Device Classification

Depending on the classification of measurement devices used for automation of the data production method, different association rules may apply that reflect how much a method relies on the capability of the performing agent. National regulators may provide such guidelines for classification of measurement devices.

The United States Food and Drug Administration (FDA) classifies medical devices into Class I, II, and III. Regulatory control increases from Class I to Class III and the device classification regulation defines the regulatory requirements for each device class. Device classes II and III have the highest level of regulatory control for devices. These devices are fully automated. Device class I is subject to a lower level of regulatory control. These devices are semi-automated since they also rely on the capability of the device user for functionality. Devices that are not subject to medical device classification support manual methods that only rely on the capability of the

performing agent. For example, blood pressure measurement devices are regulated under *21 CFR 870 (B) Cardiovascular Diagnostic Devices* and *21 CFR 870 (C) Cardiovascular Monitoring Devices*. Table 3.4 provides the classification of several blood pressure measurement devices and we map our suggested fuzzy association rules based on the device classification. If no device is used, then data quality trust is only based on the capability of agent that performs the measurement (manual measurement). A blood pressure measurement made with a manual stethoscope requires the agent to be capable of correctly using the stethoscope, and it requires the stethoscope to be reliable (semi-automated measurement). Data quality trust of a blood pressure measurement made with an electronic medical device is mostly based on the reliability of that medical device (fully-automated measurement).

Table 3.4: Mapping of Fuzzy Association Rules to FDA Device Classification

Device Type	**Regulation**	**Class**	**Association Type**
No device	-	-	Manual Method
Stethoscope (Manual)	§ 870.1875	I	Semi-Automated Method
Stethoscope (Electronic)	§ 870.1875	II	Fully-Automated Method
Blood Pressure Cuff	§ 870.1120	II	Fully-Automated Method
CNAP	§ 870.1130	II	Fully-Automated Method
Oscillometer	§ 870.2675	II	Fully-Automated Method

3.2.2 Trust Inference System

Fuzzy inference allows a fuzzy inference system to aggregate individual fuzzy sets to a fuzzy result set through approximate reasoning. Trust in the data quality of a data entity is the aggregation of trust in the reliability of transformation method, and trust in the capability of performing agent. Process steps required for fuzzy

inference are fuzzification of input sets, inference based on the fuzzy association rules of the data production method (manual, semi-automated, fully-automated), and the defuzzification of output sets. The min-max method proposed by Mamdani [60] has been widely adopted for its simplicity in reasoning and graphical expression. Applying fuzzy inference to two trust variables r for reliability trust and a for agent trust, the input fuzzy sets R (Reliability Trust), and A (Agent Trust), with the membership degrees $\mu_A(a)$ and $\mu_R(r)$ of the trust input values a and r respectively, the Mamdani method provides the output fuzzy set D (Data Quality Trust). Fig. 3.5 shows examples for trust membership functions for $\mu_A(a)$, $\mu_R(r)$, and $\mu_D(d)$. In this example for a semi-automated data production method, we show a possible combination of triangular and trapezoidal membership functions associated with trust labels *low*, *medium*, *high* to visualize the fuzzy result shape resulting from such an inference.

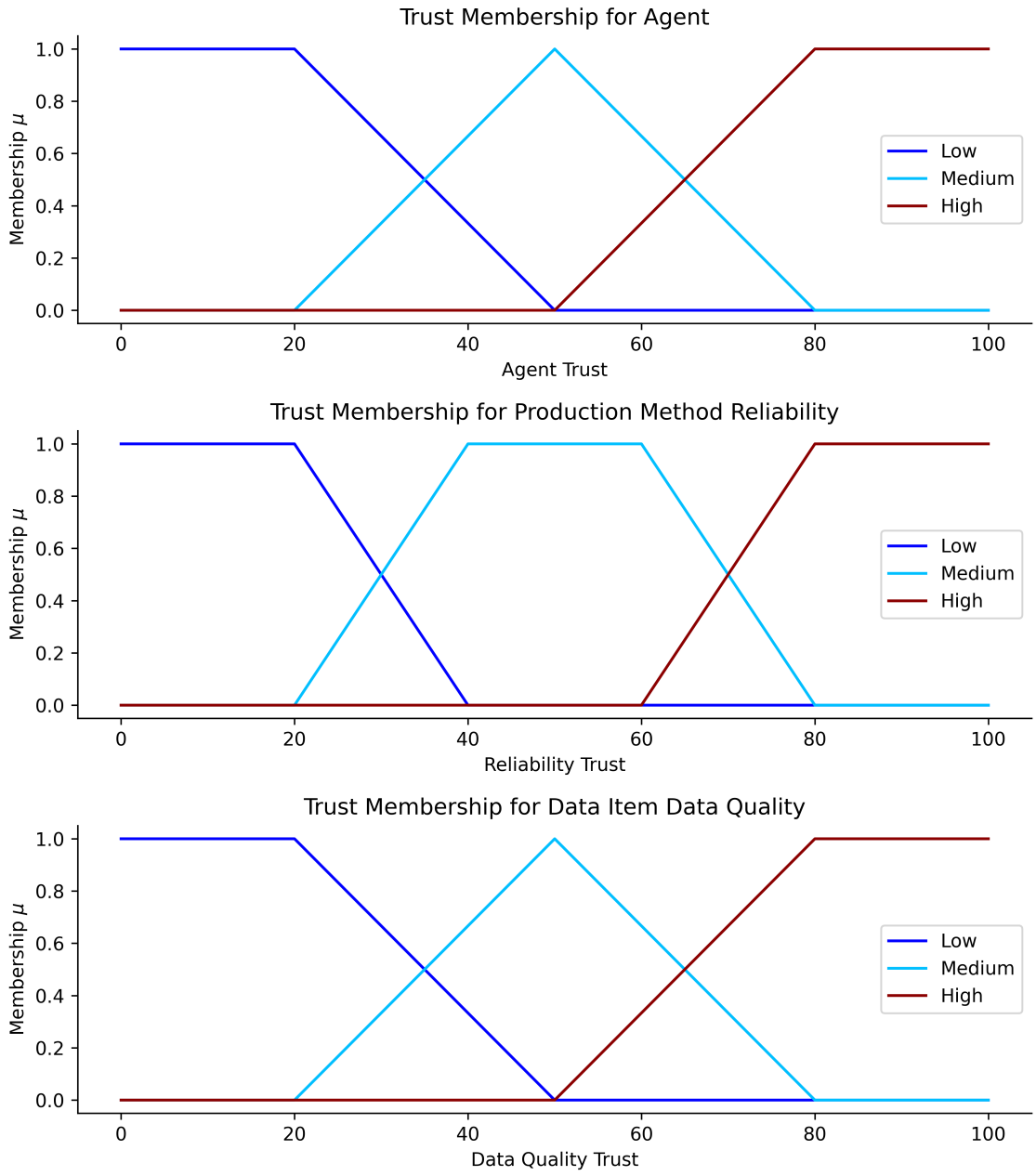


Figure 3.5: Fuzzy Membership Functions for Trust

The fuzzy association rules described in the previous sections form the basis of this inference, and they must be defined for all label combinations in R and A. The

method calculates the minimum degrees of the input fuzzy sets in $\mu_R(r)$ and $\mu_A(a)$ for each fuzzy association rule. It then combines the maxima of these membership functions to an output fuzzy set that forms the result of the fuzzy inference.

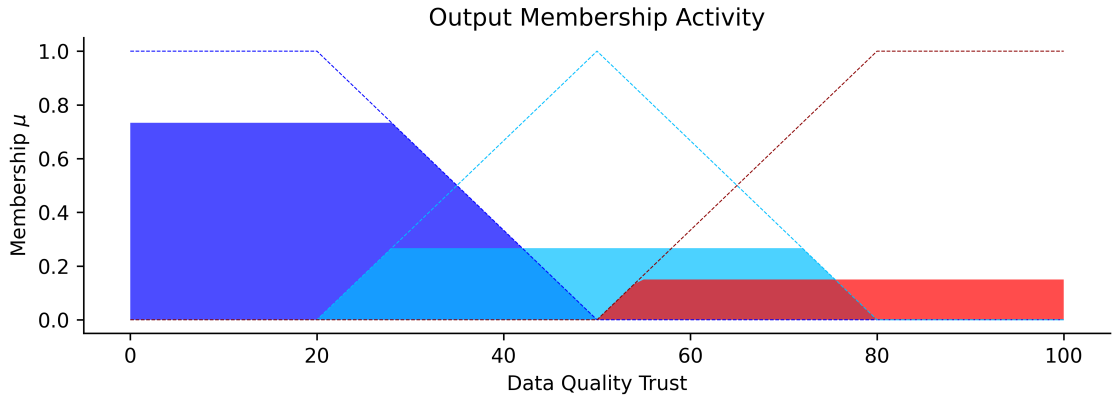


Figure 3.6: Fuzzy Output Membership Activity

The resulting shape can be difficult to process for further computation. To facilitate further processing, the shape is converted into a crisp value by way of defuzzification. Among other available methods, the centroid defuzzification method has been used successfully in other settings and will be applied in this model. The centroid d of D is a crisp value which represents the inferred data quality trust for the data entity. The corresponding label is the data quality trust label that maximizes $\mu_D(d)$. Figure 3.6 and Figure 3.7 show an example for a semi-automated method with $r = 63$ and $a = 28$ that has a trust inference result $d = 34$ which maximizes μ_D for the data quality trust label *low*. While this calculation has demonstrated how data quality trust can be computed from a mix of triangular and trapezoidal membership functions with our model, we will focus on only basic triangular membership functions (Figure 3.8) that limit complexity in describing this inference for the remaining chapters of this dissertation.

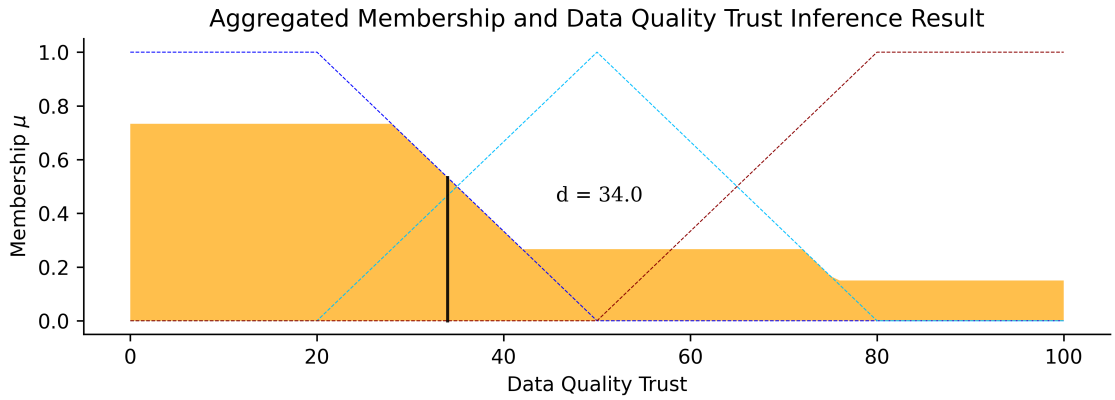


Figure 3.7: Fuzzy Aggregated Membership and Trust Inference Result

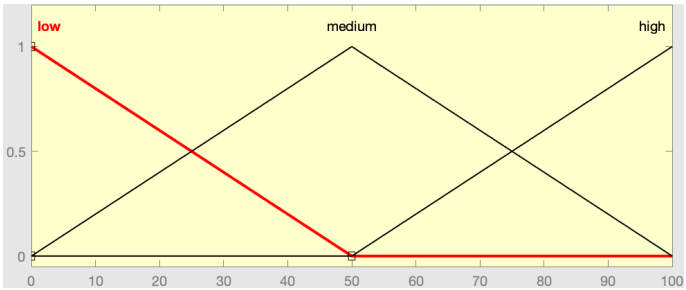


Figure 3.8: Basic Triangular Membership Functions in Matlab

3.2.3 Simulation of Fuzzy Inference

To validate that the Mamdani inference can compute data quality trust for our model with crisp input values without unexpected behavior, we devise a simulation experiment that provides visual confirmation that output from the fuzzy inference is monotonically increasing with increasing reliability trust and agent trust inputs. We also seek to experiment with different combinations of membership functions and trust labels.

Data Quality Trust Inference

Inferring a trust result through Mamdani inference is a method of approximate reasoning. Fuzzy sets used in the definition of trust require careful review to ensure that they meet the requirements for fuzzy inference [61]. Our proposed data quality trust model allows for the usage context-specific definition of trust labels with fuzzy membership functions. This requires configuration of trust membership for a particular implementation context before the model can be simulated and tested for efficacy in that context. We use three triangular membership functions with labels *low*, *medium*, *high* as the most basic representation for our model, as shown in Figure 3.8. Figure 3.12 shows our Matlab simulation setup with triangular membership functions. This model is applied to three different sets of nine association rules for manual, semi-automated, and fully-automated data production methods.

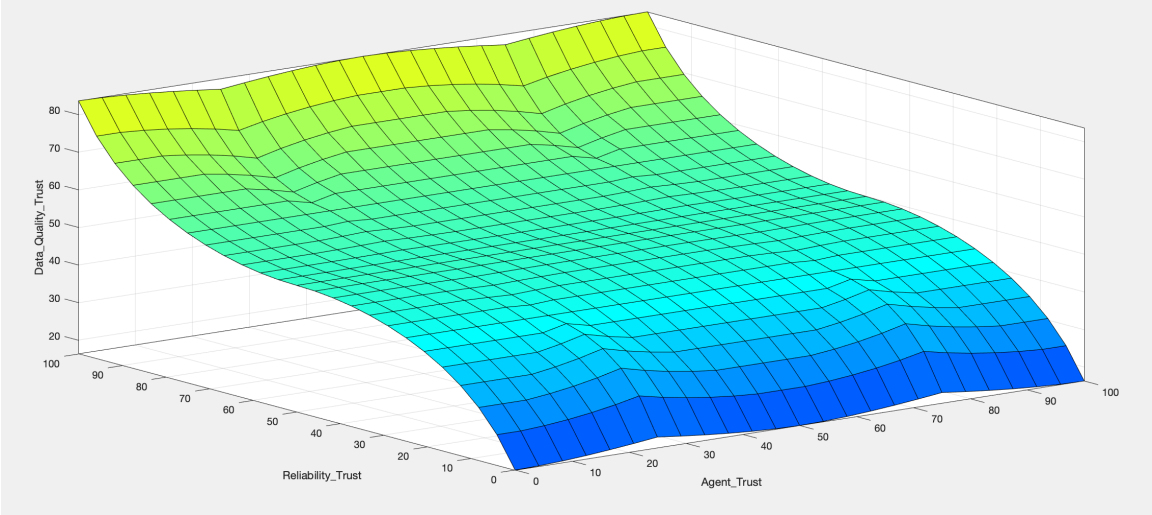


Figure 3.9: Matlab Simulation for Automated Method

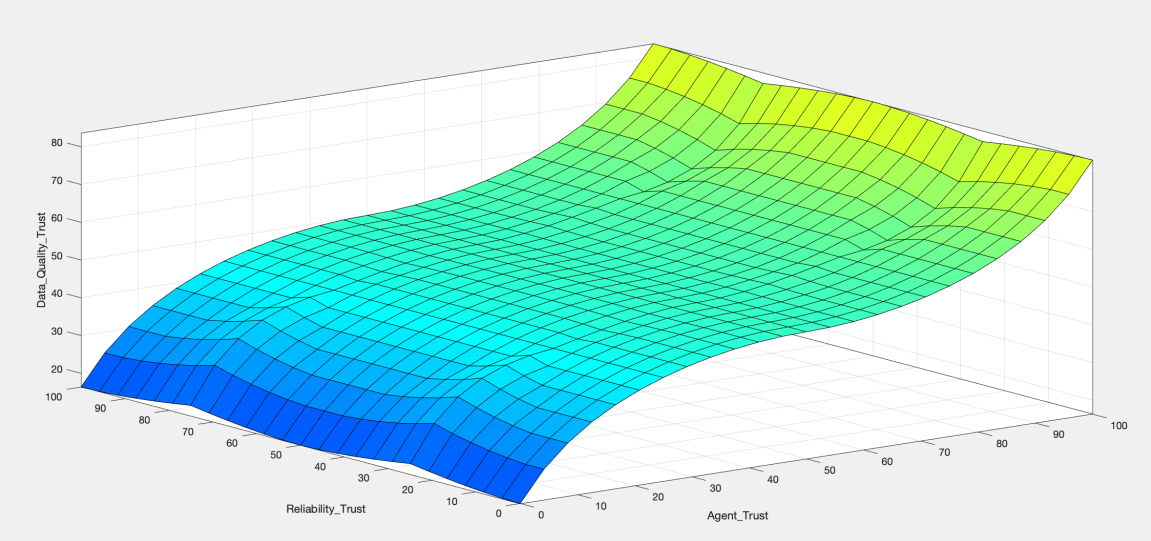


Figure 3.10: Matlab Simulation for Manual Method

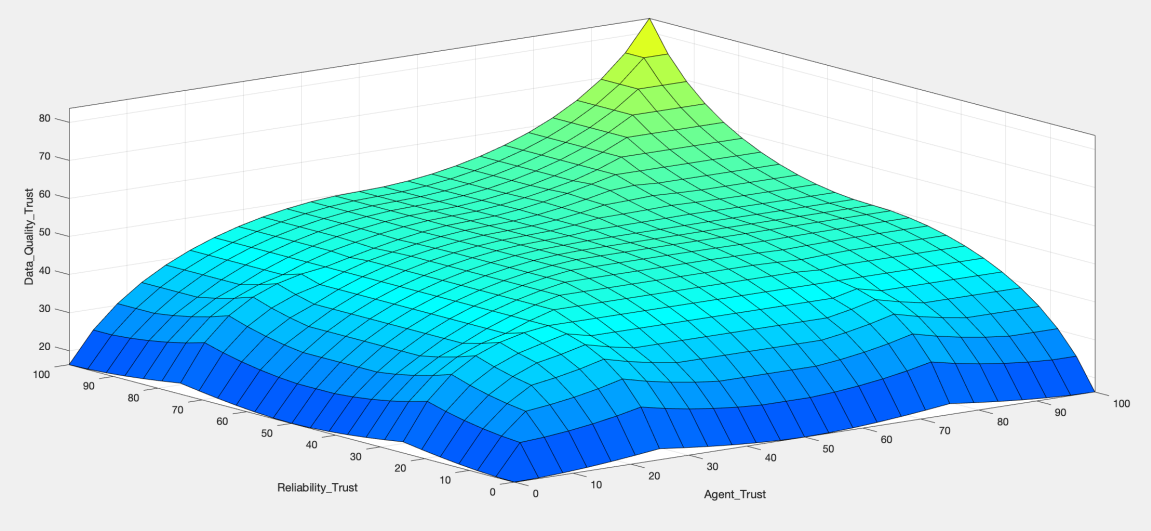


Figure 3.11: Matlab Simulation for Semi-Automated Method

Monotonic Inference Function

After configuring and running the simulation, we visually inspected the resulting graphical representations of crisp output values for crisp inputs [0,100]. Figure 3.9 shows the simulation of data quality trust for a data production method that is fully automated. Figure 3.10 shows the simulation of data quality trust for a data production method that is manual. Figure 3.11 shows the simulation of data quality trust for a semi-automated data production method. We also tested the Mamdani inference for other possible combinations of membership functions and could not find any combinations that caused unexpected behaviour. In particular, we wanted to visually confirm that the data quality trust output from the fuzzy inference was monotonically increasing with increasing reliability trust and agent trust inputs.

Minimum Number of Trust Labels

Our data quality assessment method is not limited to the three input and output trust labels of *low*, *medium*, *high* that were used for description of the method so far. Membership functions for a data quality trust result may define a smaller or larger number of labels that fit the requirements for a simplified assessment result in a particular use case. We have tested the method in Matlab for different label and membership function combinations. A basic implementation of the method uses only two input labels *low* and *high* with triangular fuzzy membership functions for agent trust and reliability trust. For these two input labels, Table 3.5 shows fuzzy association rules to reflect four possible associations for each type of data production method. The data quality trust result must be mapped to two or more output labels with different membership functions. Figure 3.13 shows an example of five output result labels.

```
[System]
   Type='mamdani'
   Version=2.0
   NumInputs=2
   NumOutputs=1
   NumRules=9
   AndMethod='min'
   OrMethod='max'
   ImpMethod='min'
   AggMethod='max'
   DefuzzMethod='centroid'

[Input1]
   Name='Agent_Trust'
   Range=[0 100]
   NumMFs=3
   MF1='low':'trimf',[0 0 50]
   MF2='medium':'trimf',[0 50 100]
   MF3='high':'trimf',[50 100 100]

[Input2]
   Name='Reliability_Trust'
   Range=[0 100]
   NumMFs=3
   MF1='low':'trimf',[0 0 50]
   MF2='medium':'trimf',[0 50 100]
   MF3='high':'trimf',[50 100 100]

[Output1]
   Name='Data_Quality_Trust'
   Range=[0 100]
   NumMFs=3
   MF1='low':'trimf',[0 0 50]
   MF2='medium':'trimf',[0 50 100]
   MF3='high':'trimf',[50 100 100]
```

Figure 3.12: Data Quality Trust Model Simulation Setup in Matlab

Table 3.5: Fuzzy Association Rules for Two-Label Input

Rule	**Reliability Trust**	**Agent Trust**	**DQ Trust**
Semi-Automated Production Method			
1	high	high	high
2	high	low	low
3	low	high	low
4	low	low	low
Fully-Automated Production Method			
1	high	high	high
2	high	low	high
3	low	high	low
4	low	low	low
Manual Production Method			
1	high	high	high
2	high	low	low
3	low	high	high
4	low	low	low

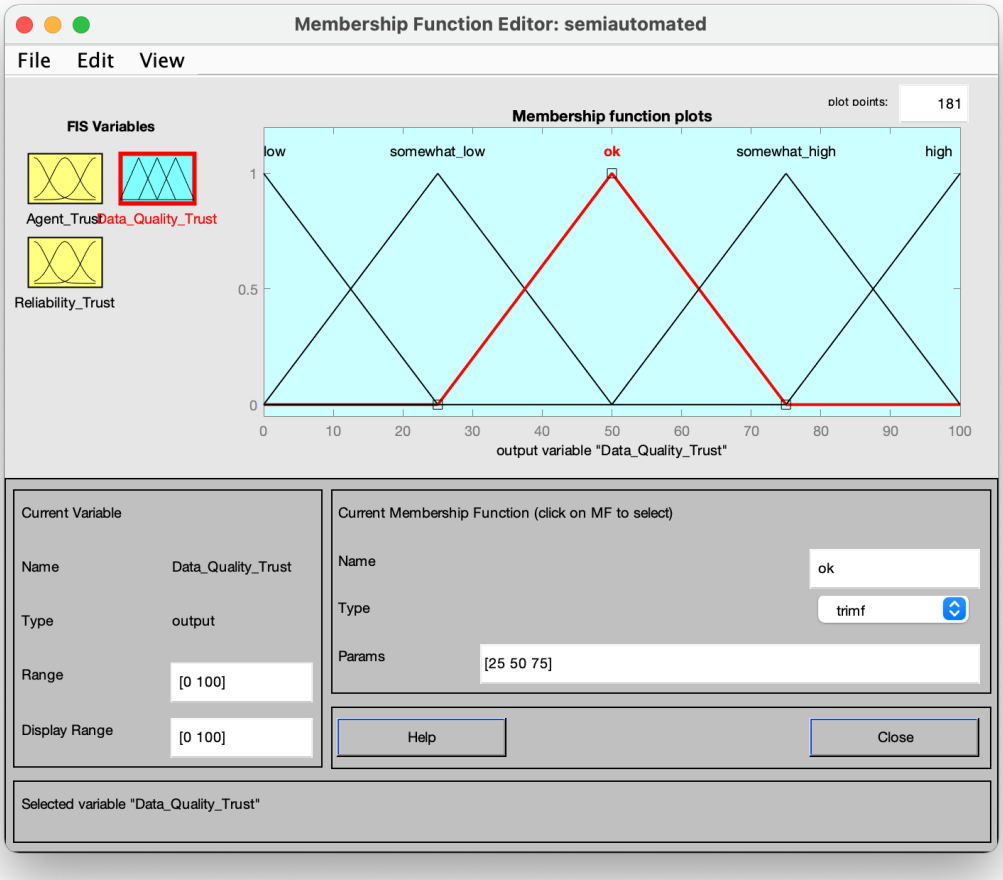


Figure 3.13: Trust Output Fuzzy Membership Function Definition in Matlab

3.2.4 Recursive Trust Inference for Provenance Trees

Trust inference described so far has shown how to infer data quality trust for a single data entity without antecedents. Our data quality trust model is not limited to single level data provenance and requires computation of trust from multi-level provenance trees, as shown in Figure 3.4. We introduce a recursive function dqt, and a fuzzy inference function fis for calculating data quality trust of a given data entity d from

its provenance tree. Data quality trust is inferred from agent trust a, and production method reliability r. It is constrained by the data quality trust of the lowest trusted antecedent data entity c_j used in the generation of d. If no antecedent data entity was used in the generation of d then the inference is based on a and r only. A logical conjunction between the inferred data quality trust of d and its antecedents c_j is represented as the minimum of $dqt(c_j)$ and $dqt(d)$. For $j \in \mathbb{N}$ let data entity c_j be an antecedent to data entity d and let k be the number of antecedents c_j to d then

$$dqt(d) = \begin{cases} fis(a,r) & \text{if } k = 0 \\ fis(a,r) \wedge \bigwedge_{j=1}^{k} dqt(c_j) & \text{if } k > 0 \end{cases} \tag{3.2}$$

Section 5.3 provides a detailed computation example of a recursive trust inference in a multi-level provenance tree.

3.2.5 Simulation of System Response

To validate that the assessment method meets the system response requirements of real-time use during clinical decision making, an evaluation of the code execution profile for recursive trust inference can highlight potential performance constraints of the assessment method. Nielsen's usability engineering guidance [62] provides that a system response time of less than one second allows for a user's flow of thought to remain uninterrupted, even though the user will notice a delay. A user may perceive a system as reacting instantaneously if the response time is less than 0.1 seconds [62]. Since computational complexity of the Mamdani inference algorithm is linear, run-time execution of the inference should not result in exponentially increasing processor run times and system response times.

We simulate the data quality assessment method in a simple experiment using

Scientific Python scripts and scikit-fuzzy function libraries [63] to evaluate execution performance of the fuzzy inference. The built-in Python profiler tool provides deterministic profiling of Python programs. A profile is a set of statistics that describes how often and for how long various parts of the program executed. These statistics can be formatted into reports. Figure 3.14 shows the execution profile of a script that infers data quality trust for 10,000 combinations of agent trust and reliability trust values on the intervals [0,100]. We executed this script on a current generation desktop computer (MAC MINI 2018 with 3.2 GHz 6-Core Intel Core i7 and 64 GB RAM) resulting in a code execution time of 4.3 seconds. Most applications of the trust model and method require computations of a much smaller number of data items with execution times less than one second. For example, computation of data quality trust of 22 data items in the BP Centiles Trust Provenance Tree in Figure 5.7 completed in 0.008 seconds.

```
         940001 function calls (880001 primitive calls) in 4.311 seconds

   Ordered by: cumulative time

   ncalls  tottime  percall  cumtime  percall filename:lineno(function)
    10000    0.317    0.000    4.311    0.000 /Users/jpstoldt/Documents/PhD Dissertation/DQTPerfEval.py:9(dqt)
    10000    0.027    0.000    3.654    0.000 /opt/anaconda3/lib/python3.8/site-packages/skfuzzy/defuzzify/defuzz.py:211(defuzz)
    10000    3.555    0.000    3.578    0.000 /opt/anaconda3/lib/python3.8/site-packages/skfuzzy/defuzzify/defuzz.py:38(centroid)
    60000    0.031    0.000    0.341    0.000 /opt/anaconda3/lib/python3.8/site-packages/skfuzzy/fuzzymath/fuzzy_ops.py:531(interp_membership)
    60000    0.029    0.000    0.310    0.000 <__array_function__ internals>:2(interp)
120000/60000    0.040    0.000    0.274    0.000 {built-in method numpy.core._multiarray_umath.implement_array_function}
    60000    0.049    0.000    0.250    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/lib/function_base.py:1278(interp)
    60000    0.105    0.000    0.105    0.000 {built-in method numpy.core._multiarray_umath.interp}
    60000    0.023    0.000    0.073    0.000 <__array_function__ internals>:2(iscomplexobj)
    10000    0.004    0.000    0.040    0.000 {method 'sum' of 'numpy.ndarray' objects}
    10000    0.003    0.000    0.036    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/core/_methods.py:36(_sum)
    10000    0.033    0.000    0.033    0.000 {method 'reduce' of 'numpy.ufunc' objects}
    60000    0.022    0.000    0.028    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/lib/type_check.py:280(iscomplexobj)
    60000    0.012    0.000    0.023    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/core/_asarray.py:16(asarray)
    60000    0.011    0.000    0.011    0.000 {built-in method numpy.array}
    10000    0.011    0.000    0.011    0.000 {method 'astype' of 'numpy.generic' objects}
    10000    0.009    0.000    0.011    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/core/getlimits.py:365(__new__)
    60000    0.006    0.000    0.006    0.000 {built-in method builtins.issubclass}
    60000    0.006    0.000    0.006    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/lib/function_base.py:1274(_interp_dispatcher)
    60000    0.006    0.000    0.006    0.000 /opt/anaconda3/lib/python3.8/site-packages/numpy/lib/type_check.py:207(_is_type_dispatcher)
    20000    0.006    0.000    0.006    0.000 {method 'ravel' of 'numpy.ndarray' objects}
    40000    0.004    0.000    0.004    0.000 {built-in method builtins.len}
    10000    0.002    0.000    0.002    0.000 {method 'get' of 'dict' objects}
    10000    0.001    0.000    0.001    0.000 {method 'lower' of 'str' objects}
        1    0.000    0.000    0.000    0.000 {method 'disable' of '_lsprof.Profiler' objects}
```

Figure 3.14: Execution Profile for Fuzzy Trust Inference in Python

It should be noted that for a large number of antecedent nodes in multiple levels of a provenance tree, the recursive function properties of our inference may cause

system memory issues. This potential algorithmic performance constraint should be addressed during system implementation and system configuration for a particular usage context.

3.3 Model and Method Extension for Verification

The core data quality trust model is limited in that trust in a multi-level provenance tree cannot be improved from a lower trust antecedent data entity since data quality trust is constrained by the lowest trust antecedent data entity. By extending the core model with a new type of activity for verification, a data entity can be highly trusted despite lower trusted antecedent data entities in its provenance tree. Verification can also lower data quality trust by probing for data quality issues with a data item. A verification activity verifies data quality trust of a given data entity, and produces a new data entity with inferred data quality trust properties. Figure 3.15 shows the possible sequencing and data flow between data entities, production activities, and verification activities.

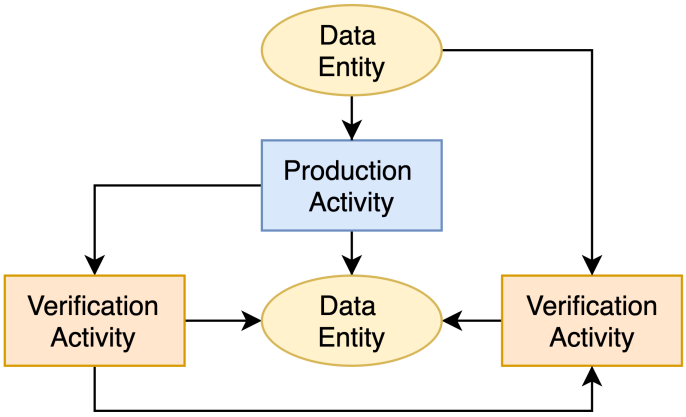


Figure 3.15: Data flow between Entities and Activities

Trust inference in a verification activity is based on a data user's agent trust in

verification agents, and reliability trust in data verification methods. This inference result can be combined with additional inputs such as other data entities, production activities, or verification activities that may influence data quality trust of a given data entity. Verification activities may result in higher or lower data quality trust than the trust in the antecedent data entity. Figure 3.16 illustrates the elements of a data verification activity.

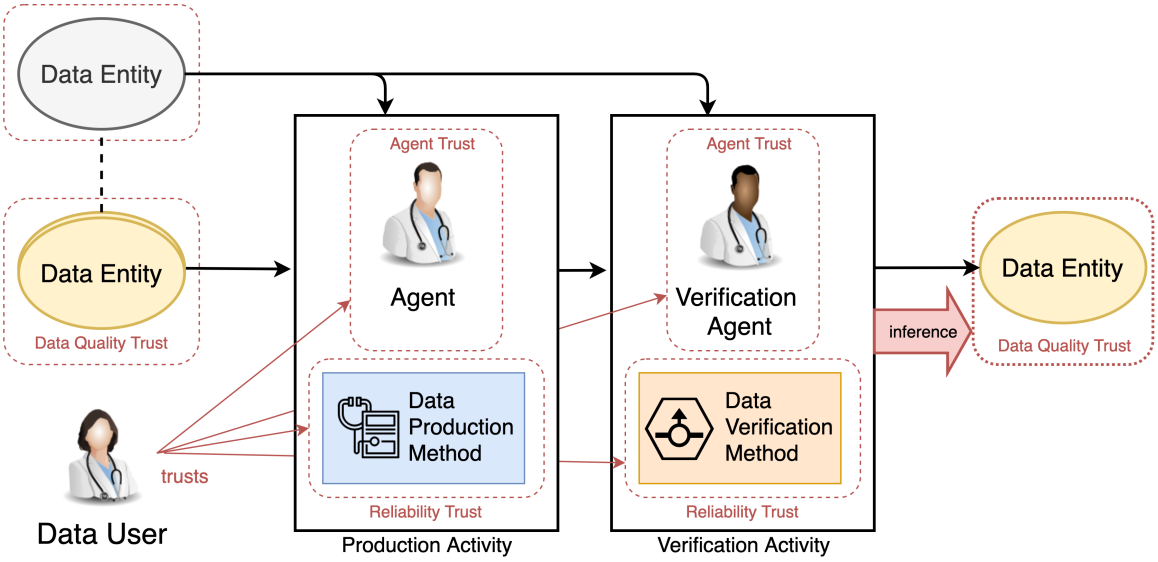


Figure 3.16: Illustration of Extended Trust Model with Verification

3.3.1 Trust in the Capability of Verification Agents

Verification agents vary in their capabilities. We define capability of a verification agent as the degree to which a verification agent is able to perform a data verification method correctly to verify data quality properties of an antecedent data entity. Agent trust in verification agents uses the base variable a and fuzzy set agent trust A from our core model.

3.3.2 Trust in the Reliability of Data Verification Methods

Data verification methods vary in their reliability. We define reliability of a data verification method as the degree to which a verification method can be depended on for precisely, and correctly or accurately, verifying the data quality of a given data entity. Reliability trust in data verification methods uses the base variable r and fuzzy set reliability trust R from our core model.

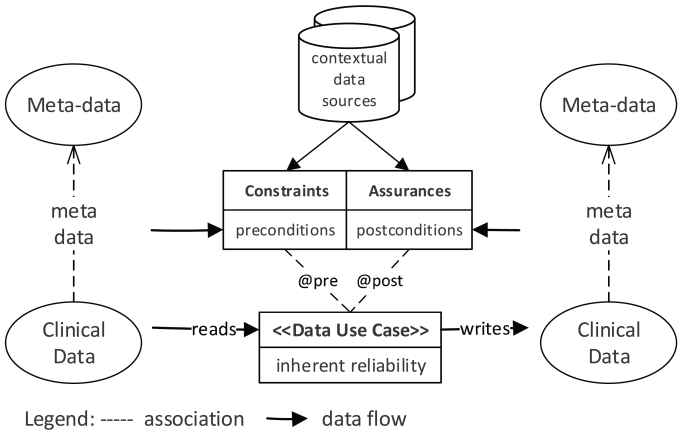


Figure 3.17: Functional Model for Data Quality Reasoning (Weber, 2015)

3.3.3 Rules for Verified Data Quality Trust

Fuzzy trust properties of verification agents and data verification methods can be combined to a fuzzy output set that characterizes the degree by which a verification activity should increase or decrease the trust level of an antecedent data entity. Depending on the level of automation in the data verification method, data verification relies on the verification method based on the association rules defined in Section 3.2.1.

3.3.4 Implementation of Data Verification Methods

Data verification can be implemented with methods that use intrinsic or extrinsic data to verify the data quality of a given data entity. This may be through probing for data quality issues, or corroboration with additional supporting data entities that confirm a data quality trust level. In particular, the concept of data quality probes and prior research about their application in monitoring and assuring data quality "by-design" [64] can help with the design of data verification methods. Data Quality by Contract (DQbC) is based on "Design by Contract" [65], a component-based software engineering paradigm, which allows formalizing and monitoring of expectations and guarantees of the interfaces among software components. DQbC models data use cases as functions that map a domain of input data to a range of output data. Figure 3.17 shows the DQbC functional model for reasoning about data quality with pre-conditions (constraints on input data) and post-conditions (assurances on output data) that must be met so data quality "contracts" can be fulfilled [64]. DQbC envisions data quality probes to be deployed for monitoring and assuring data quality during system operation. Probes can be classified based on the scope of data considered (Table 3.6).

Table 3.6: Classification of Data Quality Probes (adapted from Weber, 2015)

Type	**DQ Probe**	**Data Considered**	**Practical Example**
0	Intrinsic (conformance)	Input data (use case)	Is the medical problem encoded?
1	Intrinsic Meta (provenance)	Input data and meta data	Is the lab test recent?
2	Extrinsic Internal (internal concordance)	Input data, meta data, other data in same HIS	Does the prescription agree with the problem?
3	Extrinsic External (external concordance, currency)	Input data, meta data and other data in other HIS	Do active medications listed in the EMR agree with those listed in the pharmacy system?
4	Statistical (plausibility)	Input data, meta data and statistical data	Does the prevalence of coded Diabetes agree with national statistics?

3.3.5 Simplification with Binary Trust Assignments

In some usage contexts, it may not be practical or desired to use fuzzy sets for the assignment of trust to data verification methods. Instead, a data user may choose to trust data verification methods via binary trust assignments. A data quality probe used for verification may be activated or deactivated by a data user. If a data verification method received a binary trust assignment, then the verified data quality trust result will be crisp since fuzzy inference is not possible. For example, if a trusted

data quality probe identifies a data quality issue as part of a verification activity, the resulting value of the trust variable t may be 0 for $\mu_T(t)$ mapped to a *low* data quality trust label.

3.3.6 Multi-level Data Provenance with Trust Verification

Verification activities can be added to the trust provenance tree for a data entity, and combined with other data entities, production activities, and verification activities shown in 3.15. Figure 3.18 provides an example of a trust provenance tree of a diabetes diagnosis data entity with its antecedents and several transformations for data entity trust verification. A network graph of transformations t_n and v_n represents how data production and verification affect trust of node and leaf data entities as antecedents for the root node. Transformation activities and verification activities can be represented as edges in this graph. A transformation activity t_n can generate root, node and leaf data entities. A verification activity v_n can generate root and node data entity, but not a leaf data entity. t_{1-3} generate leaf data entities that require verification. Verification activities v_{1-4} each take one antecedent data entity as an input, and generate one output data entity.

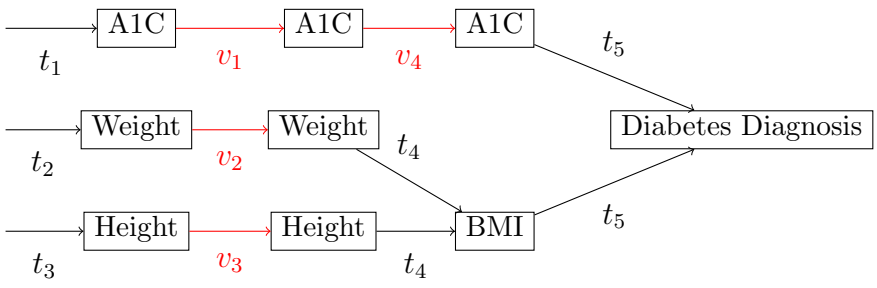


Figure 3.18: Data provenance tree with verification

3.4 Model and Method Extension for Certification

A second extension to the core model allows for the certification of agents or data production methods that participate in transformation activities. In Figure 3.19, a trusted certification authority certifies an agent or a data production method. The data user has the ability to assign a trust level to any number of certification authorities. If a certification authority has received a trust level by the data user, then a certified agent or a certified data production method receive this trust level for the data quality trust inference.

Certification Authorities vary in their capabilities. We define certifier trust in a certification authority as the degree to which a data user trusts the capability of a certification authority to certify data production methods and agents correctly. We express certifier trust as a fuzzy set C with the labels *low*, *medium*, *high*. Certifier trust in our model is a relative trust level assigned by a data user. A membership function μ_C maps the base certifier trust variable c to overlapping trust labels in C.

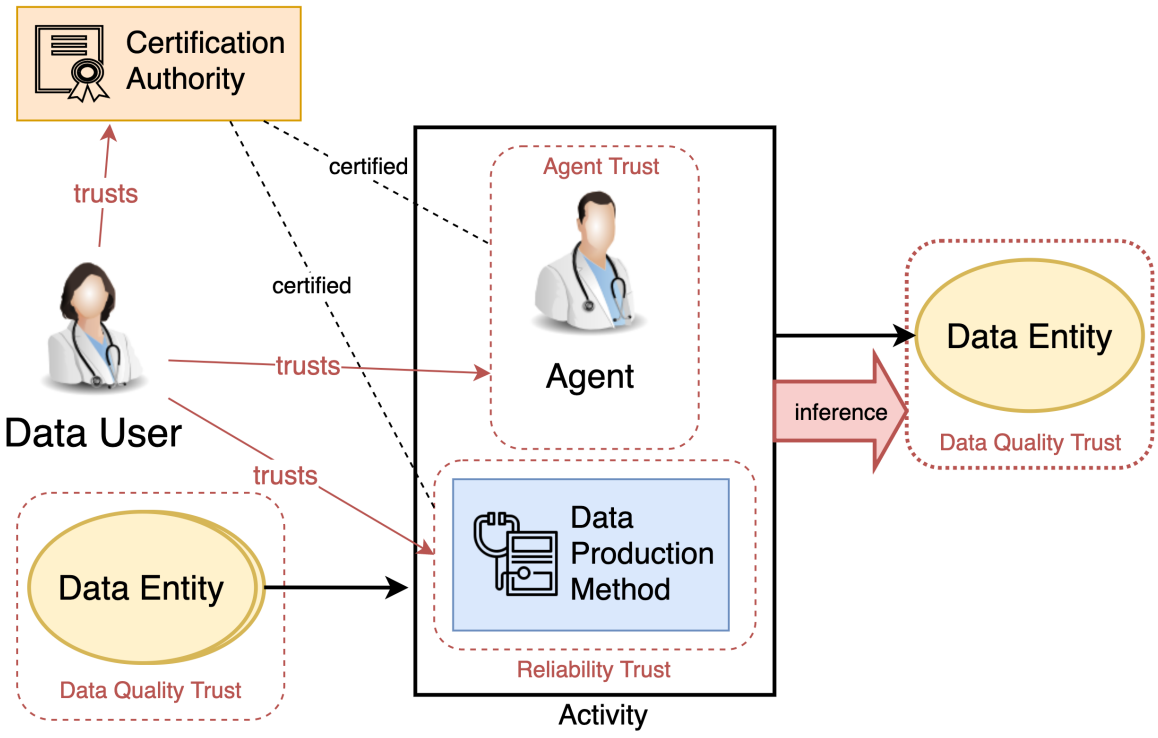


Figure 3.19: Illustration of Extended Trust Model with Certification

3.5 Summary

This chapter described a novel approach to provenance-based data quality assessment inferred from a data user's trust in individuals and methods that generated the data. A core model representation of data quality trust, an automated method to assess trust, and two model and method extensions for verification and certification features were presented to address the research objectives for a contextual data quality assessment method. The new method provides a contextual data quality assessment result that represents the relative degree of data quality trustworthiness for individual data items. Assessment results are provided in near real-time so they are available for reasoning about data quality during clinical decision making at the point of care.

Assessment results are presented with a simplified contextual result using labels of *high*, *medium*, *low*.

The next chapter will transfer model and method to the EMR system domain with a design pattern. It will show data integration with both interoperable and several legacy EMR systems are possible, and it will develop visualization guidance for extending EMR system user interfaces with data quality trust.

Chapter 4

Integration with Electronic Medical Record Systems

This chapter describes a design pattern for the integration of data quality trust with both interoperable and legacy EMR systems. We introduce the FHIR and Substitutable Medical Applications and Reusable Technologies (SMART) standards, and we develop FHIR interface extensions that allow for the transmission of data quality trust information to achieve platform interoperability research objectives. We extend the data models of several legacy EMR applications to integrate data quality trust. Next, we review current user interface design principles from the literature, and we discuss how data quality trust can be integrated with EMR user interfaces while supporting a user in two cognitive processing modes. We propose several screens for visualization of data quality trust with EMR applications. As part of the validation research process (Figure 1.2), the chapter draws applicable knowledge about EMR Systems from the knowledge base to develop/build and justify/evaluate a design pattern that is iteratively assessed and refined. Main research methods embedded in this chapter are conceptual modeling, prototyping, and lab testing.

4.1 Integration with Interoperable EMR Systems

There are many advantages to building interoperability features into modern healthcare applications. Earlier work has explored how the safety of SMART apps can be improved by applying a "Data Quality by Contract" approach to pre- and post-conditions of data use cases [49]. The SMART on FHIR architecture can support platform interoperability objectives for the integration of a novel data quality assessment method with FHIR-enabled EMR systems, as described in the following sections.

4.1.1 Fast Healthcare Interoperability Resources

Health Level Seven (HL7) established the FHIR standard to provide a Application Programming Interface (API) for healthcare using a simple Hyper Text Transfer Protocol (HTTP) interface for clinical, administrative, and infrastructure data [5]. FHIR combines lessons learned from several healthcare data exchange and information modeling standards (HL7 v2, HL7 v3, Reference Information Model (RIM), Clinical Document Architecture (CDA)) with emerging industry approaches [66]. FHIR data is organized using "resources" as building blocks with a Representational State Transfer (REST) API to access those information resources. FHIR profiles define higher level conformance packages. Several initiatives are underway to facilitate the adoption of FHIR, including the HL7 Argonaut project [67].

4.1.2 SMART on FHIR Client Apps

SMART [5, 68] originated from an interoperability project at Harvard Medical School and Boston Children's Hospital with the goal of creating medical applications that could be run unmodified across different HIS platforms. SMART apps are EMR-

platform agnostic and feature modular reusable components built around a set of core services. Combining SMART app architecture principles with FHIR APIs to a "SMART on FHIR" architecture standard provides the previously missing standards-based, interoperable apps platform for EMR systems.

A SMART on FHIR system implements the SMART on FHIR specification. The SMART App Launch Framework supports use cases for patient and provider apps that launch either from a portal or standalone [69]. The current generation of SMART apps uses HTML5 with JavaScript technologies for their API [68]. SMART on FHIR uses FHIR profiles to define resources, Open Authorization 2.0 (OAuth2) [70] for authorizing resource access, and OpenID Connect (OIDC) [71] for user authentication. Defined FHIR Resources form the building blocks of datasets that are transmitted over REST APIs. OAuth2-based access authorization enables patients or clinicians to approve third party SMART apps to access specific EMR datasets. OIDC is an authentication protocol that uses simple JSON Web Tokens that can be obtained based on OAuth2 authorization flows. It allows end users to sign into SMART apps through external identity providers.

SMART on FHIR has seen increasing adoption from EHR vendors in recent years as the potential of third-party apps has been recognized as an important technology trend.

4.1.3 FHIR Interface Extensions

Interface extensions to existing FHIR resources allow for the transmission of data quality trust information to achieve platform interoperability. The following sections show how FHIR Practitioner and FHIR Observation resources can be extended to encode data quality trust in patient observations, such as blood pressure measurements. We also consider interface integration and low network bandwidth environments.

User Trust Settings with FHIR Practitioner

The core data quality trust model leverages the FHIR Practitioner resource to encode data user-specific trust settings. Three new sub-entities *Agent Trust*, *ReliabilityTrust*, and *TrustLabelAssociations* are added as extensions to the FHIR Practitioner resource. Reliability trust is assigned to a production method defined with an identifier code. A more granular reliability trust assignment is possible by targeting an optional reference to a FHIR Device. A new valueset *fuzzy-trust-label* contains the data structure that stores characteristics of the trapezoidal membership functions for each fuzzy trust label.

Extension for Verification

FHIR-enabled transfer of data for verification in our extended model requires an additional sub-entity *VerificationTrust* to be added to the FHIR Practitioner resource. Verification status of a patient observation can be stored in existing FHIR Provenance data structures. For verification, a practitioner (*Provenance.agent.who*) is coded as "verifier" (*Provenance.agent.type*) for an observation (*Provenance.target*). *Provenance.activityType* specifies the verification method used.

Extension for Certification

A trusted certification authority can be added to the FHIR Practitioner through an additional sub-entity *CertifierTrust*. Transmission of certification status for practitioner and device, as part of a data production method, uses existing FHIR resources. FHIR Device contains properties for the regulatory certification status of medical devices.

Certification of measurement devices is represented with a unique device identifier. The International Medical Device Regulators Forum (IMDRF) UDI Working

Group published a UDI System for Medical Devices which is the base specification for Unique Device Identifiers. In the United States, the Food and Drug Administration has produced an implementation guide for Unique Device Identifiers which implements the IMDRF specification and other jurisdictions may produce similar IMDRF implementation guides as well. Assigned UDIs are encoded in the FHIR specification ([72] as part of the FHIR Device resource. The FHIR standard encodes the qualification of practitioners as part of the FHIR Practitioner resource. It stores certification, licenses, or training pertaining to the provision of care encoded along with the certifying organisation (*Practitioner.qualification.issuer*). Table 4.1 shows an integrated view of FHIR extensions required for transmitting trust settings of core model, and extensions with verification and certification.

Table 4.1: FHIR Extension: FHIR Practitioner

Identifier [Cardinality]	**FHIR Extension Type**
Practitioner.AgentTrust [0..*]	BackboneElement
.agent [1..1]	Reference (Practitioner)
.agent-trust [1..1]	Decimal
Practitioner.ReliabilityTrust [0..*]	BackboneElement
.methodcode [1..1]	CodeableConcept
.device [0..1]	Reference (Device)
.reliability-trust [1..1]	Decimal
Practitioner.CertifierTrust [0..*]	BackboneElement
.certification-authority [1..1]	Reference (Organisation)
.certifier-trust [1..1]	Decimal
Practitioner.VerificationTrust [0..*]	BackboneElement
.verificationmethod [1..1]	CodeableConcept (Provenance.activityType)
.verification-trust [1..1]	Decimal
fuzzy-trust-label	ValueSet

Trust in Observations with FHIR Observation

Data quality trust results for a patient observation can be encoded through extension of the FHIR Observation resource. A new entity *DQTrust* specifies the data quality trust for a practitioner data user. Figure 4.2 shows the additional fields required for the extension.

Table 4.2: FHIR Extension: FHIR Observation

Identifier [Cardinality]	**FHIR Extension Type**
Observation.DQTrust [0..*]	BackboneElement
.practitioner [1..1]	Reference (Practitioner)
.dq-trust [1..1]	Decimal

Data Provenance with FHIR Provenance

In addition to using the FHIR Provenance resource for encoding verification of observations, the data structure of this resource is flexible to encode provenance trees and multi-level data provenance for any data use case represented with FHIR resources.

Interface Integration with HAPI FHIR

Data quality trust information can be stored within the EMR or separately from other EMR data. However, due to complexities of modifying EMR data models and existing interfaces, an implementation with a separate Trust Preferences Server may be advantageous. Open source EMR implementations may be able to modify the EMR data model directly to incorporate database schema extensions.

Extensions to the FHIR standard for data quality trust create the necessary data structure and file format for exchanging data quality trust through standardized health information exchange interfaces. A HAPI FHIR reference implementation [73] includes the HAPI FHIR library and an open-source implementation of the FHIR specification in Java. Several features of HAPI FHIR can be utilized to intercept and customize FHIR REST API service calls for data quality trust FHIR interface extensions. Rivera Sanchez et al. [74] and Hussain et al. [75] offer practical guidance on how to implement request interceptors for REST API calls with FHIR HAPI.

Low Network Bandwidth Environments

Several implementation options for data exchange and storage of data item trust properties when exchanging trust properties over FHIR interfaces can alleviate low bandwidth concerns of implementations. Trust properties for a data item can be computed at run-time within the usage context, or a pre-computed copy of the trust properties can be retrieved from persistent database storage. Since data quality trust is contextual and user-specific, persistent storage of trust requires exchange and storage of trust properties for every combination of user and data item in a data item collection. A large collection of data items with many users may increase the data base memory requirements of the original table structure significantly. FHIR interface limitations and network latency may require caching of data if pre-computed trust properties need to be transmitted between database server and SMART app client. An alternative is to exchange user trust settings over interfaces and to only store user trust settings persistently, requiring the computation of data quality trust to occur at run-time in the SMART app. Since proposed model and assessment method focus on assessing the data quality of individual data items in a particular usage context, trust computation at run-time will be more beneficial in most usage scenarios.

4.2 Integration with Legacy EMR Systems

Interoperability characteristics form an integral part of the SMART on FHIR architecture standard and were driving factors for its inception [76]. Data quality trust implemented within a SMART on FHIR app meets our research objectives for platform interoperability. However, there are still many legacy EMR applications in use that have not been upgraded with FHIR interfaces yet. Healthcare systems are facing challenges in upgrading their old applications, in part due to difficulties in migrating

legacy data. We have evaluated feasibility of integrating data quality trust with some of those non-FHIR EMR applications.

4.2.1 Legacy EMR Systems without FHIR Interfaces

As a first step, we sought to identify suitable legacy systems that would allow us to analyze the data models and develop data model extensions with data quality trust. A recent survey of open-source projects [77] provided a listing of popular systems and we selected three open-source EMR applications (OSCAR McMaster, OpenMRS, OpenEMR) and the OpenEHR implementation standard to evaluate the technical feasibility of extending those systems with data quality trust functionality.

We downloaded the source code for the three projects and installed the applications in a development environment using the provided documentation. To understand the data models and the semantic meaning of tables and relations, we searched for data dictionaries and analyzed the table structures. In particular, we wanted to understand potential integration points for data quality trust as it related to patient observations. A review of the project documentations provided that with the notable exception of OpenEHR, documentation of the existing database schemata appeared incomplete. The projects required re-engineering of the code base to establish the semantic meaning of database tables and fields before data quality trust integration could be evaluated.

4.2.2 Generic Database Schema Extension

We developed a generic database schema extension for data quality trust of patient observations. The following sections show how this schema can be integrated with legacy applications. Figure 4.1 provides the schema with nine tables and foreign key relationships that define the data quality trust relation between an existing *provider*

(data user) and *observation* (data item).

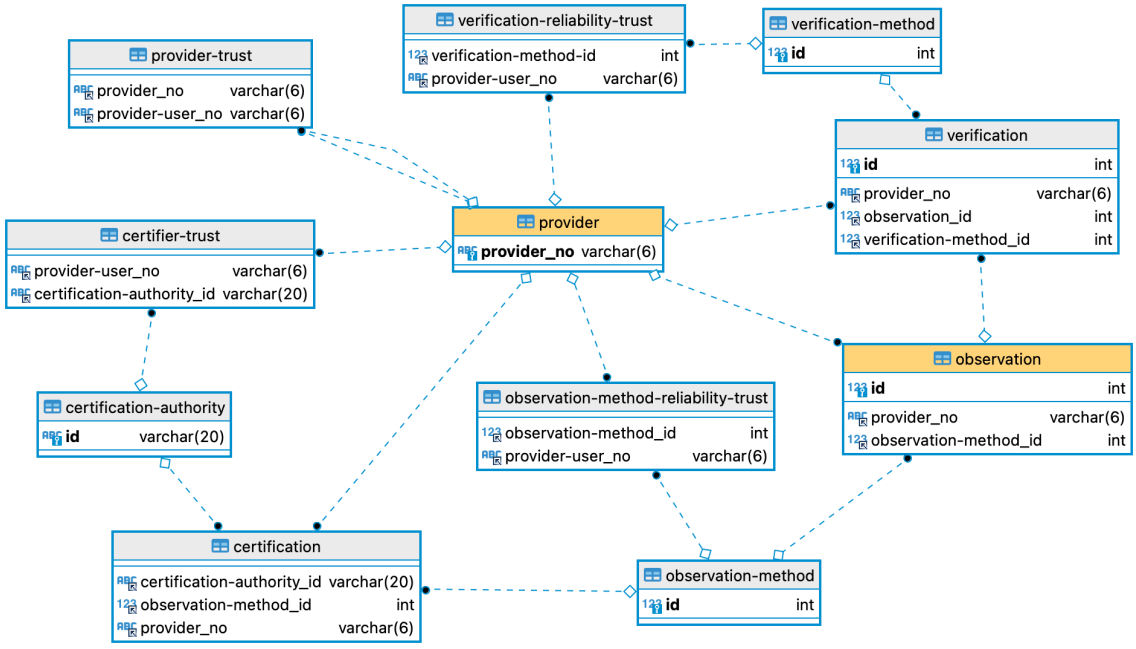


Figure 4.1: Generic Database Schema Extension for Data Quality Trust

The trust relation between a data user and a healthcare provider is represented in a table *provider-trust*. The trust relation between a data user and a observation data production method is represented in a table *observation-method-trust*. The trust relation between a data user and a healthcare provider is represented in a table *certifier-trust*. The trust relation between a data user and a observation verification method is represented in a table *verification-reliability-trust*. The schema extension does not include a new table for data item trust. A data user's trust in a observation data item is computed at run-time from trust assignments to agent, production method reliability, certifier, and verification method reliability.

4.2.3 OSCAR McMaster

OSCAR McMaster (OSCAR) was the first system analyzed. OSCAR is an open-source EMR system designed by doctors for doctors, mainly used in primary care settings in Canada and several other countries. OSCAR is built using the Java programming language, a MySQL or MariaDB database, and it runs in an Apache Tomcat web container. The OSCAR architecture leverages Struts, Spring, JPA, Angular and jQuery application frameworks. Basis for our research is OSCAR EMR version 15.

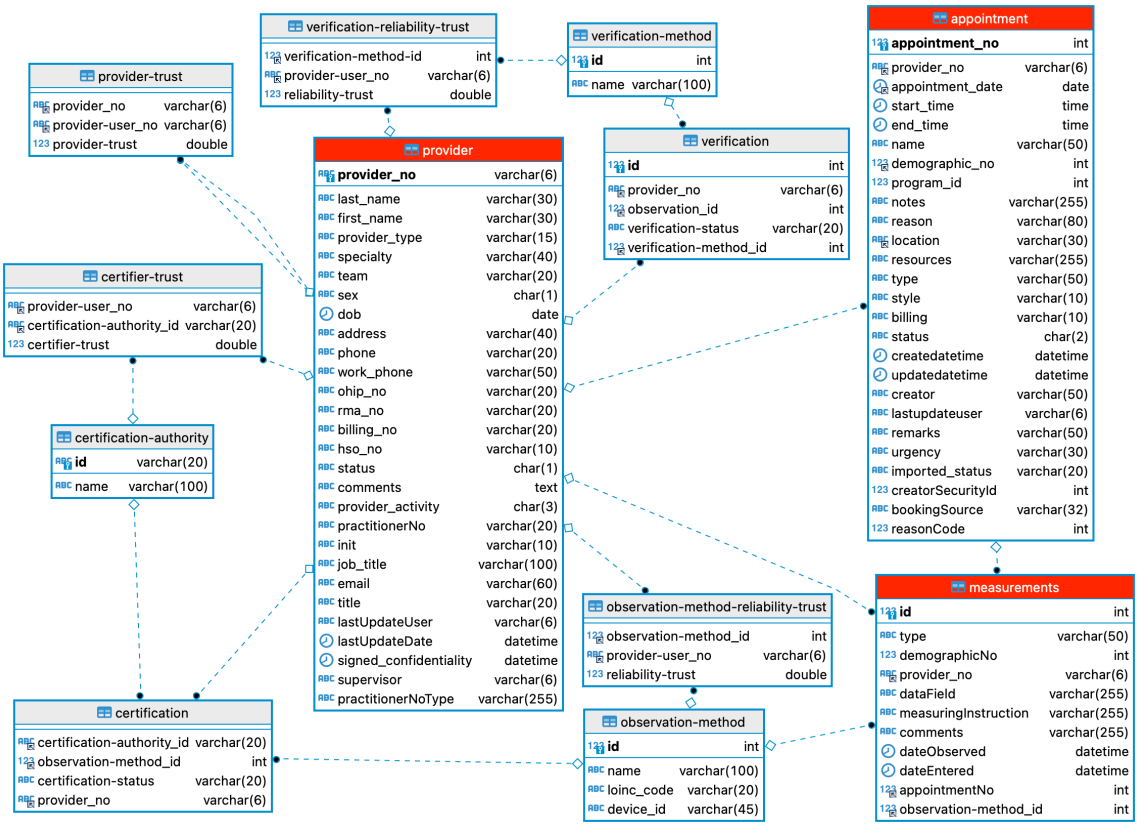


Figure 4.2: Oscar Database Schema Extension

Existing Data Model

OSCAR required re-engineering of the code base to establish the semantic meaning of database tables and fields before data quality trust integration could be evaluated. To find the tables that stored patient observations, we executed the application in our development environment in debug mode and manually traced Java source code segments that generated the HTML code to display the application user interface in the web browser client. The functional application area allowing a user to enter and organise patient measurements led us to several code segments that queried database tables. Further inspection and testing confirmed that patient observations are stored in a *measurements* table that organises measurements made in the context of a patient encounter with a healthcare provider. Observations can be grouped by types and meta information can be attached to an observation through unstructured text fields *measuringInstruction* and *comment*. Patient encounters are stored in a table *appointment* which may be linked to multiple patient observations. The healthcare provider is stored in a table *provider* which contains demographic information. In Figure 4.2, we have color coded existing tables *provider*, *appointment* and *measurements* in red. These were identified as suitable integration points for data quality trust.

Data Model Extension

Figure 4.2 shows table extensions to the data entity relationship model colored in grey. The existing table *provider* forms the basis of several new data entity relationships that store a healthcare practitioner's data quality trust preferences. The existing table *measurements* was extended with an additional field *observation-method-id* that links a patient observation with the observation method used to generate the observation. This extended schema can support implementation of data quality trust functionality

by augmenting the existing OSCAR user interface, or it could serve as EMR data store for trust preferences with a future FHIR-enabled implementation of OSCAR with SMART apps.

4.2.4 OpenEMR

OpenEMR is a free and open-source EMR and medical practice management application that features fully integrated EMRs based on a MySQL database as data store. After installing the application in our development environment, we executed and debugged to manually trace Java source code segments that generated the HTML code to display the application user interface in the web browser. Review of the MySQL database structure definition files provided that foreign key constraints between database schema tables were not defined. Such definition would have allowed the creation of an entity relationship model for the application without further code analysis. Instead, we analyzed the application source code and found likely entity relations and foreign key constraints through SQL database access operations in code segments that control vital signs forms.

Existing Data Model

The OpenEMR data model closely aligns data tables to application forms used by the user interface. This approach differs from other applications that have a clear separation of presentation, application logic and data layers. A patient encounter record and its user interface form are bound to a table *form_encounter*. During a patient encounter, a healthcare provider can attach additional information to the encounter record by linking other forms. A *form_vitals* table is linked to *form_encounter* through a *forms* table entry. In Figure 4.3, we have color coded these existing tables *form_vitals*, *form_encounter*, *forms* and *form_vitals* in yellow. These were identified

as suitable integration points for data quality trust.

Data Model Extension

Figure 4.3 shows table extensions to the data entity relationship model colored in grey. The existing table *users* forms the basis of several new data entity relationships that store a healthcare practitioner's data quality trust preferences. The existing table *form_vitals* was extended with an additional field *observation-method-id* that links a patient observation with the observation method used to generate the observation. This extended schema can support implementation of data quality trust functionality by augmenting the existing OpenEMR user interface, or it could serve as EMR data store for trust preferences with a future FHIR-enabled implementation of OpenEMR with SMART apps.

4.2.5 OpenMRS

OpenMRS [78] is a collaborative open-source project focused on supporting the delivery of health care in developing countries. It is based on the Java programming framework and a MySQL relational database model. We installed the application in our development environment and manually traced Java source code segments that generated the HTML code to display the application user interface. We found several tables in the database schema that appeared to store information related to patient observations. Further analysis of the MySQL database showed that foreign key relations were present which helped our analysis.

Existing Data Model

The database schema stores patient observations in a *obs* table with a foreign key relation to a patient encounter table *encounter*. The healthcare provider is encoded in

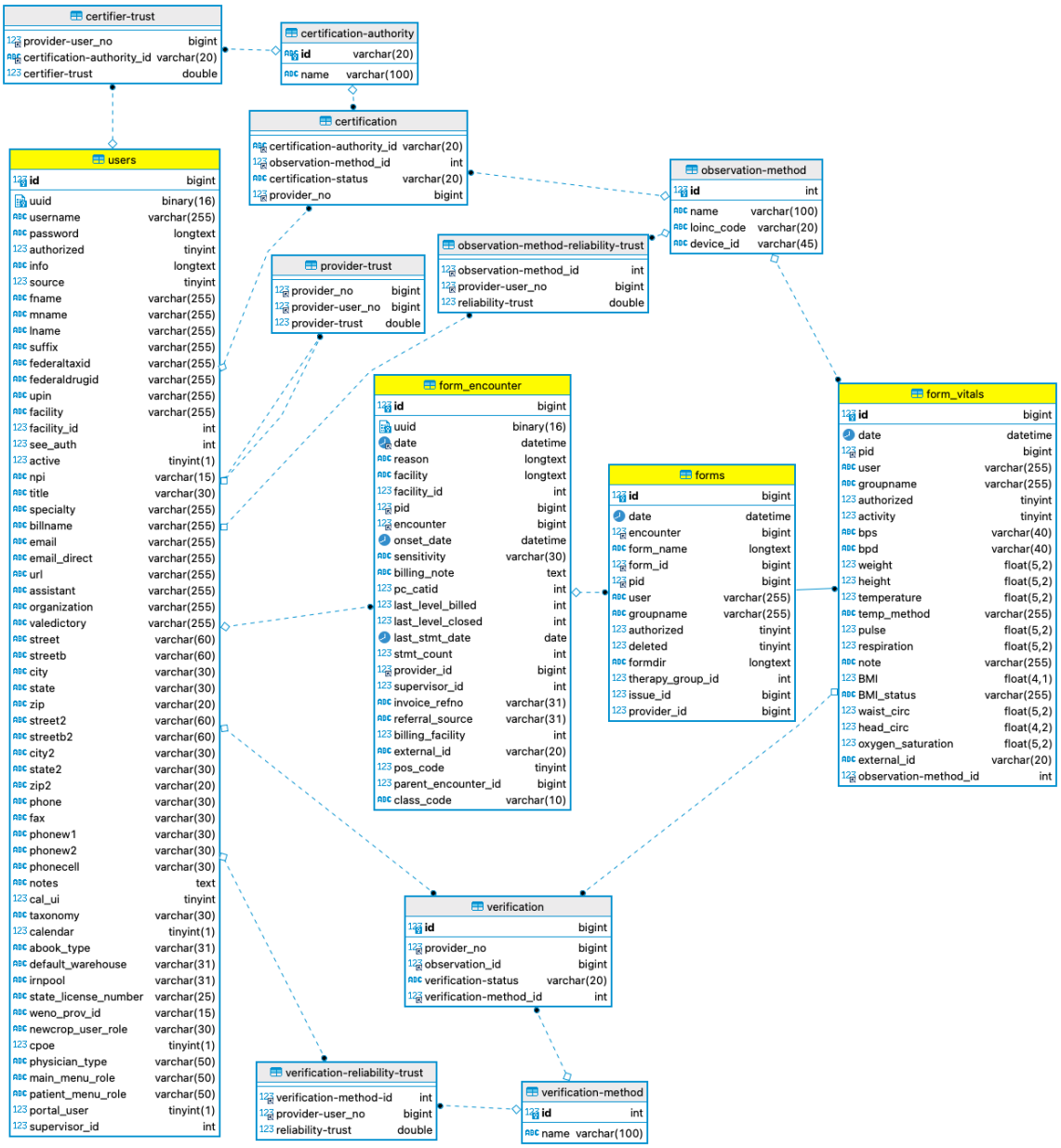


Figure 4.3: OpenEMR Database Schema Extension

a table *person* which is linked to a patient observation in table *obs* through a patient encounter in table *encounter*. In Figure 4.4, we have color coded these existing tables *form_vitals*, *form_encounter*, *forms* and *form_vitals* in dark grey. These tables were identified as suitable integration points for data quality trust.

Data Model Extension

Figure 4.3 shows table extensions to the data entity relationship model colored in grey. The existing table *person* forms the basis of several new data entity relationships that store a healthcare practitioner's data quality trust preferences. The existing table *obs* was extended with an additional field *observation-method-id* that links a patient observation with the observation method used to generate the observation. Existing table *encounter* links the patient observation to a practitioner. This extended schema can support implementation of data quality trust functionality by augmenting the existing OpenMRS user interface, or it could serve as EMR data store for trust preferences with a future FHIR-enabled implementation of OpenMRS with SMART apps.

FHIR Integration for OpenMRS

Review of the OpenMRS documentation also provided that OpenMRS has already mapped several FHIR resources and supports integration with the SMART on FHIR standard. Select resources are available through the FHIR interface [79], including FHIR Practitioner, FHIR Observation and FHIR Encounter. Every OpenMRS object stores basic provenance information about when it was created and when it was last updated in *Object.dateCreated*, *Object.creator*, *Object.dateChanged* and *Object.changedBy*. These fields can be made available through the FHIR interface using the FHIR Provenance resource. OpenMRS does not automatically capture prove-

nance chains of derived data entities. Work on mapping additional FHIR resources to the OpenMRS data model is currently in progress by the open-source developer community.

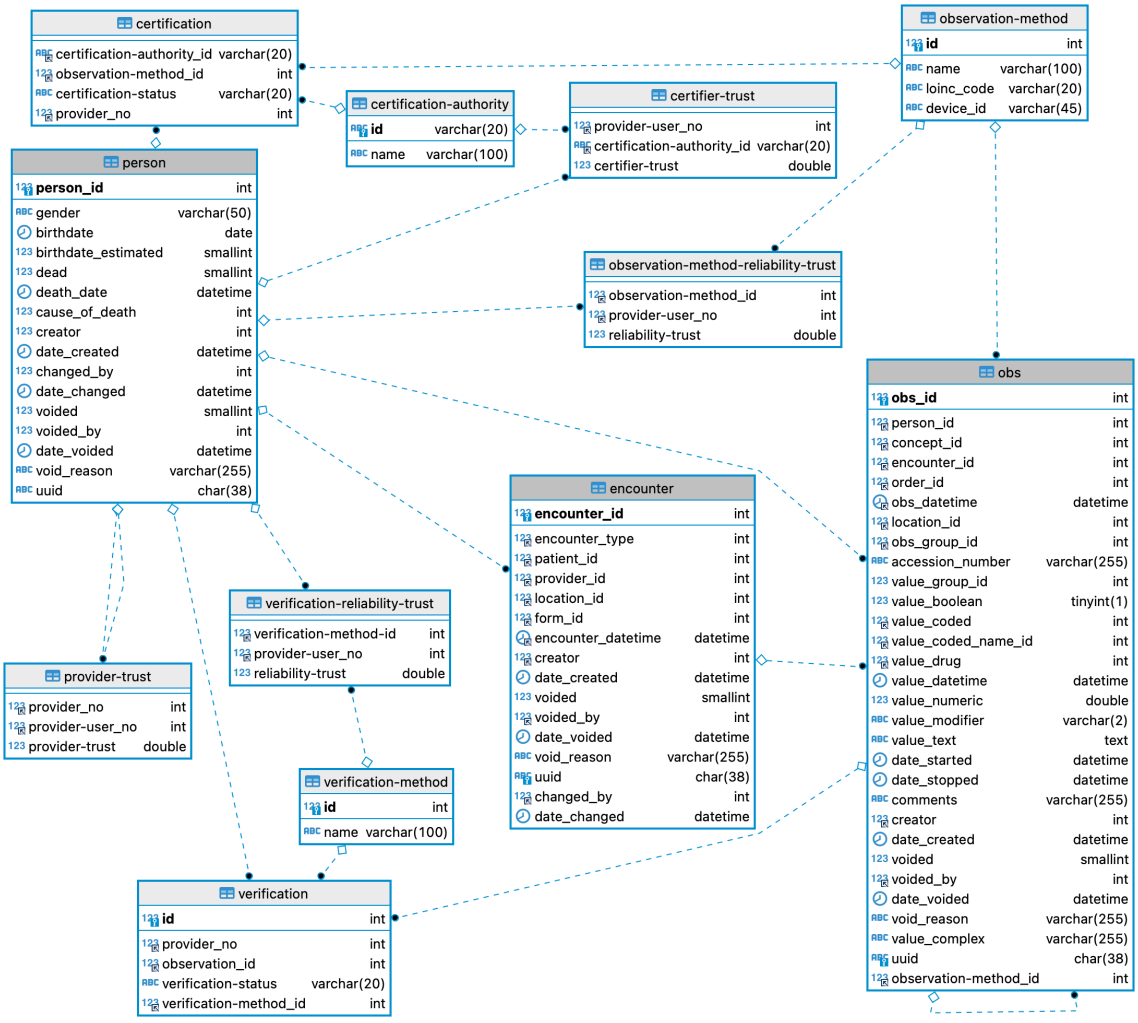


Figure 4.4: OpenMRS Database Schema Extension

4.2.6 OpenEHR Standard

Integration with OpenEHR-based systems follows a different approach. OpenEHR is an open standard created and maintained by the openEHR Foundation to facilitate

interoperability of electronic healthcare data. The openEHR architecture is based on a two-level modeling approach that separates domain concepts as defined by domain experts in archetypes and templates, from the technical implementation as defined by software engineers with reference models. [80]

Archetypes and Templates

OpenEHR Archetypes consist of formal models of a domain concept that can be translated to any implementation programming language [81]. Archetypes are expressed in the Archetype Definition Language and can be modeled using an iterative process with six steps [82]: collecting data requirements, normalizing data elements, organizing domain concepts, searching corresponding archetypes, editing archetypes, and reviewing archetypes. Archetypes contain the maximal data set of a clinical domain concept. Templates are then used to assemble and constrain archetypes for context-specific purpose. For example, most clinical use cases will not require the complete blood pressure archetype with all listed elements so a use case template defines which elements are required and what use-case specific data constraints and terminology bindings apply.

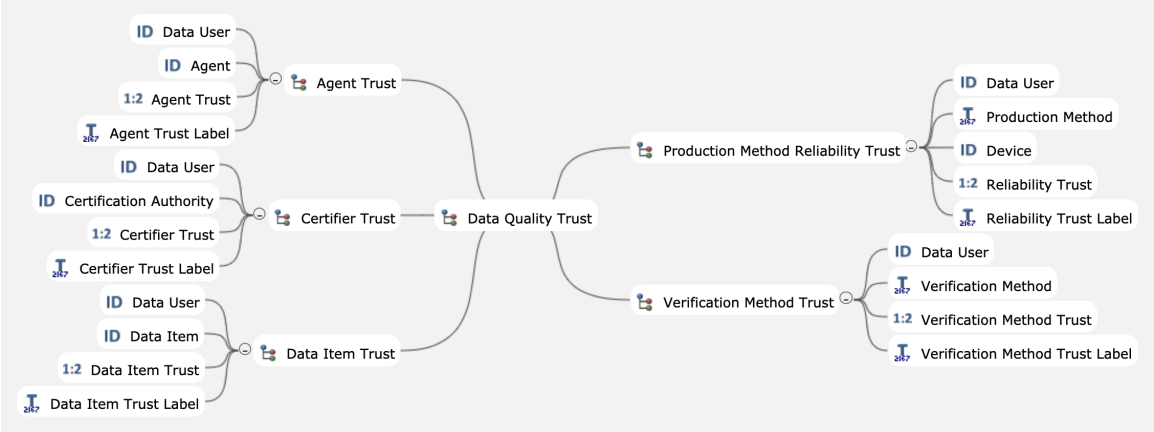


Figure 4.5: OpenEHR Data Quality Trust Archetype Mind Map

Next, we show how the domain concept of data quality trust can be modeled with an archetype and templates. In Figure 4.5, we provide a mind map of the elements required to conceptually represent data quality trust. Figure 4.6 further details this mind map into an archetype tree structure, using the Archetype Designer tool and repository provided with OpenEHR. Data quality trust is modeled as a OpenEHR Cluster which contains several sub clusters with child elements that are arranged in a tree structure. The cluster defines all data elements required to represent data quality trust, as the maximal data set for all use cases.

For each of the sub-clusters, a template defines the use case specific constraints and terminology bindings, including the membership functions that define associations of trust labels with levels of trust. These OpenEHR templates can then be used to generate APIs, XML schema definitions, user interface forms, or other application elements. An archetype can also be extended with other archetypes. For example, the data item trust concept and its representation in the Data Item Trust archetype cluster can be added to an existing Blood Pressure archetype to specify a data user's data quality trust in a blood pressure observation. Compositions of several archetypes to model data provenance in a particular clinical use case are possible.

Reference Model and Provenance

The technical implementation of OpenEHR-based applications is driven by a common reference model [80]. This reference model contains a predefined set of classes that represent the structure of an EHR, and the archetypes as specific concepts that restrict reference model classes. The reference model for *Demographic* which defines system actors with their roles and capabilities in OpenEHR. A reference model for *Entry* provides the general definition of EHR data elements committed to the EHR. Based on these classes, OpenEHR keeps detailed provenance meta data on agents and

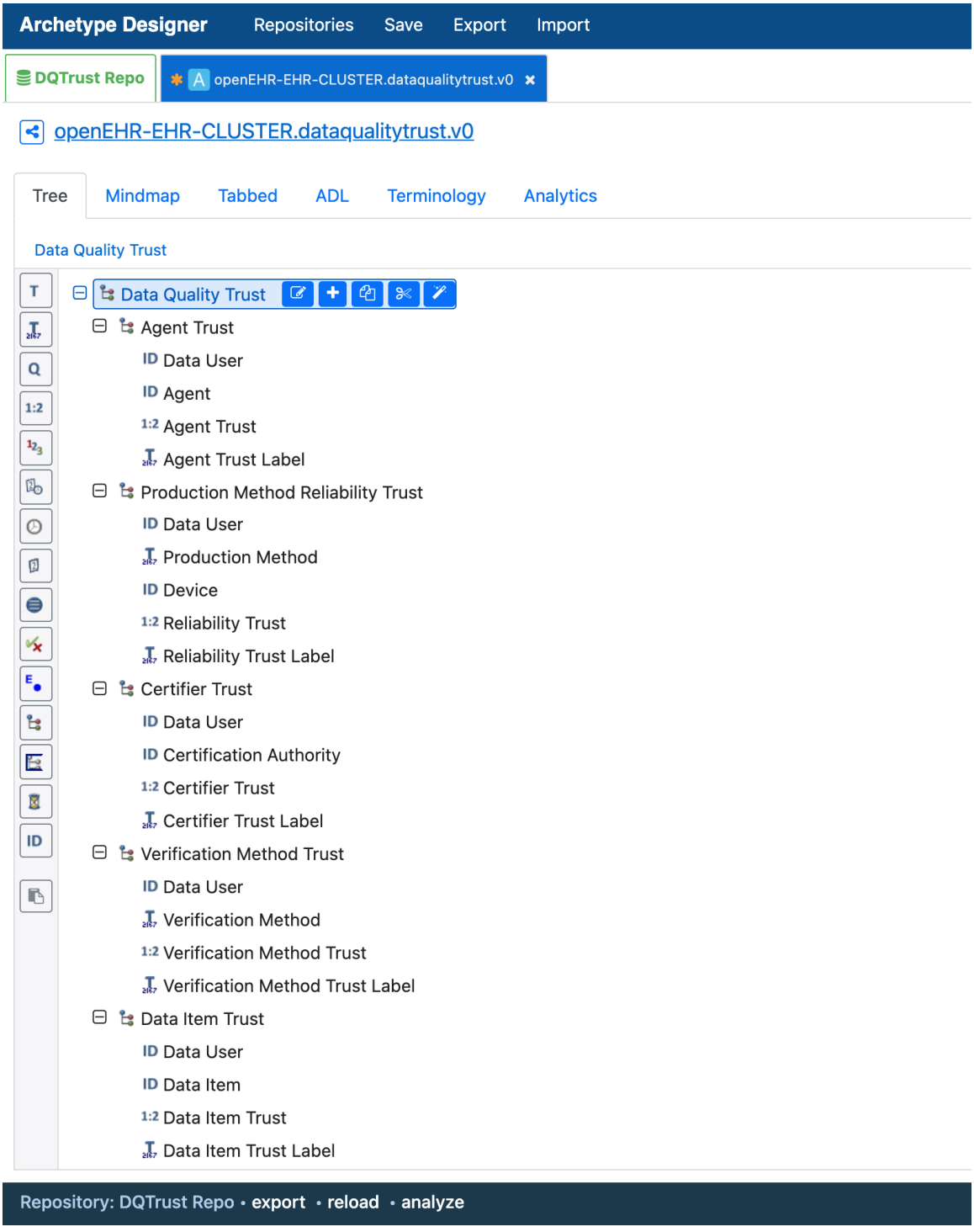


Figure 4.6: OpenEHR Data Quality Trust Archetype Tree

data transformation activities for data generated or modified by systems that implement the OpenEHR reference model. In addition, the *FEEDER_AUDIT* class defines the semantics of an audit trail for EHR data obtained from external non-OpenEHR feeder systems. This approach to provenance meta data allows for the construction of detailed provenance trees based on the audit trail and versioning data captured when information is committed to the EHR. Availability of detailed provenance information differentiates the OpenEHR approach from many other legacy EMR systems that do not support detailed object provenance. The combination of detailed data item provenance meta data of agents and data production or verification methods enables complex applications of data quality trust with multi-level provenance trees, as described in Chapter 3.

4.2.7 Trust Inference Computation in Legacy Systems

Proposed data model extensions of OSCAR, OpenEMR and OpenEMR have addressed data integration of data quality trust with legacy systems. To compute data quality trust in these legacy systems, algorithms for fuzzy inference need to be implemented either within the existing Java application environments or through function calls to external system components. In particular, those legacy applications that are already in the process of creating FHIR integration points may choose to create a REST API architecture with a data quality trust web service. Application components of such a web service could likely be reused for a FHIR interface server once the legacy application is fully FHIR-enabled. Java libraries with fuzzy logic functions that implement the Mamdani method are readily available as part of several open-source projects.

4.3 User Interface for Data Quality Trust

We shift our focus to the visualization of data quality trust as part of EMR system user interfaces. Zahabi et al. [83] reviewed existing literature on EMR safety and usability to provide guidelines for designing EMR user interfaces. We combine this research with prior research on visualization and cognitive user interface design to integrate contextual data quality assessment results with EMR user interfaces. We also searched for visualization guidelines beyond the medical domain, but there is only limited existing research on user interaction support for assessing contextual data quality [84].

Cognitive engineering principles by Gerhardt-Powals [85] were developed in the context of submarine warfare to support users in high stress situations. The findings are still relevant, despite recent advances in computer technology and user experience engineering. Ripalda and colleagues [86] have linked Nielsen's heuristics with the principles of perception of Gestalt, offering references to developers and usability experts that allow for the generation and evaluation of mockups and prototypes. Marcilly et al. [87] consolidated existing research evidence to six high-level design guidelines for designing medication alerts. While primarily targeted towards medication alerts in clinical decision support, these guidelines can be useful for designing data quality trust notifications during clinical decision making as well. Research by Ricca et al. [88] has shown the effectiveness of using screen mockups in software engineering and the improvements in terms of understandability of functional requirements when screen mockups are present.

4.3.1 Dual Processing and the Heuristic Systematic Model

Watts et al. [89] developed a theoretical model for understanding contextual information quality assessment based on a family of dual process theories. The heuristic systematic model distinguishes human decision making in heuristic and systematic processing modes. While in heuristic processing mode, humans tend to use intuition to make quick decisions using limited cognitive effort. Lack of expertise for the decision task, time constraints, and distractions are factors that can increase reliance on heuristic processing. Systematic processing requires higher cognitive effort to process ambiguous decision tasks. Individuals are not able to process all informational stimuli in their environment and are continuously faced with trade-off decisions about what tasks can be processed heuristically, and what tasks require the additional cognitive effort for systematic processing. Taft and colleagues [90] leverage the insights from dual process theory to adapt and map Nielsen's design heuristics to clinical decision support systems.

4.3.2 Development of User Interface Screens

Development of user interface screens followed an iterative prototyping and experimentation process. We created an initial set of two screens that followed the "list-detail" screen pattern of a list view of data items and a detail view to display data quality trust meta-data for a selected individual data item in the list, as shown for *Data Item 2* in Figure 4.7. Next, we reviewed adaptations of this basic pattern and found that overview screens that plot individual data items on a timeline graph are commonly used to visualize the history of measurements in EMR systems. We devised and refined screen types in several iterations based on the literature for heuristic and systematic cognitive processing modes, and we applied the usability guidelines above.

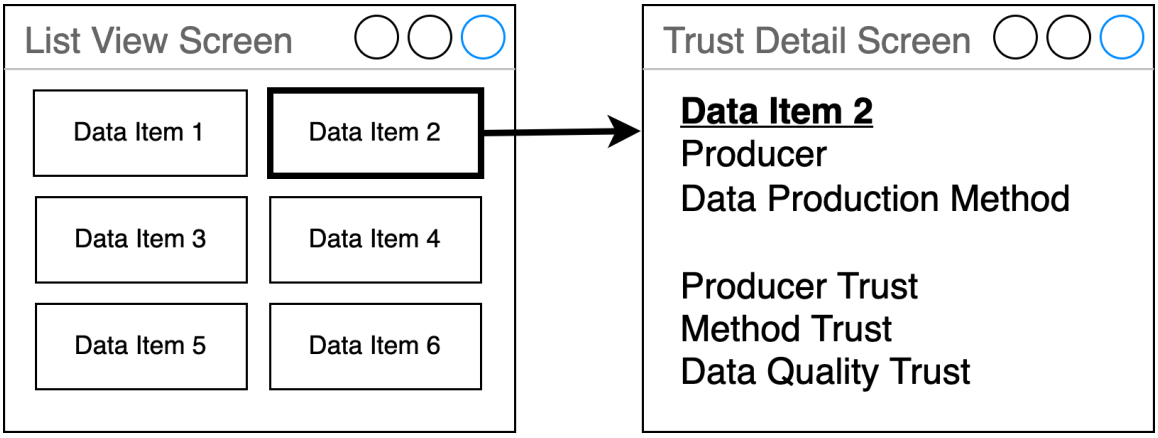


Figure 4.7: Mockup Screen: List View and Trust Detail View

4.3.3 Visualization for Heuristic Processing

We created two screen types to support a user in heuristic processing mode by integrating data quality trust with a user interface. The principles that we applied to these two screen types can be extended to other screens as well which support other application contexts for data quality trust beyond EMR systems.

Quick View Screen

The main interface for heuristic processing is a *Quick View* screen that supports a practitioner during a patient encounter allowing the practitioner to engage with and focus on the patient. Layout of the screen is driven by the need to simplify all human computer interactions so the practitioner is not unnecessarily distracted away from the patient. Clinical data items needed for a clinical decision are displayed with a simplified result of a data quality trust check on these items. In this situation, the practitioner uses the application in heuristic processing mode and does not have the time to systematically analyze further details about clinical data items, and how trust

of data items is calculated. If the data quality trust check detects one or more data items with low trust, a warning message should be displayed advising the practitioner to proceed with caution and to seek additional detail and context for the data items considered. The warning effectively pulls the user from heuristic processing into systematic processing mode to ensure focused attention and alertness. The *Quick View* interface seeks to address the requirements of Nielsen's design heuristics and other cognitive design principles. This screen should be augmented by a legend briefly explaining all *Quick View* screen elements. Additional help resources should be linked from this screen for a detailed explanation of graph elements.

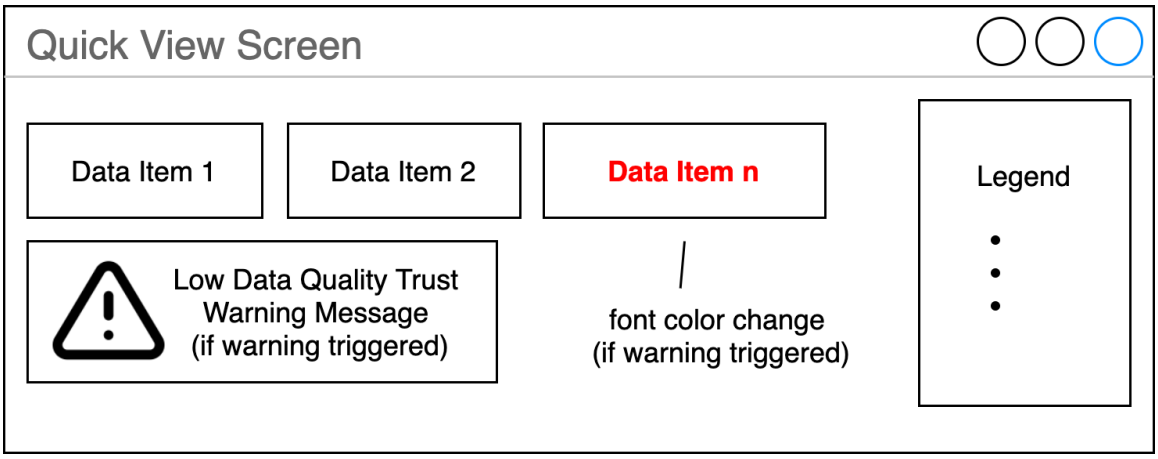


Figure 4.8: Mockup Screen: Data Quality Trust Quick View

Historical Overview Screen

Should a practitioner require additional data items to make a decision while in heuristic processing mode, a *Historical Overview* screen can display a chart of data items over time with historical trend information that allows for heuristic pattern analysis of general trends and potential outliers in the dataset "at a glance".

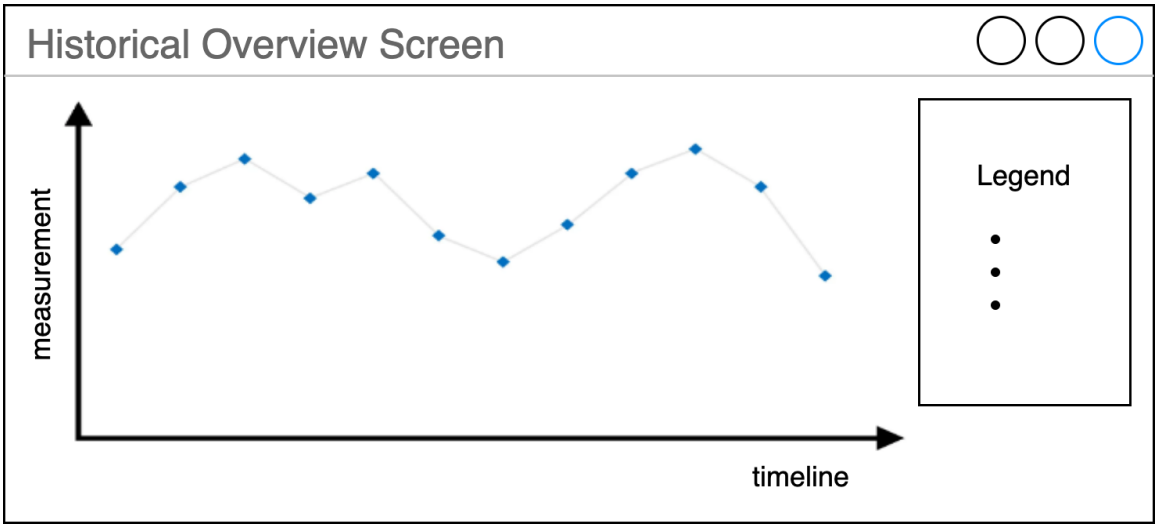


Figure 4.9: Mockup Screen: Data Quality Trust Historical Overview

4.3.4 Visualization for Systematic Processing

We created two additional screen types to support a user in systematic processing mode. If the information needs of the user cannot be satisfied through information in the *Quick View* or *Historical Overview* screens, more detailed screens allow for systematic processing and further analysis of data quality information in context. Prior research has identified numerous ways to display detailed data quality information which may be suitable for a particular implementation context. Heat maps, density plots or box plots for confidence intervals can provide detailed visualizations of data quality measures that support systematic processing [52, 53, 54, 55]. A simple filtered overview screen based on trust labels and detailed view screen displaying data item provenance can help support systematic processing by a user.

Historical Filtered Screen

We created a *Historical Overview* screen with a data quality trust filter to filter out data items with different data quality trust labels, as selected by a user. While this requires additional cognitive effort from the practitioner, analysis of the filtered view can quickly remove or show groupings of lower data quality trust allowing the practitioner to proceed with the clinical decision once satisfied that a sufficient number of data items with acceptable data quality trust support the clinical decision. This functionality may be particularly important for a high risk clinical decision where only high trust data should be used. It can also be used to identify low trust data items that may be the result of a defective data production method or low trust healthcare practitioner.

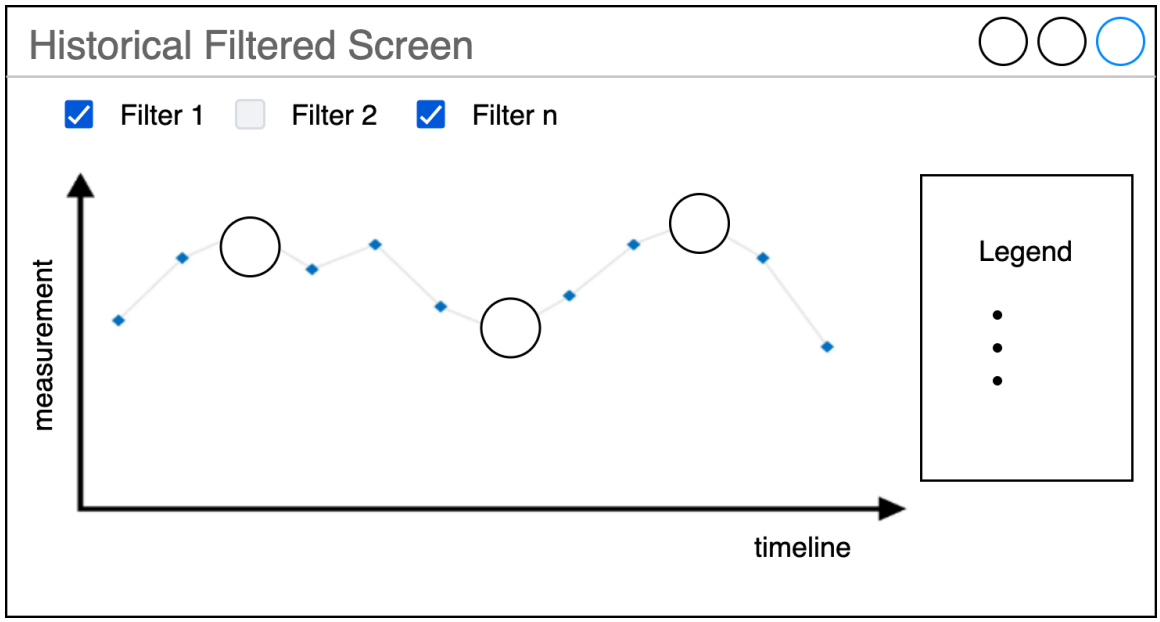


Figure 4.10: Mockup Screen: Data Quality Trust Historical Filtered View

Detail Provenance Screen

A *Detail Provenance* screen provides contextual data quality information for a data item in focus. This screen displays meaningful information about data item provenance including the producer of the data item, the data production method, trust assignments made for producer and method, and the resulting data quality trust result. For data items that were derived from multiple levels of antecedent data items with provenance information, only the immediate antecedents should be displayed.

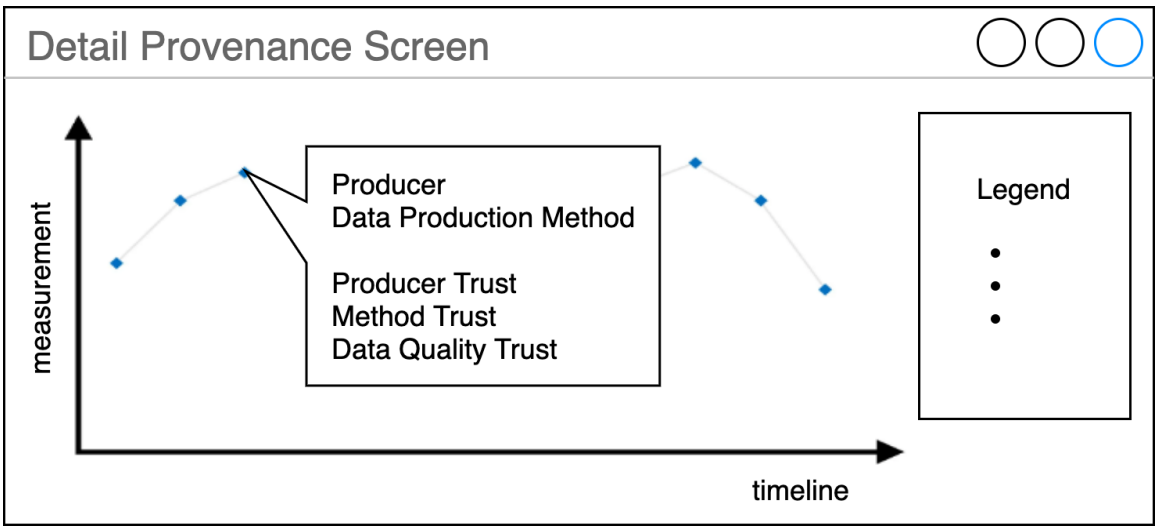


Figure 4.11: Mockup Screen: Data Quality Trust Detail Provenance

4.3.5 User Interface Validation with Design Principles

We proposed four user interface screen types aimed at visualizing data quality trust assessment results in ways that may allow a clinical user to consider data quality trust during clinical decision making. Normally, such guidance could be validated through a user study but such a study was outside the scope of our research project. Instead, we reviewed Marcilly's six meta-principles [87] for usability design (Table

4.3) and Gerhardt-Powals' ten cognitive engineering principles [85] to demonstrate how design decisions for proposed visualization screens align with these principles. A key limitation of this review was that it was conducted by a single author without independent verification. This may have introduced researcher bias where additional authors could have improved the review process.

Table 4.3: Usability Design Principles and Design Decisions

Meta Principle	**Design Decision**
A. Improve the system's signal-to-noise ratio	A low data quality trust warning message is only displayed if a warning is triggered.
B. Support collaborative work, advocate a team approach, and make the system a team player	This principle was not applied to the four proposed screens but should be considered for system implementation.
C. Fit with clinicians' workflow and their mental model	Differentiated screens for heuristic and systematic cognitive modes support this principle.
D. Display relevant data within the alert	The data quality trust alert and label color coding direct the user's attention to the data item with low data quality trust.
E. Make the system transparent for the user	A legend provides explanations of the screen elements and additional help resources are linked to provide detailed explanations of screen elements.
F. Include actionable tools within the alert	This principle was not applied to the four proposed screens but should be considered for system implementation.

Ten cognitive engineering principles [85] are listed in bold, followed by the design decisions made when developing the four screen types.

Principle 1. Automate Unwanted Workload. Visualization of data quality trust seeks to automate unwanted workload to free cognitive resources for high-level tasks. This was reflected in all four screen types.

Principle 2. Reduce Uncertainty. A minimalist *Quick View Screen* interface allows the user to operate in heuristic processing mode as much as possible to reduce cognitive load. Uncertainty is reduced by displaying data on that screen in an intuitive manner to reduce decision time and error.

Principle 3. Fuse Data. Lower level data, such as data producer and production method reliability meta data, as part of the *Detail Provenance Screen*, are fused into a higher level data quality trust aggregation to reduce cognitive load.

Principle 4. Present New Information with Meaningful Aids to Interpretation. Screens present new information with meaningful aids to interpretation. A legend on each screen guides the user with the interpretation of color coding for data quality trust concerns.

Principle 5. Use Names that are Conceptually Related to Function. Display names and labels are context-dependent to improve recall and recognition. Data quality trust labels are specific to the user context. Implementation of recall and recognition features will depend on the app that data quality trust is embedded in.

Principle 6. Group Data in Consistently Meaningful Ways. Data is logically grouped within screens and consistently grouped across screens to decrease information search time. This principle was addressed in all proposed screens.

Principle 7. Limit Data-Driven Tasks. Colors and graphics illustrate the *Quick View* and *Detailed History View* to reduce the time spent assimilating raw data.

Principle 8. Display Only the Information Needed at a Given Time. Proposed data quality trust screens display only the trust information needed by the data user at a given time. This is achieved by showing varying levels of provenance detail needed for heuristic and systematic processing modes.

Principle 9. Provide Multiple Coding of Data. Proposed data quality trust screens provide data in varying formats and levels of detail to promote cognitive flexibility and satisfy user preferences. This principle is addressed through the differentiation of screens for heuristic and systematic processing modes, as well as through the user-specific definition of trust preferences.

Principle 10. Practice Judicious Redundancy. Some judicious redundancy is provided across proposed screens, for example the same *Data Quality Trust Detail* could be accessed from both *Quick View* and *Detailed History View* screens.

4.4 Summary

This chapter showed how data quality trust can be integrated with EMR systems. We introduced several FHIR interface extensions to allow for the transmission of data quality trust over FHIR interfaces. We also analyzed four legacy EMR applications and showed how data quality trust can be integrated with applications that have not been FHIR-enabled yet. To visualize data quality trust in EMR systems, we proposed four user interface screen types aimed at visualizing data quality trust assessment results in ways that support decision making for a clinical data user.

Research objectives for visualization were addressed to support decision making in heuristic and systematic cognitive processing modes.

The next chapter will apply data quality trust to a clinical example context with a detailed computation example and visual prototype. It will demonstrate feasibility of integrating the data quality trust approach with an existing clinical workflow and several usage scenarios. We will show how data quality trust could also be applied to more complex clinical examples.

Chapter 5

Clinical Example with Prototype

This chapter introduces pediatric hypertension as a clinical example for a data quality trust slice of life example. An existing SMART on FHIR app provides the basis for a prototype that demonstrates that the integration of data quality trust with the user interface of an interoperable EMR application is feasible. After introducing the Blood Pressure Centiles App, the chapter describes how the existing app architecture can be extended with additional components to compute and visualize data quality trust for blood pressure patient observations. A detailed computation example demonstrates clinical workflow integration and simulates the data quality trust computations of the app step by step. We show how data quality trust model and method could also be applied to more complex clinical examples. As part of the validation research process (Figure 1.2), the chapter assures relevance of the novel design pattern by applying it to a clinical example context. Main research methods embedded in this chapter are conceptual modeling, prototyping, simulation, lab testing, and scenario evaluation.

5.1 Clinical Example

We begin by introducing pediatric hypertension as a simple and relatable real-world example to demonstrate the data quality assessment method. The data provenance tree for pediatric hypertension diagnosis has multiple levels, yet it is compact enough allowing us to demonstrate the data quality trust computations for each leaf and node.

5.1.1 Pediatric Hypertension

Diagnosis and treatment of pediatric hypertension is a growing problem in many countries. For participants aged 12-19 years in the 2001-2016 US National Health and Nutrition Examination Survey, prevalence of hypertension was between 1.3%-4.2% [91]. 75% of cases of pediatric hypertension and 90% of cases of pre-hypertension in children from 3 to 18 years of age go undetected [92]. Diagnosis of the disease at the point of care is complicated by the need to account for constantly changing body size when interpreting blood pressure measurements. Manual calculations are time-consuming, yet assessment of blood pressure percentiles is medically recommended from the age of 3 onward [92].

Because hypertension is a condition that is prone to misdiagnosis due to data quality issues with underlying blood pressure patient observations, it could benefit from a data quality assessment solution for clinical data users that reduces the risk of misdiagnosis. Blood pressure measurement is a simple routine test performed in primary care physician offices, hospitals, pharmacies, and home care settings, especially for hypertension control and diagnosis. Numerous factors can influence blood pressure measurements, such as the environment, the method or technique, the patient, and the equipment used [93]. Observed differences between ambulatory and

physician-office blood pressure measurements suggest that cases of true hypertension are often not identified [94]. Several studies have evaluated the impact of measurement uncertainty and inaccuracies in blood pressure measurement on hypertension diagnosis [95]. These data quality concerns with measurements and interpretation are exacerbated when diagnosing hypertension in pediatric patients [96, 92, 97].

At the core of a pediatric hypertension diagnosis is the calculation and assessment of a blood pressure percentile, or "blood pressure centile". Using percentile rank rather than the absolute blood pressure measurement is necessary for children because their blood pressure changes substantially during normal growth and development during childhood.

5.1.2 Blood Pressure Centiles App

Boston Children's Hospital developed a SMART on FHIR application ("BP Centiles") that automates blood pressure percentiles calculations at the point of care. The BP Centiles app [98] reads a child's relevant vitals and calculates systolic and diastolic blood pressure percentiles normalized by age, sex, and height. The last three blood pressure percentile pairs form the basis of a hypertension diagnosis. The app also includes a calculator and a graphical history of the child's blood pressure percentile to enable full screening at each visit. It is available under open source license which makes it attractive as the foundation for an artifact prototype in this research project. The following sections describe the screens found as part of the original application before modifications for data quality trust.

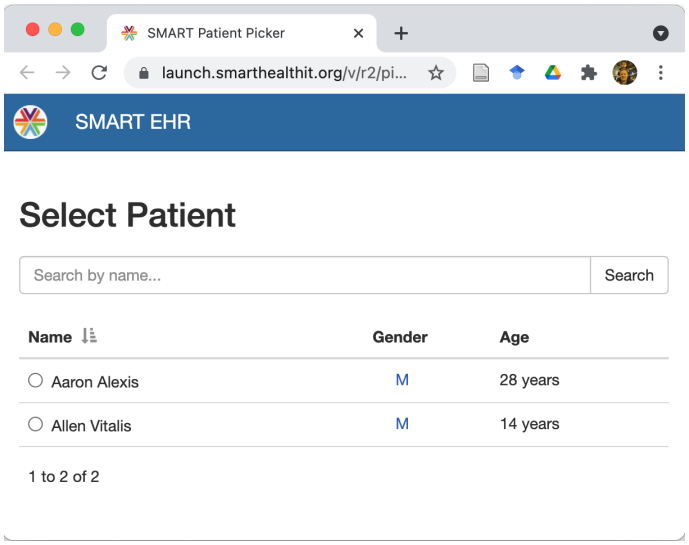


Figure 5.1: Original Screen: SMART EHR Select Patient

Patient Selection Screen

Figure 5.1 shows the initial patient screen that allows for a clinician to search for and select the patient in scope for the encounter.

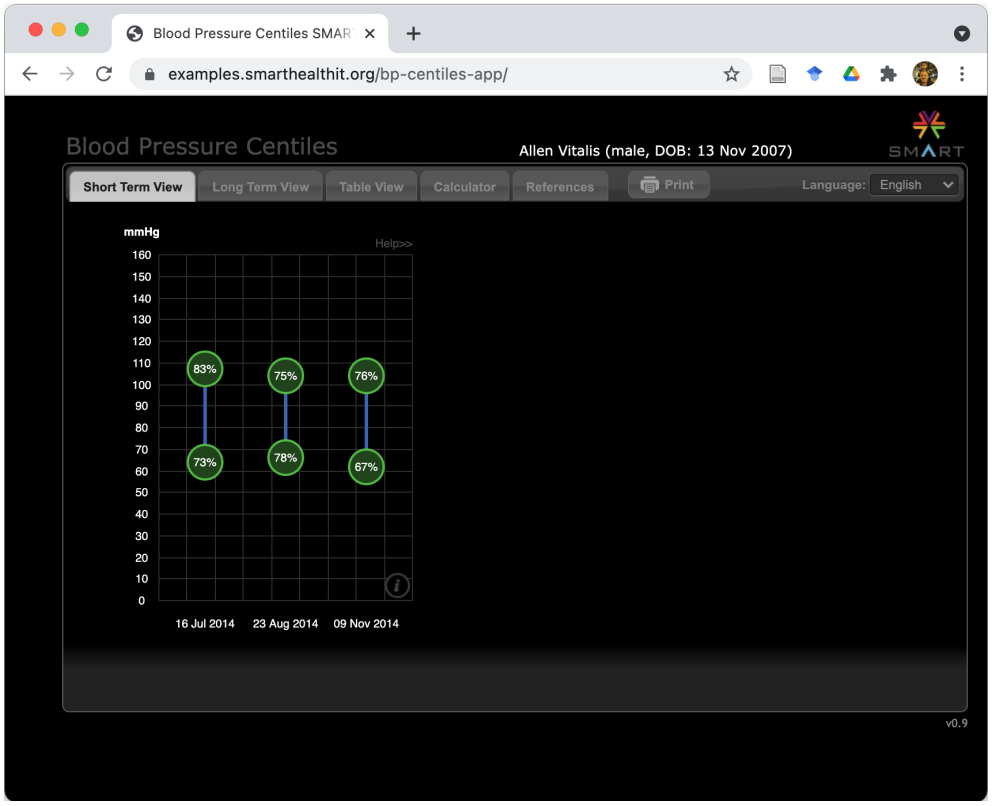


Figure 5.2: Original Screen: Short Term View

Short Term View Screen

Figure 5.2 is the short term view screen. It shows the last three blood pressure patient observations and percentiles. Each circle represents a systolic or diastolic blood pressure measurement. Hovering on a circle reveals more detail about the measurement. Color coding of the circle reveals at a glance whether individual readings are normal (green), prehypertensive (yellow), hypertensive (red), or hypotensive (blue). Hovering the mouse over an information icon in the bottom right corner displays a legend for the graph (Figure 5.3).

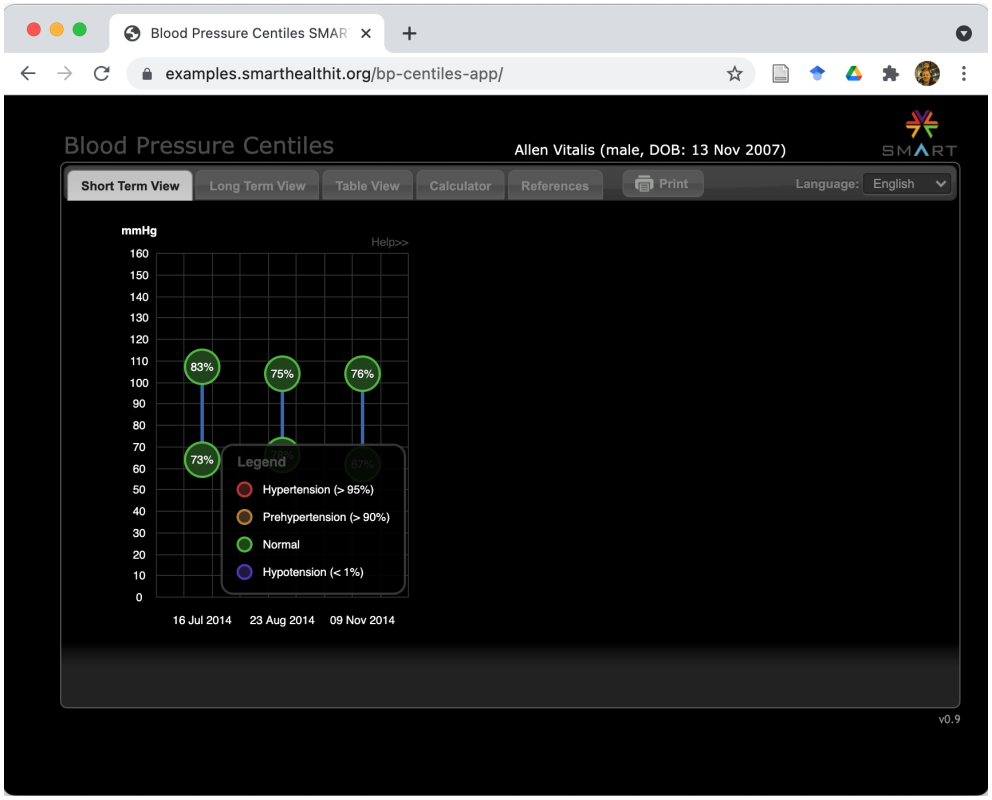


Figure 5.3: Original Screen: Short Term View with Legend

Calculator View Screen

Figure 5.4 shows the calculator screen that allows a clinician to perform blood pressure percentile calculations manually.

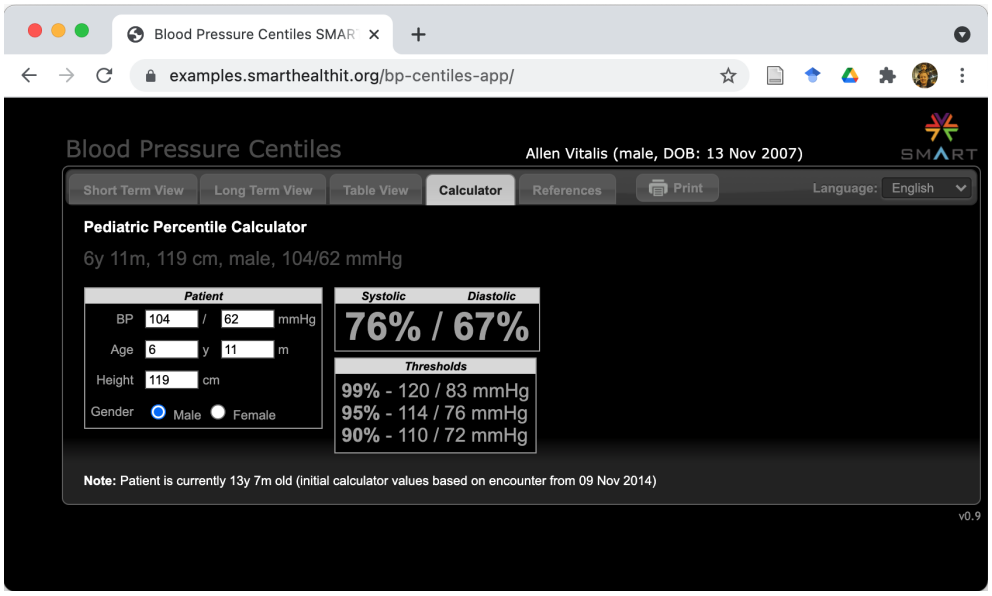


Figure 5.4: Original Screen: Calculator

Long Term View Screen

Figure 5.5 shows the long term view screen that provides the historical view of blood pressure percentiles. Each circle represents a blood pressure measurement. Hovering on a circle reveals more detail about the measurement (Figure 5.6). Hovering over an icon at the bottom right of the screen brings up a legend with color-coding information.

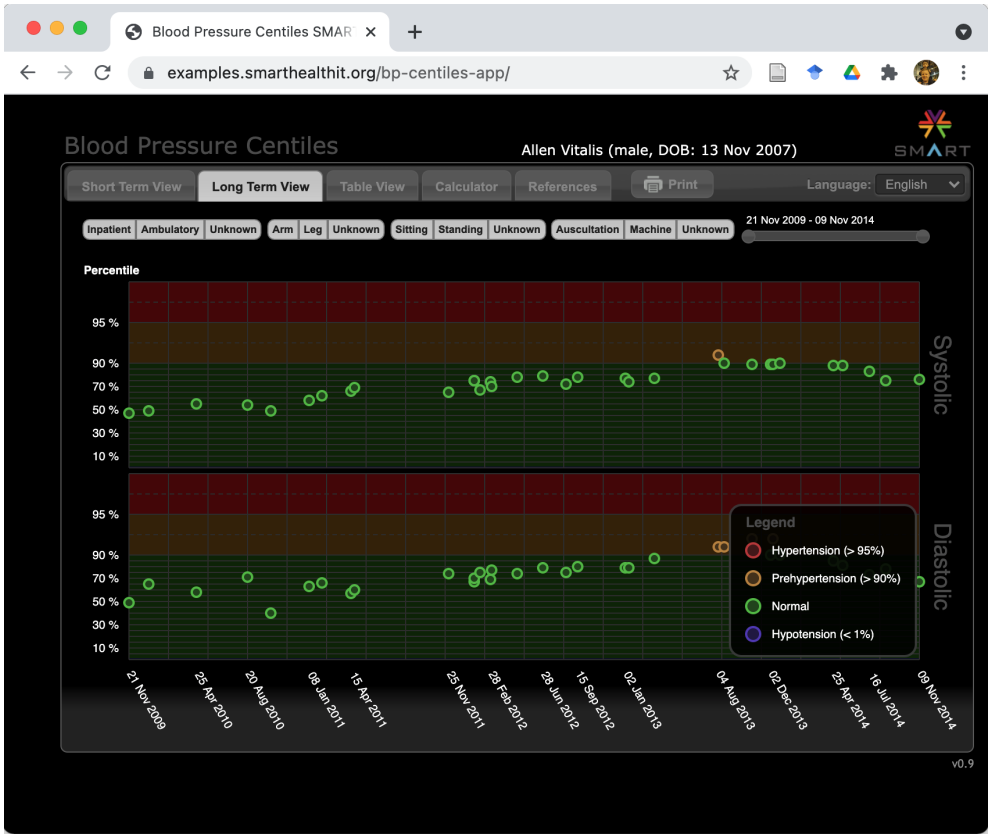


Figure 5.5: Original Screen: Long Term View with Legend

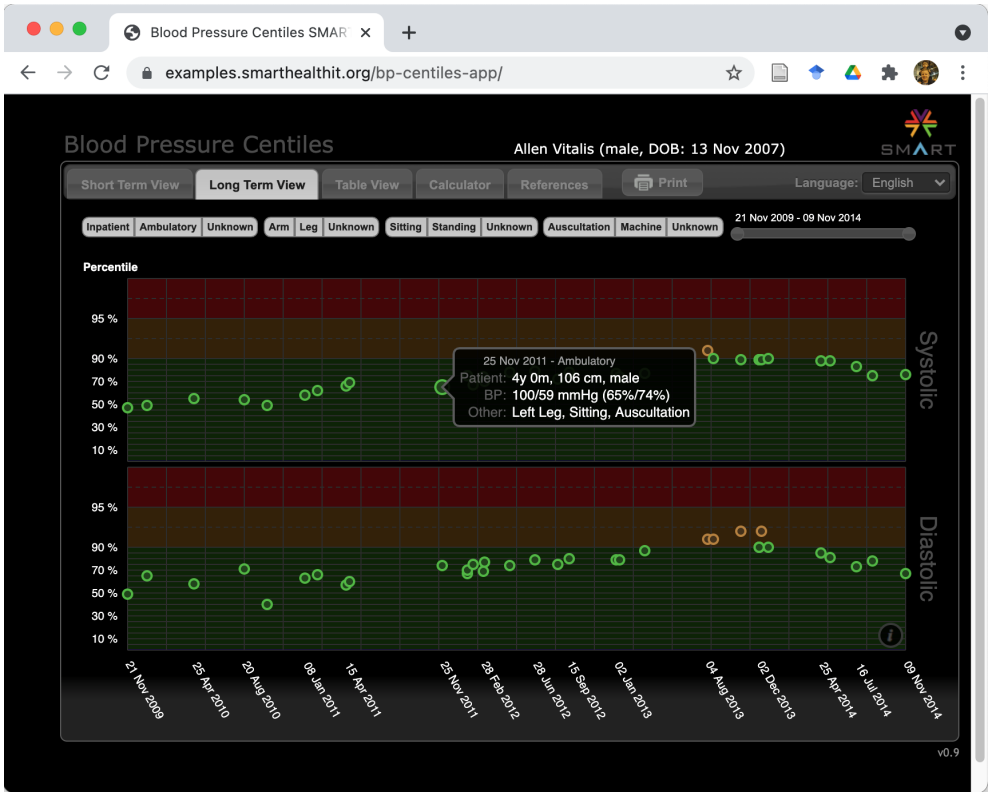


Figure 5.6: Original Screen: Long Term View with Measurement Detail

5.1.3 Hypertension Diagnosis Provenance Tree

Figure 5.7 shows the trust provenance tree for a typical hypertension diagnosis data item using our model and the BP Centiles app. Transformations t_{1-5}, t_{6-10}, t_{11-15} describe age, sex, height, and blood pressure patient observations that were recorded as part of three separate patient encounters. These observations were generated by agents and data production methods without antecedents. Transformations t_{16-21} describe the blood pressure centile calculation with four antecedent data entities (age, sex, height, diastolic/systolic blood pressure). This calculation is performed by an agent and an automated data production method within the app. Transformation t_{22} describes the diagnosis of hypertension based on three blood pressure percentile value pairs.

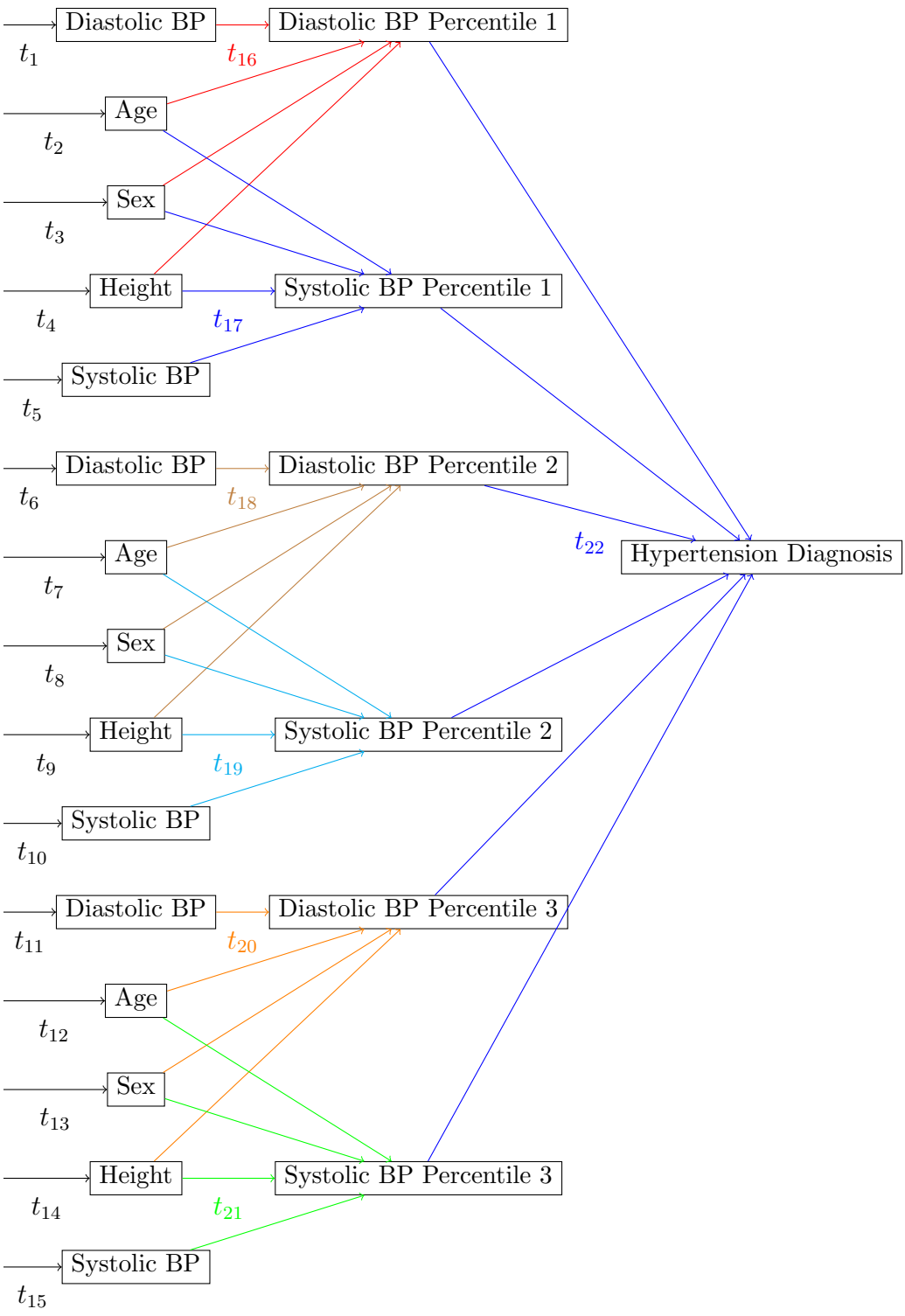


Figure 5.7: Hypertension Diagnosis Provenance Tree

5.2 Visual Prototype

In this section, we describe how the existing BP Centiles App is extended to serve as a prototype for the visual and clinical workflow integration of data quality trust with an existing app user interface and clinical example. We differentiate extensions required for functional integration from those that extend the user interface.

5.2.1 Prototyping Approach

Based on our research objective of demonstrating feasibility of the new data quality assessment approach with a clinical example, we searched for suitable clinical examples and existing SMART on FHIR applications from the SMART App Gallery [99]. In addition to the characteristics of the pediatric hypertension provenance tree mentioned earlier, we selected the BP Centiles App because it had a range of existing interface screen types that could be extended for visual integration with data quality trust.

Our prototyping approach consisted of downloading the source code from the Github repository and installing it in our local development environment. IntelliJ IDEA served as our integrated development environment with a Vagrant virtual machine management platform. To identify integration points within the existing BP Centiles app, we analyzed the source code and located the respective Java source files that generated particular user interface components. We iteratively prototyped source code modifications in JavaScript and HTML and debugged the code with the IntelliJ IDEA debugger to generate desired data quality trust visualizations. MySQLWorkbench and DBeaver served as database management systems for testing data schema extensions.

5.2.2 Functional Integration

We begin with describing the functional integration of data quality trust with the BP Centiles app. For data quality trust to be fully integrated with the app, several extensions to app architecture, app data model, user interface screens, and FHIR interfaces are required. We briefly cover extensions to the FHIR interface and a new Trust Preferences Server that would be required for a fully functioning prototype.

Extended App Conceptual Architecture

BP Centiles implements the "SMART on FHIR" architecture [5]. Figure 5.8 shows a conceptual architecture for an extended BP Centiles app.

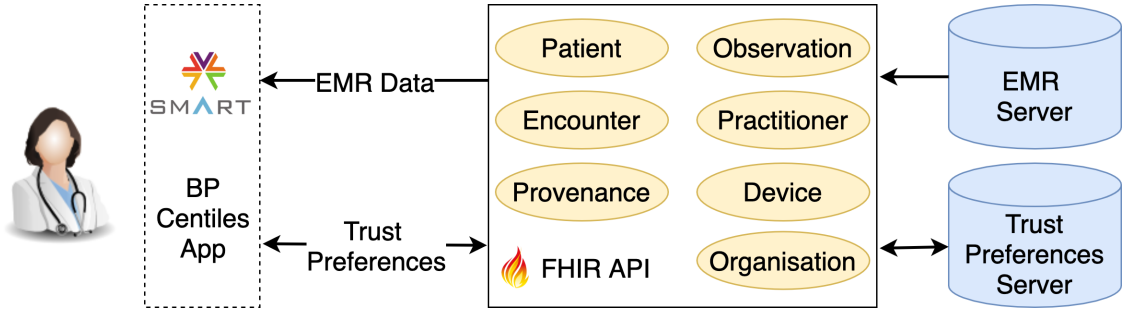


Figure 5.8: Extended App Conceptual Architecture

Model Terminology with Extended App Data Model

In its unmodified form, the app accesses the FHIR API and reads FHIR resources with EMR data. Blood pressure measurements are vital sign patient observations stored in FHIR Observation resources that are profiled with a FHIR Blood Pressure Profile. A FHIR Encounter resource links observation data items (FHIR Observation) with a patient (FHIR Patient) and clinician (FHIR Practitioner). FHIR Device provides information about any device used to generate a FHIR Observation.

In the extended app data model, the data user is a healthcare practitioner analyzing clinical data for a pediatric hypertension diagnosis at the point of care. Data production method is the method that was used to generate a patient observation. Agent is the practitioner that made the patient observation. Certification authorities are organisations that can certify production methods through the devices used for blood pressure observations. The data user can assign varying levels of trust to practitioners that produced patient observations, assign varying levels of trust to data production methods, and assign trust to certification authorities that certify agents or devices that are used as part of data production methods. Data quality probes serve as data verification methods which can be activated or deactivated by the clinical data user. We assume that a set of patient age, sex, height and blood pressure recordings made during a single patient encounter visit were made by the same practitioner.

Trust Fuzzy Membership Functions

The app uses fuzzy membership functions to define data quality trust labels of *high*, *medium* and *low*. A triangular membership function is used for the three labels by instantiating the function from Equation 3.1 with parameters for a, b, c, and d for each of the trust types (Table 5.1).

Table 5.1: Trust Membership Functions for SMART App Prototype

Trust Type	**Label**	**a**	**b**	**c**	**d**
Agent Trust	Low	0	0	0	50
Agent Trust	Medium	0	50	50	100
Agent Trust	High	50	100	100	100
Reliability Trust	Low	0	0	0	50
Reliability Trust	Medium	0	50	50	100
Reliability Trust	High	50	100	100	100
Certification Trust	Low	0	0	0	50
Certification Trust	Medium	0	50	50	100
Certification Trust	High	50	100	100	100
DQ Trust	Low	0	0	0	50
DQ Trust	Medium	0	50	50	100
DQ Trust	High	50	100	100	100

Fuzzy Association Rules

To compute data quality trust, we let the app differentiate between several data production methods based on their degree of automation in alignment with the safety classification of medical devices for blood pressure measurement (Table 3.4). Fully automated, semi-automated and manual data production methods have trust association rules that reflect their reliance on the capability of a human agent (Table 5.2).

Table 5.2: Fuzzy Association Rules for Prototype

Data Production Method	**Automation Level**
System Calculation	Fully-Automated
System Copy	Fully-Automated
Hypertension Diagnosis	Manual
BP Auscultation (Auto)	Fully-Automated
BP Auscultation (Manual)	Semi-Automated
BP Oscillometry	Fully-Automated
BP CNAP	Fully-Automated
BP Invasive	Fully-Automated
Height Measurement (Manual)	Manual

Trust Assignment

Trust computation requires a user to assign trust to agents and data production methods. This trust assignment is expressed as a number between 0 and 100. Membership functions, as defined in Table 5.1, calculate the corresponding trust label. Trust assignment to data verification methods in the app is binary, a user can choose to activate or deactivate individual data quality probes for verification.

Verification of Blood Pressure Measurements

Data quality trust in patient observations can be verified with data quality probes. For this prototype, we are interested in factors that may have temporarily increased or lowered blood pressure making a measurement unsuitable to be included for a diagnosis. The first probe makes a concordance check to verify if the patient may have had an acute illness at the time of the blood pressure measurement in question.

Figure 5.9 shows a simplified view of how a data quality probe verifies concordance of active problems and observations between FHIR resources. A type 2 data quality probe, as defined by Weber et al. [64], is implemented as a pre-condition to Systolic BP Centile display on the *Quick View* app screen. This probe checks for acute illnesses in the list of active conditions for the patient. If an acute illness is found for one of the blood pressure observations used, the measurement is marked as "not trusted" to reflect the potential issue with the measurement for use in a diagnosis.

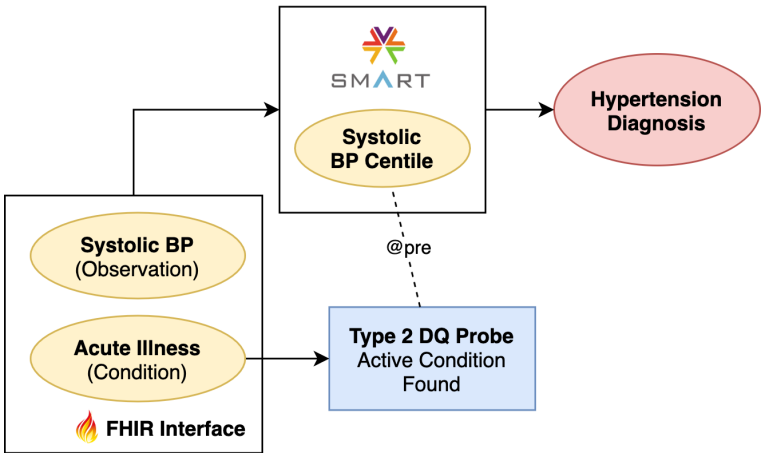


Figure 5.9: Data Quality Probe for Acute Illnesses or Allergies

A second probe verifies if the patient may have taken medication at the time of the blood pressure measurement. For example, a decongestant prescribed against a common cold illness may have temporarily increased systolic blood pressure of the pediatric patient [100]. Figure 5.10 shows a simplified view of how a data quality probe verifies concordance of active medication orders and observations between FHIR resources. If an active medication order for a certain medication is found at the time that the blood pressure observation was made, the measurement is marked as "not trusted" to reflect the potential issue with the measurement for use in a diagnosis.

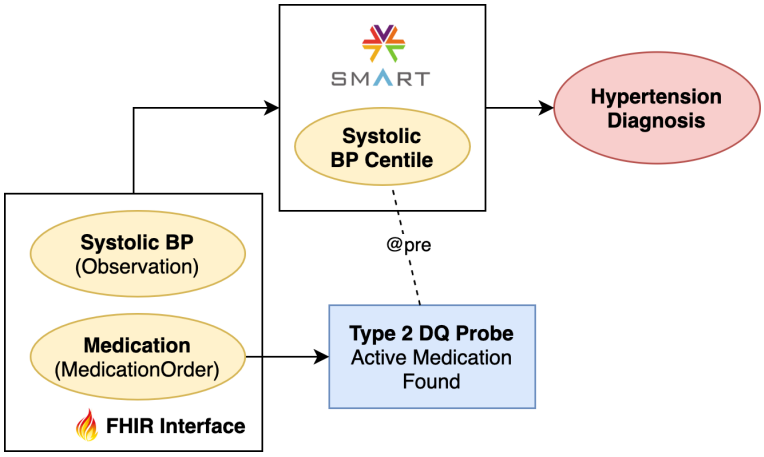


Figure 5.10: Data Quality Probe for Active Medication

Three other data quality probes check statistical plausibility of blood pressure or height observations. If the observations are found to be statistically implausible, the measurement is marked as "not trusted" to reflect the potential issue with the measurement for use in a diagnosis. Table 5.3 shows an overview of the five data quality trust probes applied as part of verification activities v_{1-5}.

Table 5.3: Probes for Trust Verification of Blood Pressure Observations

Id	**DQ Probe**	**DQ Probe Type**	**Trust Type**
v_1	Systolic BP Plausible	Statistical Plausibility	Binary
v_2	Patient Height Plausible	Statistical Plausibility	Binary
v_3	Diastolic BP Plausible	Statistical Plausibility	Binary
v_4	Check for Acute Illness	Internal Concordance	Binary
v_5	Check for Active Medication	Internal Concordance	Binary

The trust provenance tree for hypertension diagnosis (Figure 5.7) is extended with data quality probes for verification of data quality trust. Data Items resulting from t_1, t_4 and t_5 are verified with data quality probes that are applied as part of verification

activities v_{1-5}.

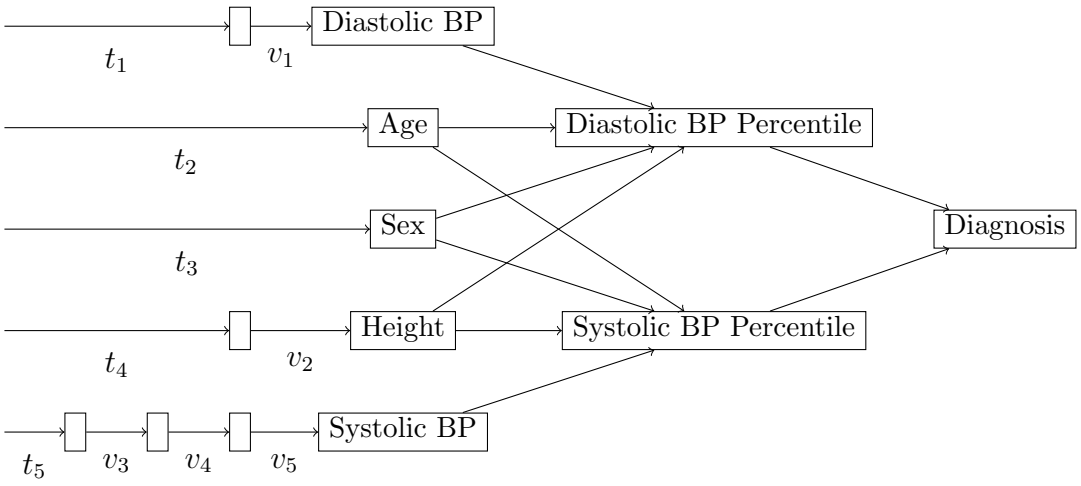


Figure 5.11: BP Centiles Trust Provenance Tree with Verification

Certification of Agents and Production Methods

Data quality trust in patient observations can be certified through practitioners and observation methods. Our prototype allows a data user to assign trust to organisations that certify blood pressure measurement devices. The premise of this is that a blood pressure measurement method that uses a certified device is trusted to be of higher reliability. This extension also allows for an organisation to certify practitioners that perform blood pressure measurement methods. A certified practitioner is trusted to have higher capability to perform blood pressure measurements. Figure 5.12 illustrates the link between certified practitioners, certified measurement methods, certification authorities, and blood pressure observations encoded in FHIR resources, as part of the data flow that ultimately results in a hypertension diagnosis process.

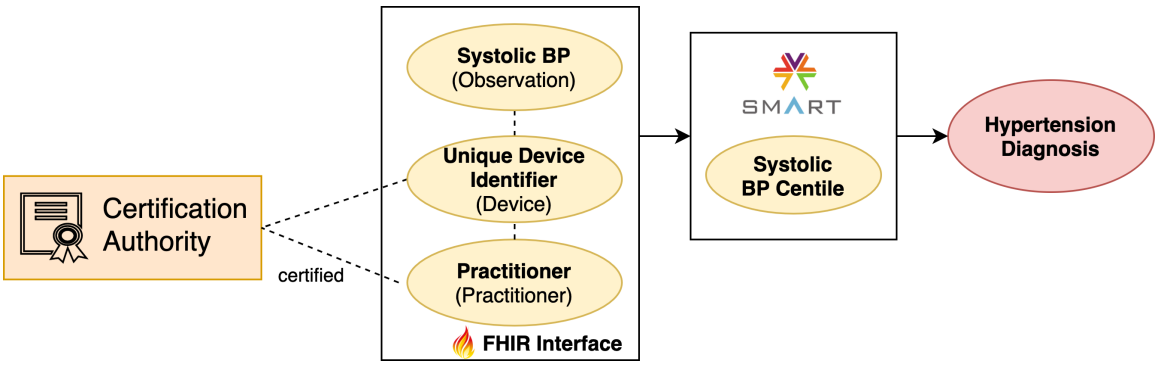


Figure 5.12: Certification of Measurement Devices and Practitioners

Recall of Defective Medical Devices

A certified medical device, as part of a certified measurement method, may be found defective prompting a trusted national regulator, such as the FDA, to revoke certification and recall the device. EMR data integration with the FDA device recall database [101] allows for automatic data quality trust recalculation of any patient observations that were affected by such a recall. Based on the data item provenance, lowered trust from antecedent patient observations that originated from a defective device would automatically cascade to other data that was derived from it or relied on it for a diagnosis. Action alerts in the EMR could prompt a clinician to review this data quality trust issue, allowing them to re-evaluate a diagnosis and correct treatment as required.

FHIR Interface and Trust Settings Database

As part of the extended app architecture, a new Trust Preferences Server is required to store user trust settings separate from the existing EMR database server. Data integration of EMR data and data quality trust occurs through an extended FHIR API which serves as an intermediary between extended BP Centiles App and the

existing data stores for EMR data and a new Trust Preferences database. A data quality trust extension to the standard FHIR API allows for standard FHIR Resource request responses to be augmented with additional required data elements containing data quality trust information. Data schema integration between EMR data and Trust Settings database can be achieved with a HAPI FHIR server. The HAPI FHIR server receives HTTP-based requests from the extended app, queries for required data elements from EMR system and Trust Preferences Server, forms a FHIR response document compliant with the FHIR extension, and responds to the app with a HTTP response including FHIR document payload.

5.2.3 User Interface Integration

We have modified existing BP Centiles user interface screens to demonstrate visualization of data quality trust with our prototype. For the data example used, detailed information on the background computations required to generate the data elements on the screens is provided in Section 5.3.

Quick View Screen

We have renamed the *Short Term View* screen to a *Quick View* that displays the result of a data quality trust check on the three most recent blood pressure observations (Fig. 5.13). This view supports a patient encounter where the practitioner engages with and focuses on a patient. In this situation, the practitioner uses the app in heuristic processing mode and does not have the time to systematically analyze the detailed history of blood pressure measurements, and how trust of data items is calculated. If the data quality trust check detects one or more blood pressure measurements with low trust, the measurement font color is changed to red, and a warning message is displayed advising the practitioner to proceed with caution and to review detail

measurement history for any safety concerns. The warning effectively pulls the user from heuristic processing into systematic processing mode to ensure focused attention and alertness.

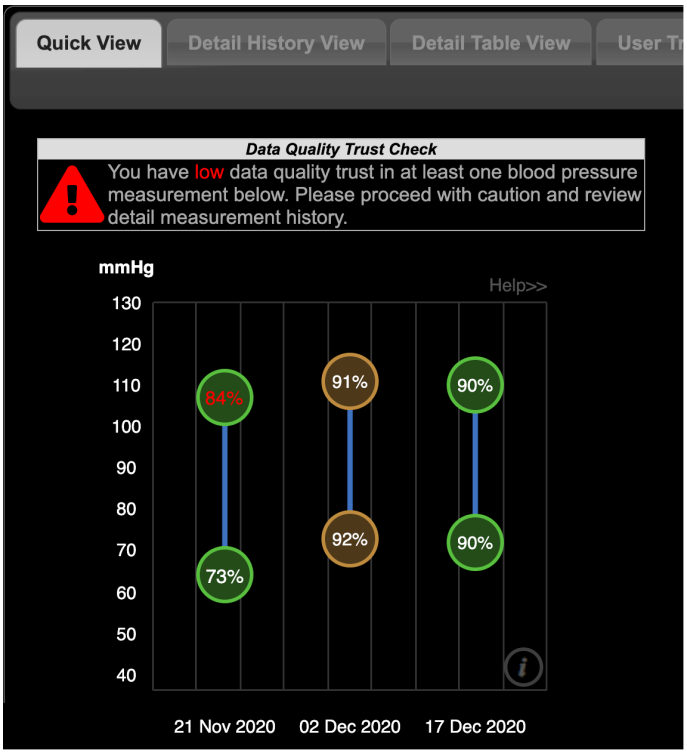


Figure 5.13: Prototype Screen: Quick View with Low Trust Warning

Screen Example (Figure 5.13): The practitioner reviews the three most recent blood pressure observation pairs using the *Quick View*. In the background, the system computes data quality trust for each observation displayed.

The first observation pair was made on 21 Nov 2020. Systolic blood pressure was in the 84th percentile and diastolic blood pressure was in the 73th percentile. Because the observations were at the 90th percentile or less, they are circled in green color.

The second observation pair was made on 2 Dec 2020. Systolic blood pressure

was in the 91th percentile and diastolic blood pressure was in the 92th percentile. Because the observations were between the 90th percentile and 95th percentile, they are circled in orange color.

The third observation pair was made on 17 Dec 2020. Systolic blood pressure was in the 90th percentile and diastolic blood pressure was in the 90th percentile. Because the observations were at the 90th percentile or less, they are circled in green color.

A data quality trust check detects that the systolic blood pressure taken on 21 Nov 2020 is of low data quality trust. Font color for the observation is changed to red and a warning message is displayed to advise the practitioner to proceed with caution and review detail measurement history.

Filtered View Screen

Should the practitioner wish to process additional historical information systematically beyond the three observations quick view, an existing screen was modified for a *Detail History View* of measurements with a data quality trust filter (Fig. 5.14)). This filter allows the practitioner to filter historical observations based on data quality trust levels of $High$, $Medium$, and/or Low. This functionality may be important for a clinical decision where only high trust observations should be used. It can also be used to identify low trust measurements that may be the result of a low trust measurement method, a low trust healthcare practitioner or the result of a data quality probe that identified a potential data quality issue.

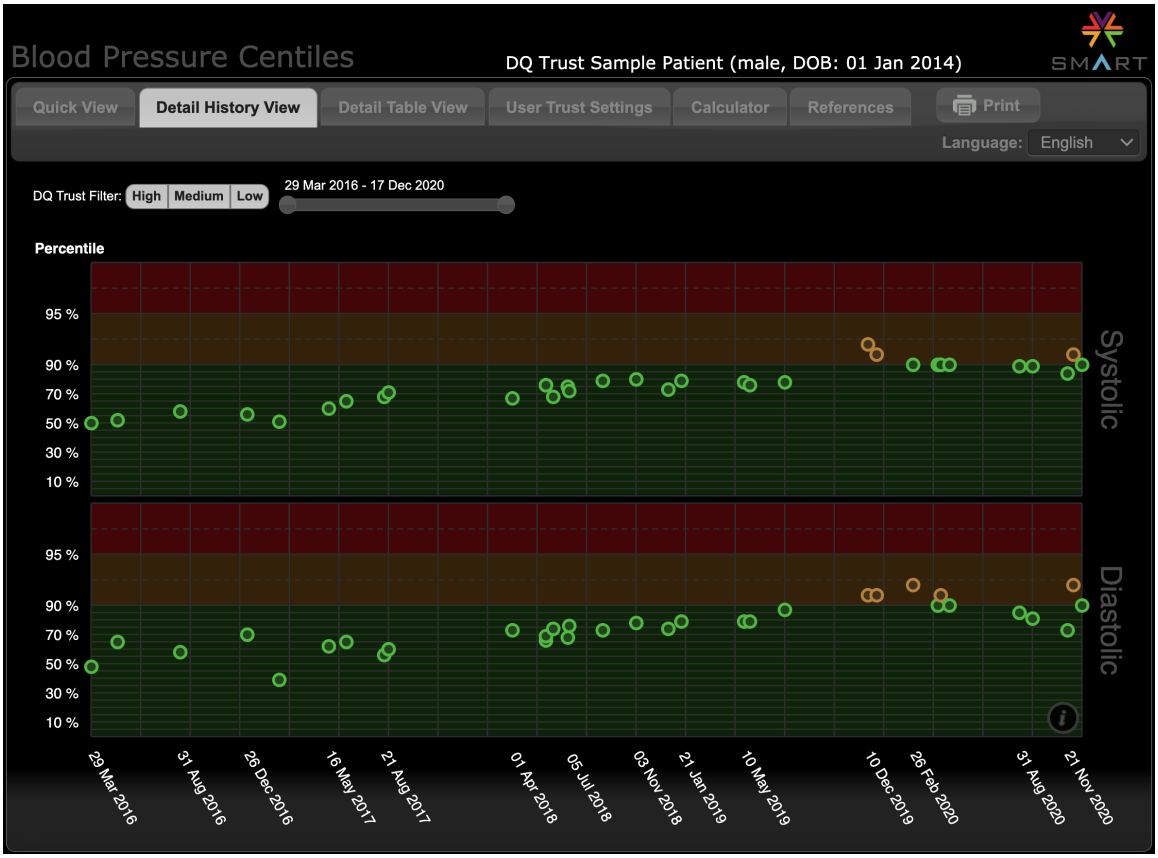


Figure 5.14: Prototype Screen: Detail History View with Modified Filters

Screen Example (Figure 5.14): The practitioner reviews historical blood pressure measurements for the patient using the *Detail History View*. She uses the filter buttons to activate a filter to hide blood pressure measurements with low data quality trust. She finds that the three most recent measurement pairs do not appear to be outliers. In fact, a measurement pair from 23 Sep 2020 with high data quality trust confirms her diagnosis.

Data Quality Trust Detail

A practitioner in systematic processing mode may also want to analyze details of how data quality trust of a blood pressure measurement was inferred from agent trust and reliability trust of the measurement method. Hovering over a data point on both *Detail History View* and *Quick View* opens a pop-up screen displaying meta-data for the patient observation. The existing popup screen was augmented with data quality trust details on practitioner, measurement method, verification status, certification status, and resulting data quality trust (Figure 5.15).

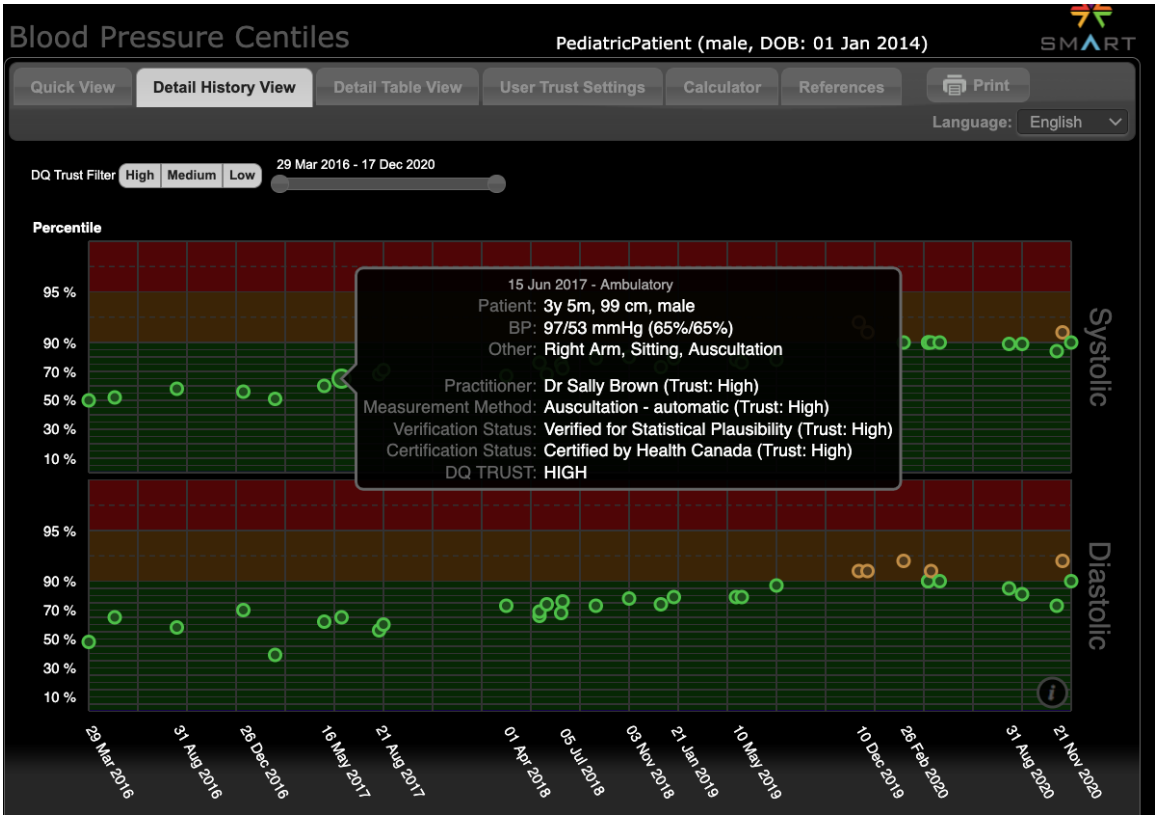
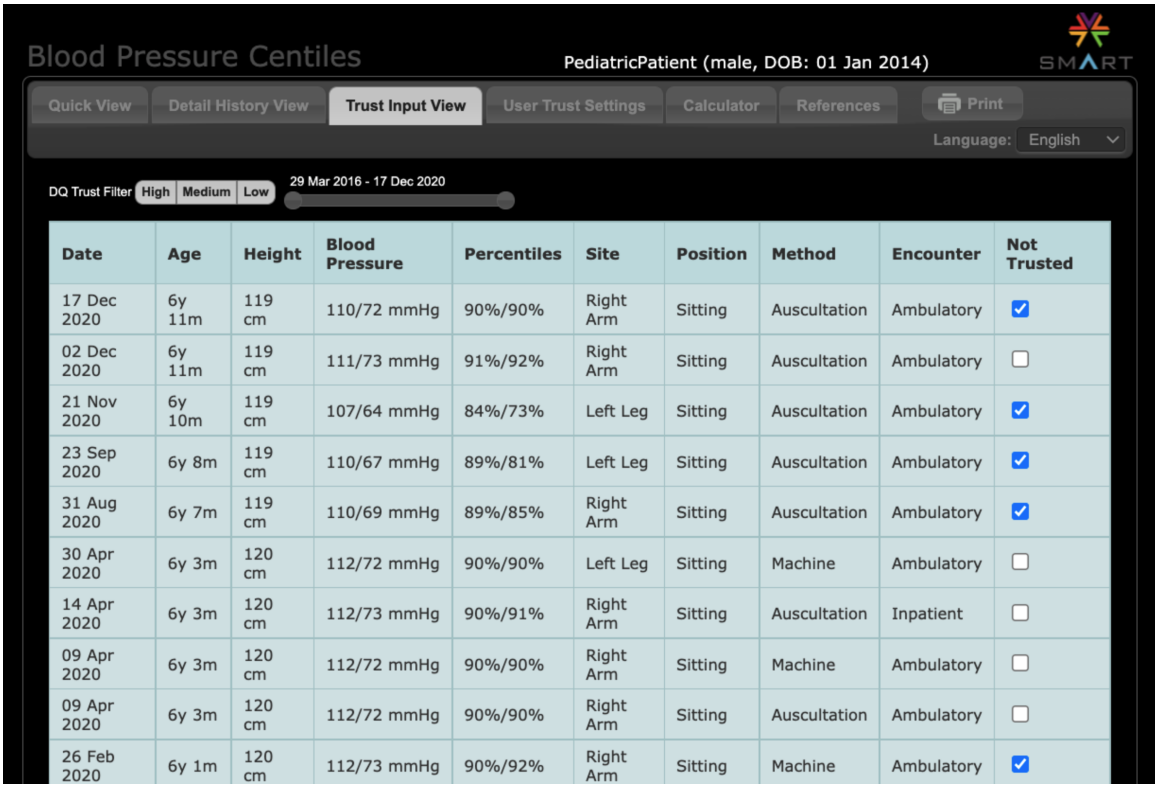


Figure 5.15: Prototype Screen: Data Quality Trust Meta-Data

Trust Input View

After reviewing data quality trust details for one or more blood pressure measurements, the practitioner may want to mark a measurement as not trusted so it is not displayed as part of the *Quick View* and excluded from diagnosis going forward. The existing *Detail Table View* screen was modified to include a checkbox input field for each blood pressure measurement listed (Figure 5.16). Placing or removing a check mark in this input field results the measurement to be marked as "Not Trusted" by the practitioner.



Date	Age	Height	Blood Pressure	Percentiles	Site	Position	Method	Encounter	Not Trusted
17 Dec 2020	6y 11m	119 cm	110/72 mmHg	90%/90%	Right Arm	Sitting	Auscultation	Ambulatory	☑
02 Dec 2020	6y 11m	119 cm	111/73 mmHg	91%/92%	Right Arm	Sitting	Auscultation	Ambulatory	☐
21 Nov 2020	6y 10m	119 cm	107/64 mmHg	84%/73%	Left Leg	Sitting	Auscultation	Ambulatory	☑
23 Sep 2020	6y 8m	119 cm	110/67 mmHg	89%/81%	Left Leg	Sitting	Auscultation	Ambulatory	☑
31 Aug 2020	6y 7m	119 cm	110/69 mmHg	89%/85%	Right Arm	Sitting	Auscultation	Ambulatory	☑
30 Apr 2020	6y 3m	120 cm	112/72 mmHg	90%/90%	Left Leg	Sitting	Machine	Ambulatory	☐
14 Apr 2020	6y 3m	120 cm	112/73 mmHg	90%/91%	Right Arm	Sitting	Auscultation	Inpatient	☐
09 Apr 2020	6y 3m	120 cm	112/72 mmHg	90%/90%	Right Arm	Sitting	Machine	Ambulatory	☐
09 Apr 2020	6y 3m	120 cm	112/72 mmHg	90%/90%	Right Arm	Sitting	Auscultation	Ambulatory	☐
26 Feb 2020	6y 1m	120 cm	112/73 mmHg	90%/92%	Right Arm	Sitting	Machine	Ambulatory	☑

Figure 5.16: Prototype Screen: Trust Input View

Trust User Settings

Since data quality trust relies on user-specific settings to assign trust in the app, a new *User Trust Settings* screen (Fig. 5.17) was added to the user interface. Healthcare practitioners with recorded blood pressure measurements in the EMR are listed with a relative degree of trust, as represented by a numerical value between 0 and 100. Observation methods are assigned a relative degree of trust, as represented by a numerical value between 0 and 100. Trust preferences for data quality probes can be set by activating or deactivating individual data quality probes.

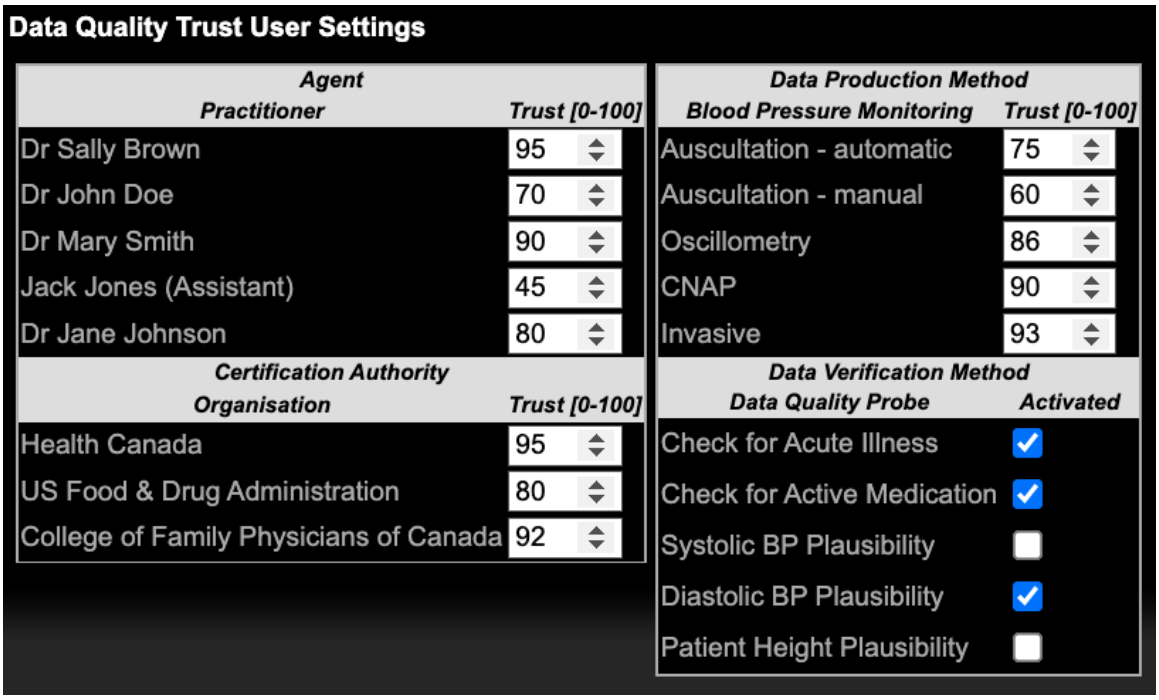


Figure 5.17: Prototype Screen: Data Quality Trust User Settings

Trust in certification authorities can be set as part of a certification trust user settings screen. For example, the user can assign a particular trust level to *Health Canada* which is reflected in the trust of any measurement device or practitioner certified by this certification authority. A certification trust settings screen is shown

in Figure 5.18.

Figure 5.18: Prototype Screen: Certification Trust Settings

5.2.4 Prototype Validation with Scenario Evaluation

A user study of acceptance and use of the prototype by clinical users was beyond the scope of our research project. The goal of our validation research with a prototype was to predict how the prototype artifact would interact with the clinical example context, without actually observing the implemented artifact in a real-world context. We developed six usage scenarios presented by a model of the context to demonstrate how the prototype would respond.

Development and evaluation of these usage scenarios was done in two iterations. The initial iteration proposed and evaluated usage scenarios that seemed plausible and realistic to us from a software engineering perspective. We sought feedback for the scenarios from a clinical domain expert who diagnoses pediatric hypertension in clinical practice. The feedback provided that our proposed scenarios required changes to reflect clinical practice. We modified the scenarios to align with the clinical feedback received. Table 5.4 shows the six scenarios with variations in availability and provenance of clinical data for use during clinical decision making.

Table 5.4: Clinical Usage Scenarios for Data Quality Trust

Scenario	**Description**
1	All data available, recorded by practitioner
2	All data available, not recorded by practitioner, certified device
3	All data available, not recorded by practitioner, not certified device
4	All data available, not recorded by practitioner, not certified device, social network of trust
5	Confounding Clinical Presentation Impacting BP Measurement
6	Incomplete Data or Known Data Quality Issue

A description of how the prototype is predicted to respond to the six usage scenarios follows. The data user and practitioner is a family physician FamilyDoc. Blood pressure measurements from prior patient encounters are available in her EMR system. FamilyDoc has access to other HISs that may store additional clinical information about her patients.

FamilyDoc wants to make a diagnosis of hypertension for a pediatric patient based on clinical practice guidelines [102]. Patient data for age, sex and height are available in the EMR. Based on patient age, sex and height the BP Centiles app calculates the corresponding percentile value for diastolic and systolic blood pressure measurements. Placement of these percentiles in a range determines a diagnosis of normal blood pressure, hypertension, pre-hypertension, or hypotension. If FamilyDoc believes that the three percentile pairs support a diagnosis, she records the diagnosis in the EMR. For her diagnosis, FamilyDoc needs to know the data quality of the three most recent blood pressure measurements.

Scenario 1

The first scenario describes a diagnosis with no data quality concerns. The three most recent blood pressure measurements were recorded by FamilyDoc. She trusts her own measurements for the three BP Centiles pairs and makes her diagnosis accordingly. The prototype shows high data quality trust for the blood pressure centile pairs.

Scenario 2

The second scenario describes a diagnosis where the blood pressure measurements were made by a different practitioner. The measurement device used is certified by a certification authority that FamilyDoc trusts. The prototype shows high data quality trust for the blood pressure centile pairs.

Scenario 3

The third scenario describes a diagnosis where the blood pressure measurements were made by a different practitioner. The measurement device used was not captured in the EMR. FamilyDoc does not know the other practitioner. The prototype shows low data quality trust for the blood pressure centile pairs.

Scenario 4

The fourth scenario describes a diagnosis where the blood pressure measurements were made by a different practitioner. The measurement device used was not captured in the EMR. However, FamilyDoc knows the other practitioner. The level of trust assigned to the other practitioner determines what data quality trust for the blood pressure centile pairs is returned by the prototype.

Scenario 5

The fifth scenario describes a diagnosis where the patient had confounding clinical presentation impacting blood measurement at a prior visit. This could be due to an acute illness or active medication taken by the patient. The prototype detects from other data stored in the EMR that a confounding clinical presentation existed. It shows low data quality trust for one or more of the blood pressure centile pairs.

Scenario 6

The sixth scenario describes a diagnosis where FamilyDoc knows that a blood pressure measurement has a data quality issue. This scenario also applies if required measurement data is incomplete. FamilyDoc makes a new blood pressure measurement that she trusts. She marks the measurement with the known data quality issue as "Not trusted" in the prototype and proceeds with her diagnosis.

5.3 Clinical Workflow and Computation Example

It can be difficult to follow the computations required for data quality trust based on app screen examples alone. We provide a detailed data example to illustrate the workflow and computations performed to serve data quality trust information to the extended app screens. The following sections instantiate the hypertension diagnosis provenance tree for user trust settings, and guide through the calculations that the prototype needs to perform. Example data and computation results were also used to illustrate user interface screen modifications for the visualization of data quality trust earlier.

5.3.1 Trust Settings and Simulation Setup

Since the trust computations rely on preset user trust preferences, individual trust settings need to be defined first for the clinical data user.

Measurement Method Trust Assignments

Table 5.5 provides an example of user trust assignments to several blood pressure and height measurement methods used by the app. System copy and automated system calculations are assumed to be highly trusted (trust value of 100).

Table 5.5: Computation Example: Method Trust Assignments

BP Measurement Method	**Trust [0-100]**	**Trust Label**
BP Auscultation (Auto)	97	High
BP Auscultation (Manual)	91	High
BP Oscillometry	95	High
BP CNAP	95	High
BP Invasive	99	High
Height Measurement (Manual)	85	High
Hypertension Diagnosis	95	High
System Copy	100	High
System Calculation	100	High

Agent Trust Assignments

Table 5.6 provides an example of user trust assignments to several practitioners that have performed blood pressure measurements for the patient. FamilyDoc's trust in her own capabilities is set at 100 (high).

Table 5.6: Computation Example: Agent Trust Assignments

Practitioner	**Trust [0-100]**	**Trust Label**
FamilyDoc (Clinical User)	100	High
Dr Sally Brown	95	High
Dr John Doe	90	High
Dr Mary Smith	92	High
Jack Jones (Assistant)	35	Medium
Dr Jane Johnson	93	High

Data Quality Probe Trust Assignments

Table 5.7 provides an example of user trust assignments to several data quality probes used for verification. Trust assignment is binary allowing a user to activate or deactivate individual data quality probes in the app.

Table 5.7: Computation Example: Data Quality Probe Trust Assignments

DQ Probe	**DQ Probe Type**	**Trust [Yes,No]**
Check for Acute Illness	Internal Concordance	Yes
Check for Active Medication	Internal Concordance	Yes
Systolic BP Plausible	Statistical Plausibility	No
Diastolic BP Plausible	Statistical Plausibility	Yes
Patient Height Plausible	Statistical Plausibility	Yes

Certification Authority Trust Assignments

Table 5.8 provides an example of user trust assignments to several organisations that can certify practitioners or data production methods through their measurement

devices.

Table 5.8: Computation Example: Certification Trust Assignments

Organisation	**Trust [0-100]**	**Trust Label**
Health Canada	99	High
US Food and Drug Admin.	93	High
College of Family Physicians of Canada	97	High

Instantiated Provenance Tree

The app computes data quality trust from the hypertension diagnosis provenance tree for the patient. Tables 5.9 and 5.10 instantiate the provenance tree with trust settings for provenance tree data elements with and without antecedents.

Table 5.9: Computation Example: Captured Data without Antecedents

Id	**Data Item**	**Agent (Trust)**	**Method (Trust)**
t_1	Diastolic BP	Dr Brown (95)	Auscultation - Automatic (97)
t_2	Age	Dr Brown (95)	System Copy (100)
t_3	Sex	Dr Brown (95)	System Copy (100)
t_4	Height	Assistant Jones (35)	Height - Manual (85)
t_5	Systolic BP	Dr Brown (95)	Auscultation - Automatic (97)
t_6	Diastolic BP	Dr Smith (92)	Auscultation - Manual (91)
t_7	Age	Dr Smith (92)	System Copy (100)
t_8	Sex	Dr Smith (92)	System Copy (100)
t_9	Height	Dr Smith (92)	Manual Measurement (85)
t_{10}	Systolic BP	Dr Smith (92)	Auscultation - Manual (91)
t_{11}	Diastolic BP	Dr Johnson (93)	Auscultation - Manual (91)
t_{12}	Age	Dr Johnson (93)	System Copy (100)
t_{13}	Sex	Dr Johnson (93)	System Copy (100)
t_{14}	Height	Dr Johnson (93)	Manual Measurement (85)
t_{15}	Systolic BP	Dr Johnson (93)	Auscultation - Manual (91)

Table 5.10: Computation Example: Derived Data with Antecedents

Id	**Data Item**	**Agent (Trust)**	**Method (Trust)**
t_{16}	Diastolic BP Percentile 1	FamilyDoc (100)	System Calc.(100)
t_{17}	Systolic BP Percentile 1	FamilyDoc (100)	System Calc.(100)
t_{18}	Diastolic BP Percentile 2	FamilyDoc (100)	System Calc.(100)
t_{19}	Systolic BP Percentile 2	FamilyDoc (100)	System Calc.(100)
t_{20}	Diastolic BP Percentile 3	FamilyDoc (100)	System Calc.(100)
t_{21}	Systolic BP Percentile 3	FamilyDoc (100)	System Calc.(100)
t_{22}	Hypertension Diagnosis	FamilyDoc (100)	Hypert. Diagnosis (95)

5.3.2 Basic Computation of Data Quality Trust

Data quality trust is computed with a recursive trust inference (Equation 3.2) applied to the hypertension diagnosis provenance tree (Figure 5.7). Data quality trust for t_{1-15} data items without antecedents (Table 5.9) must be computed first before the resulting antecedent data quality trust can be used for computing trust of t_{16-22} data items with antecedents (Table 5.10). If data quality trust of any antecedent data items is lower than the trust computed for the derived data item, then the lower trust applies. Table 5.11 and Table 5.12 show the data quality trust results for each data item.

Table 5.11: Computation Example: Trust for Data without Antecedents

Id	**Data Item**	**Association Rules**	**DQ Trust**	**DQ Trust Label**
t_1	Diastolic BP	Fully-Automated	77.4	High
t_2	Age	Fully-Automated	83.5	High
t_3	Sex	Fully-Automated	83.5	High
t_4	Height	Manual	47.7	Medium
t_5	Systolic BP	Fully-Automated	77.4	High
t_6	Diastolic BP	Semi-Automated	68.8	Medium
t_7	Age	Fully-Automated	83.3	High
t_8	Sex	Fully-Automated	83.3	High
t_9	Height	Manual	68.7	Medium
t_{10}	Systolic BP	Semi-Automated	68.8	Medium
t_{11}	Diastolic BP	Semi-Automated	68.8	Medium
t_{12}	Age	Fully-Automated	83.3	High
t_{13}	Sex	Fully-Automated	83.3	High
t_{14}	Height	Manual	69.9	Medium
t_{15}	Systolic BP	Semi-Automated	68.8	Medium

Table 5.12: Computation Example: Trust for Data with Antecedents

Id	**Data Item**	**Association Rules**	**DQ Trust**	**DQ Trust Label**
t_{16}	Diastolic BP Percentile 1	Fully-Automated	47.7	Medium
t_{17}	Systolic BP Percentile 1	Fully-Automated	47.7	Medium
t_{18}	Diastolic BP Percentile 2	Fully-Automated	68.7	Medium
t_{19}	Systolic BP Percentile 2	Fully-Automated	68.7	Medium
t_{20}	Diastolic BP Percentile 3	Fully-Automated	68.8	Medium
t_{21}	Systolic BP Percentile 3	Fully-Automated	68.8	Medium
t_{22}	Hypertension Diagnosis	Manual	47.7	Medium

5.3.3 Computation with Verification

Based on the modified provenance tree with verification (Figure 5.11), we can calculate how data quality probes affect the recursive trust calculations. For this calculation example, we assume that data quality probe v_5 detected that the patient was taking prescription medication containing the active ingredient *pseudoephedrine*. This medication may have affected the systolic blood pressure measurement resulting in t_5. Table 5.13 shows the recalculated data quality trust based on a the potential data quality issue identified with the blood pressure measurement. Trust in derived data items t_{17} and t_{22} is now also set to 0 with trust label (**low**).

Table 5.13: Computation Example: Trust with Data Quality Probe Issue

Ids	**Data Item**	**DQ Probe**	**DQ Trust**	**DQ Trust Label**
t_1	Diastolic BP	v_1	77.4	High
t_2	Age	-	83.5	High
t_3	Sex	-	83.5	High
t_4	Height	v_2	47.7	Medium
t_5	Systolic BP	v_3, v_4, v_5	**0**	**Low**
t_{16}	Diastolic BP Percentile 1	-	47.7	Medium
t_{17}	Systolic BP Percentile 1	-	**0**	**Low**
t_{22}	Hypertension Diagnosis	-	**0**	**Low**

5.3.4 Computation with Certification

The effect of practitioner certification on data quality trust computation can be shown with an example. In practice, Health Canada does not certify practitioners for blood pressure measurements but we have constructed this example to illustrate our new method. In trust assignment Tables 5.6 and 5.8, assistant Jack Jones received a trust value of 35 (Medium) and Health Canada received a trust value of 99 (High). If Jack Jones is certified by Health Canada, he receives his transitive data quality trust property from Health Canada. His trust assignment from certification is now 99 (high), superseding his previous direct trust assignment of 35. Since the height measurement t_4 is an antecedent data item to t_{16}, t_{17} and t_{22}, data quality trust for these data items must be recalculated. The result of this recalculation is shown in Table 5.14.

Table 5.14: Computation Example: Trust with Certified Practitioner

Id	**Data Item**	**Association Rules**	**DQ Trust**	**DQ Trust Label**
t_4	Height	Manual	80.1	High
t_{16}	Diastolic BP Percentile 1	Fully-Automated	77.4	High
t_{17}	Systolic BP Percentile 1	Fully-Automated	77.4	High
t_{22}	Hypertension Diagnosis	Manual	68.7	Medium

5.4 Complex Clinical Examples

Pediatric hypertension provided a simple and relatable real-world example to demonstrate feasibility for the new data quality assessment method with a prototype. While this condition is arguably part of a growing problem for pediatric patients, other clinical examples and prototypes with more complex applications of data quality trust may show additional value to individual patient care and healthcare systems. It is relatively inexpensive to perform an additional blood pressure measurement if there are concerns with the data quality of an earlier blood pressure observation. Complex clinical examples that rely on the data quality of several patient observations combined with existing diagnoses of underlying conditions may be more difficult to replicate if there are concerns with underlying data quality.

In addition to the retrospective data quality assessment of existing data, data quality trust could be used with predictive or prescriptive analytics functionality. From an initial assessment of a potential condition with low diagnostic confidence, to further verification activities that increase or reduce the confidence in the presence of the condition, to conclusive testing that may provide very high confidence in a diagnosis, the data quality trust model supports data transformations required to

compute provenance-based trust and to guide diagnostic workflows.

The following sections describe ideas for how data quality trust may be used with complex clinical examples. These ideas represent early work that is not part of the current prototype. Future work may investigate how data quality trust can be integrated with clinical concepts of diagnostic accuracy and clinical decision support systems.

5.4.1 Decision Tree for Food Allergy Diagnosis

The decision tree for a food allergy diagnosis as developed by Mailhol et al. [103] (Figure 5.19) uses results from skin prick tests (SPT), food atopy patch tests (APT), food specific s-IgE tests, and oral food challenge (OFC) tests. In their study, the authors defined a clinically relevant history as hives (urticaria), angioedema, erythema, rhinitis, bronchial obstruction, vomiting or anaphylaxis appearing within 4 hours after ingestion of an identified single food item.

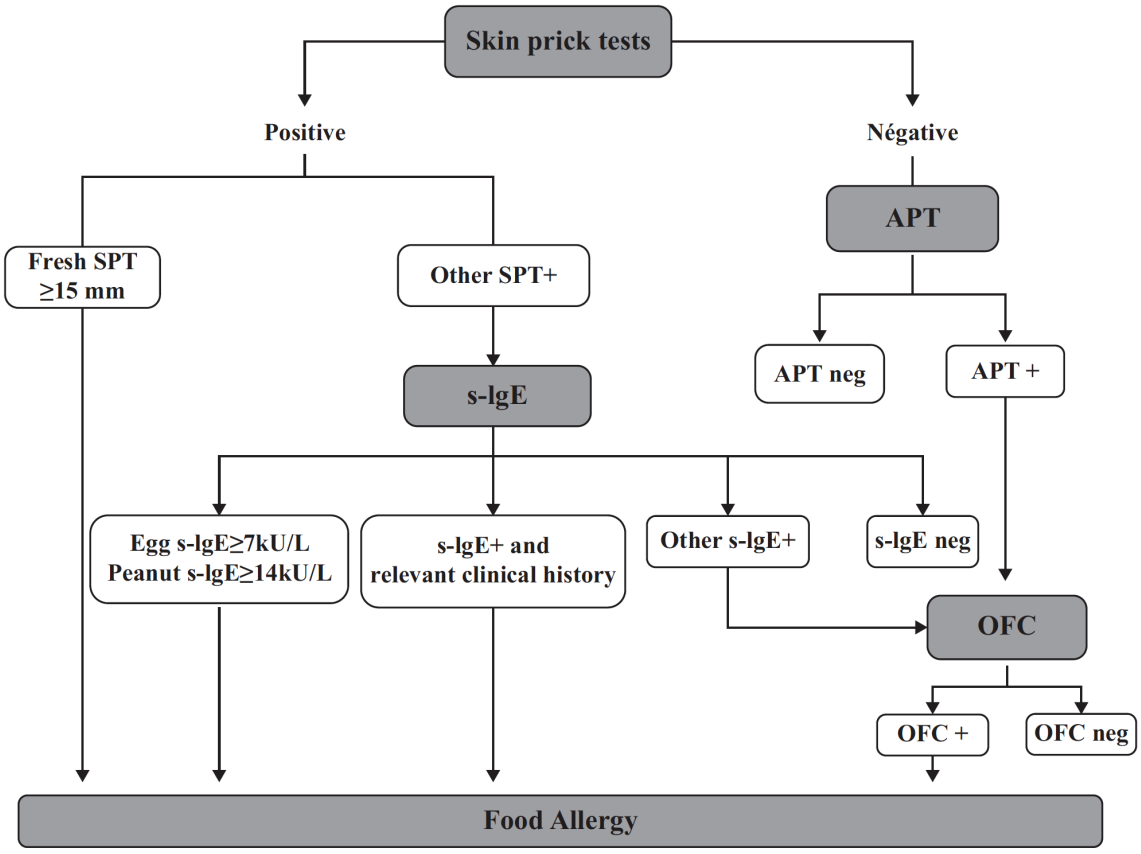


Figure 5.19: Decision Tree for Food Allergy Diagnosis (Mailhol, 2014)

5.4.2 Provenance Tree for Food Allergy Diagnosis

The provenance tree for a food allergy diagnosis reflects a single path taken through the diagnosis decision tree that was used to make a food allergy diagnosis. It combines documented occurrences of food intolerance, patient observations, and diagnosed conditions in the patient history to a food allergy diagnosis data entity. Compared to the pediatric hypertension diagnosis provenance tree which primarily relied on patient observations to compute data quality trust, this more complex provenance tree also considers data quality trust of prior diagnoses or clinical assessments for an allergy

or intolerance.

Figure 5.20 shows a trust provenance tree for a food allergy diagnosis data item using our model and possible antecedents to the food allergy diagnosis. Transformations t_{1-4} describe patient observations from skin prick test, atopy patch test, s-IgE blood test and oral food challenge. These observations were generated by agents and data production methods without antecedents. Transformation t_k describes one or more relevant prior adverse reaction caused by an allergy or intolerance in the patient's clinical history. Transformation t_n describes one or more relevant diagnoses of conditions in the patient's clinical history. These diagnoses were generated by agents and data production methods with antecedent patient observations generated by transformations t_m. Transformation t_5 describes the diagnosis of a food allergy that may have been based on skin prick test, atopy patch test, s-IgE blood test, oral food challenge, and/or the relevant clinical history for the patient with other clinical assessments and diagnoses. In our model, diagnosis of the food allergy is performed by an agent and manual data production method. Based on the different possible paths taken through the decision tree in Figure 5.19, the data quality trust provenance tree of a particular food allergy diagnosis may not contain all transformations listed. For example, a fresh skin prick test with positive result and a patch size of $\geq$ 15 mm alone may allow for a food allergy diagnosis. In this case, the provenance tree for that particular diagnosis would only be made up of transformations t_1 and t_5.

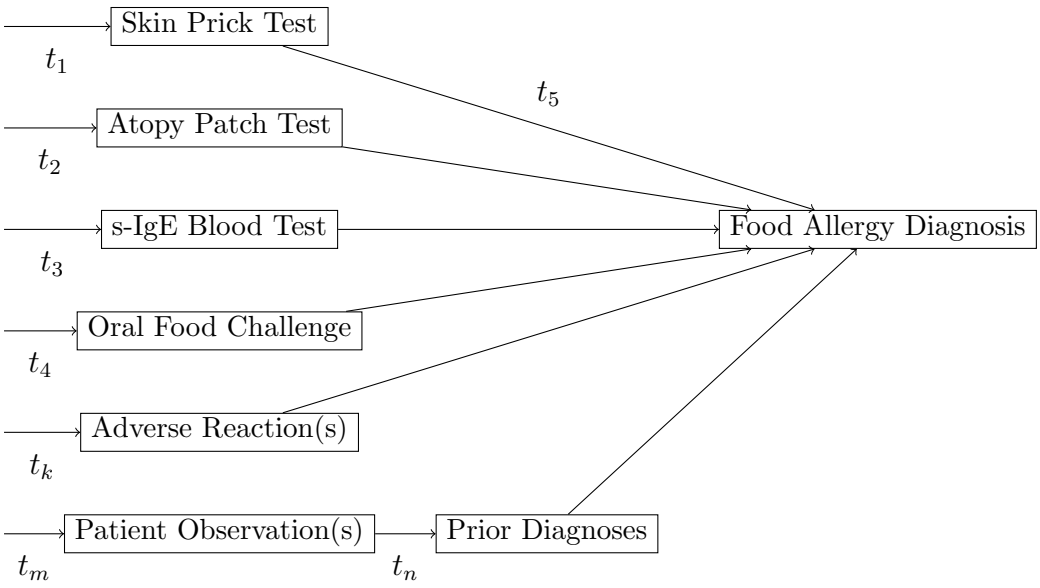


Figure 5.20: Trust Provenance Tree for Food Allergy Diagnosis

5.4.3 Allergies with Predictive Data Quality Trust

The previous section assumes that a diagnosis for a food allergy can be made in retrospective based on patient observations, recorded adverse reactions, and prior diagnoses. However, in practice it is more common that assessment of even a suspected intolerance is recorded in a patient's allergy list which informs a clinician of potential risks of an allergic reaction to a substance. The allergy list record for an intolerance to a specified substance may guide decisions by the clinician, for example by avoiding certain substances when prescribing medication. Such an allergy assessment can then be updated in subsequent patient encounters as more information about the possible intolerance becomes available. A clinical decision support system with data quality trust may predict the presence of an allergy or intolerance, using the clinician's relative trust preferences. A system with prescriptive functionality may go further by advising a clinician on what types of testing could improve their diagnostic confidence in the presence of an allergy based on data quality trust computations of the agents

and methods used to generate existing diagnostic data and allergy assessments.

Practical Example with Food Allergies

To illustrate the potential for implementing data quality trust with a simple example for predictive functionality, Figure 5.21 shows the provenance tree for an initial assessment of a suspected food allergy that was then verified with additional test results in subsequent patient encounters and assessments. It reflects a single path taken through the decision tree (Figure 5.19) and four versions of an allergy assessment record, as updated during each patient encounter. Data quality trust is computed and assigned to trust labels *low*, *medium*, *high*, and *very high* that represent a clinician's diagnostic confidence in the presence of the suspected food allergy.

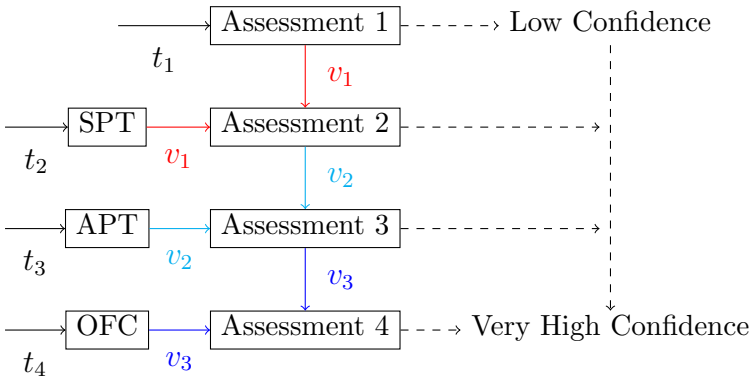


Figure 5.21: Verification Example for Confidence in Food Allergy

In this example, the patient may have reported a rash after consuming a food (t_1). Based on the decision tree for a food allergy diagnosis (Figure 5.19), a clinician would go through the tree until their confidence in the presence of an allergy was sufficiently high to make a diagnosis. During each subsequent encounter, the clinician would use the prior assessment and the new test results to evaluate the presence of an allergy. Data quality trust is computed for each allergy assessment, combined with results of different tests (t_{2-4}) to verify allergy assessments from the prior encounter. These

verification activities v_{1-3} then either increase or decrease the clinician's diagnostic confidence in the presence of a food allergy.

Medication Allergies in Primary Care

A more common application of allergy assessments in primary care is the diagnosis of medication allergies. Medication allergies may pose serious safety risks to patients that are prescribed medication containing certain substances, such as penicillin [104]. Medication allergies are often suspected but rarely tested for and diagnosed conclusively due to relatively expensive testing and limited availability of tests and allergists for referral. Patients experiencing a mild allergic reaction after taking medication may attribute the reaction to the medication when other, non medication-related, confounding issues may be the true cause. Subsequent patient observations of an allergic reaction recorded in an EMR may be the result of the patient retelling the same story of an allergic reaction in multiple clinical settings. Data quality trust allows for assessing trustworthiness of patient observations based on their provenance, which may provide a clinician sufficiently high diagnostic confidence without the need for expensive testing.

Diagnosis of Penicillin Allergy

Diagnosing a penicillin allergy requires a physical examination and diagnostic testing to establish confidence in that the allergy is present for a patient. There are different types of penicillin with differing chemical properties causing allergic reactions. A positive skin test from patient exposure to a small dose of the suspect penicillin may determine that the patient is allergic to penicillin. However, a negative skin test may not be conclusive, and it may require additional diagnostic testing for drug reactions that cannot be detected by skin tests. Graded drug challenges with up

to five increasing doses of the suspect penicillin allow a clinician to increase their diagnostic confidence in the presence of a penicillin allergy. If the therapeutic dose is reached with no reaction, the clinician may conclude that the patient is not allergic to that type of penicillin.

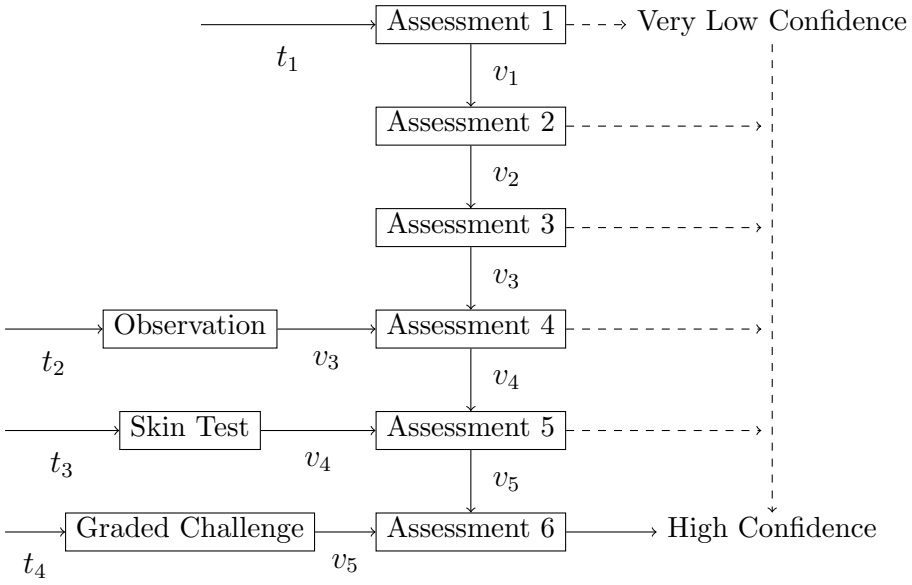


Figure 5.22: Verification Example for Confidence in Penicillin Allergy

Practical Example with Penicillin Allergy

Figure 5.22 shows an illustration for how verification activities v_{1-5} increase or decrease the diagnostic confidence in the presence of a penicillin allergy. The initial allergy Assessment 1 of a penicillin allergy (t_1) may have been recorded by a family physician years earlier based on anecdotal evidence of a skin rash, as reported by the patient after taking medication containing penicillin. Data quality trust and diagnostic confidence are very low, since provenance is unclear. Assessment 2 for the penicillin allergy occurs as part of the initial encounter with a new family physician where allergy assessments are routine. Despite the new assessment, there is no new information about factors that could have changed the diagnostic confidence so the

confidence remains very low. Assessment 3 happens during admission to the emergency room for an unrelated issue. Patient retells the story of an exposure and allergic reaction during childhood which the admitting nurse records in the EMR. Since no new evidence about the allergy is presented, diagnostic confidence remains very low.

Recently, our patient required antibiotic medication. Since a suspected penicillin allergy was present in the EMR, the attending physician prescribed *Cephalexin* instead of penicillin. *Cephalexin* is similar to penicillin in action and some side effects. *Cephalexin* carries an increased risk of allergic reaction among patients with penicillin allergy [105]. A day after taking this medication, patient presented with a rash which the clinician recorded as a new patient observation (t_2). Verification activity v_3 combines Assessment 3 and the new patient observation to Assessment 4 with increased diagnostic confidence that a penicillin allergy is present. To confirm a penicillin allergy, clinician orders a skin test that helps verify Allergy Assessment 4 and suspected allergy (v_4). The skin test is negative and inconclusive skin test results generate a new Allergy Assessment 5 with decreased diagnostic confidence that the penicillin allergy is present. A further verification activity (v_5) of Allergy Assessment 5 and a graded challenge with an observed allergic reaction (t_4) generate Allergy Assessment 6 with high diagnostic confidence that a penicillin allergy is present. Our practical example is summarized in Table 5.15.

Table 5.15: Complex Clinical Example: Diagnosis of Penicillin Allergy

Id	**Data Item**	**Validation of Suspected Allergy**	**Confidence**
t_1	Assessment 1	Suspected penicillin allergy in patient history based on patient report	Very Low
v_1	Assessment 2	Change of family physician in patient history, verbal confirmation of suspected allergy only	Unchanged
v_2	Assessment 3	Emergency Room admission, verbal confirmation of suspected allergy only	Unchanged
t_2	Observation	Clinician observes rash after prescribing Cephalexin	
v_3	Assessment 4	Due to possible cross-reactivity of Cephalexin with penicillin allergy, additional testing is required	Increased to Medium
t_3	Skin Test	Penicillin skin test is negative	
v_4	Assessment 5	Inconclusive skin test rules out some forms of penicillin allergy, additional testing is required	Lowered, still Medium
t_4	Graded Challenge	Penicillin graded challenge is positive	
v_5	Assessment 6	Positive graded challenge allows for a conclusive diagnosis, no further testing is required	High

5.4.4 FHIR Extensions for Data Quality Trust with Allergies

Platform interoperability features for data quality trust proposed in Chapter 4 allow for the transmission of trust in patient observations. Assignment of data quality trust to prior diagnoses of conditions or clinical assessments of allergies and intolerances

requires the extension of two additional FHIR resources to transmit trust preferences over FHIR interfaces. The existing trust settings stored in a FHIR Practitioner resource extension are sufficient to transmit required agent trust and method reliability trust properties. A FHIR AllergyIntolerance resource [106] stores information about *risk of harmful or undesirable, physiological response which is unique to an individual and associated with exposure to a substance.* This FHIR resource can be extended with a new entity *DQTrust* that specifies the data quality trust for a practitioner data user. Figure 5.16 shows the additional fields required for the extension.

Table 5.16: FHIR Extension: FHIR AllergyIntolerance

Identifier [Cardinality]	**FHIR Extension Type**
AllergyIntolerance.DQTrust [0..*]	BackboneElement
.practitioner [1..1]	Reference (Practitioner)
.dq-trust [1..1]	Decimal

A FHIR Condition resource [107] stores information about *a clinical condition, problem, diagnosis, or other event, situation, issue, or clinical concept that has risen to a level of concern.* This resource for a patient diagnosis can be extended with a new entity *DQTrust* that specifies the data quality trust for a practitioner data user. Figure 5.17 shows the additional fields required for the extension.

Table 5.17: FHIR Extension: FHIR Condition

Identifier [Cardinality]	**FHIR Extension Type**
Condition.DQTrust [0..*]	BackboneElement
.practitioner [1..1]	Reference (Practitioner)
.dq-trust [1..1]	Decimal

5.4.5 Visualization of Data Quality Trust with Allergies

To show how our ideas for data quality trust could be visualized with an allergy diagnosis, we create mockups for several screens that implement the visualization guidance from Chapter 4.

Allergy List with Diagnostic Confidence

The first screen *Allergy List* (Figure 5.23) shows how data quality trust could be integrated with the existing allergy list EMR user interface for a patient. A new column is added to display the diagnostic confidence for the presence of an allergy.

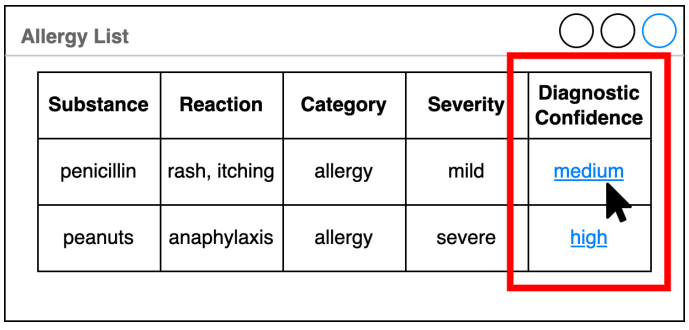


Figure 5.23: Mockup Screen: Allergy List with Diagnostic Confidence

Allergy Diagnostic Confidence

Linked from the allergy list, a *Allergy Diagnostic Confidence* screen (Figure 5.24) shows a summary of patient observations, medication prescription history, and diagnostic tests performed with a computed confidence level in the presence of an allergy, as a relative data quality trust level in our model. For an implementation with prescriptive functionality, the system could provide a clinician recommendations on how to improve the diagnostic confidence. This could be through additional diagnostic testing, or revising the trust preferences for agents and data production methods that generated underlying diagnostic data. Trade-off decisions about whether additional expensive testing is necessary may be considered if the diagnostic confidence is already sufficiently high to make a diagnosis. Recommendations for next steps could be generated based on the path already traversed through a diagnosis decision tree, data quality trust of antecedent data entities, or any remaining branches that could yield an improvement of diagnostic confidence through further verification activities.

In the clinical example shown in Figure 5.24, a patient observation of a mild skin rash, a medication prescription for *Cephalexin* a day before the rash, and a negative skin test result provide the clinician with $Medium$ diagnostic confidence that a penicillin medication allergy is present. Additional verification with a graded challenge for Penicillin is recommended to improve the diagnostic confidence from $Medium$ to an expected new confidence level of $High$ if the test is positive.

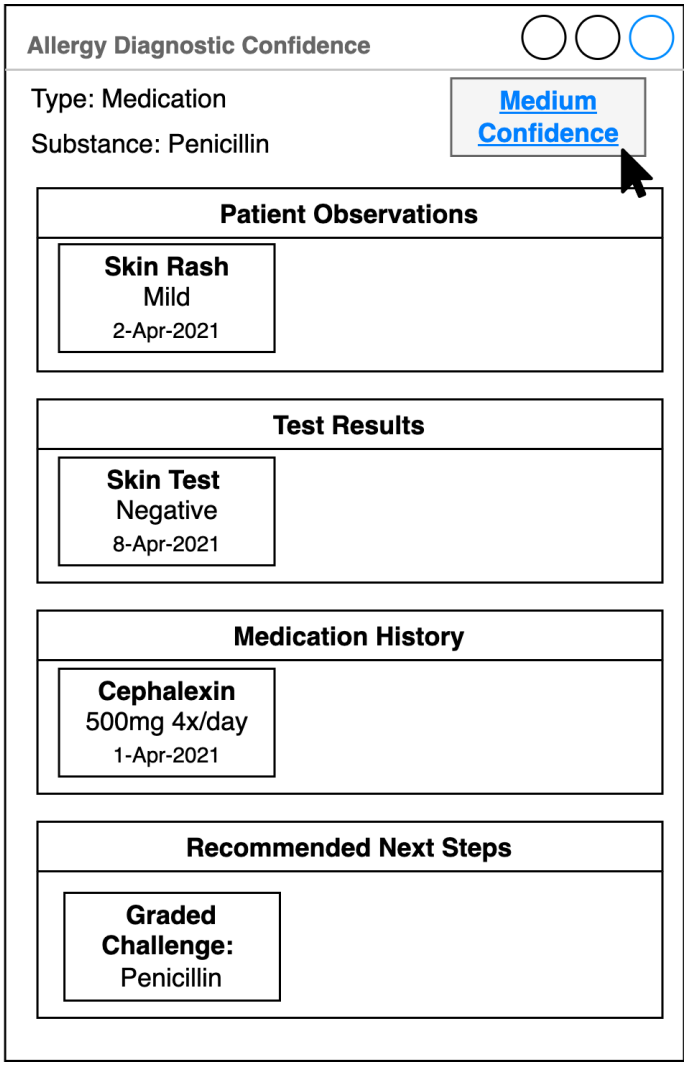


Figure 5.24: Mockup Screen: Allergy Diagnostic Confidence

Allergy Assessment History

Linked from the diagnostic confidence label of the quick view screen, a *Allergy Assessment History* screen in Figure 5.25 shows how contextual data quality trust information allows a clinician to systematically process data quality trust of allergy assessments for an allergy. Historical assessments are placed on a graph based on their diagnostic confidence of *very low*, *low*, *medium*, *high* data quality trust to show how confidence has changed with each new assessment. Hovering over an assessment data

point would show the data item provenance for the assessment including the verifying agent, the verification method, trust assignments made for agent and method, and the data quality trust result. Antecedent data items that served as inputs in the trust computation are listed as well.

In the clinical example shown in Figure 5.25, the initial Assessment A1 is shown with *very low* diagnostic confidence. Confidence does not change for Assessments A2 and A3. Confidence increases for Assessment A4 but decreases somewhat due to a negative skin test result, as reflected in Assessment A5. The placement of Assessment A6 shows the expected increase in diagnostic confidence from verification with a graded challenge.

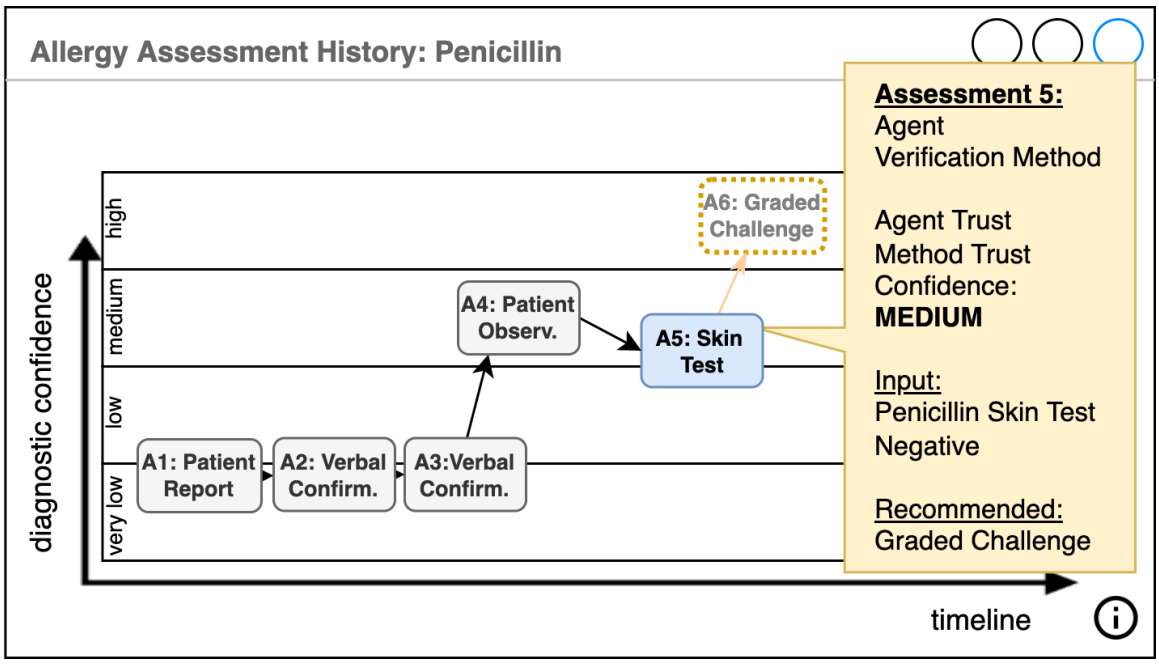


Figure 5.25: Mockup Screen: Allergy Assessment History

5.4.6 User Interface Validation with Domain Expert

To validate proposed user interface screens for a allergy diagnosis clinical example, we sought feedback from a family physician and domain expert. He provided several suggestions for improving aspects of visualization and clinical workflow integration that allowed us to iteratively improve user interface mockup screens, with the result of the final iteration shown in Figures 5.23, 5.24 and 5.25.

First Iteration

In a first iteration, based on this feedback, we made changes to our description of the food allergy clinical example, color coding of warnings, screen layout of allergies, and the display of allergy severity information as part of our allergy diagnosis screen mockups.

Second Iteration

We received feedback from a second iteration that resulted in an additional clinical example of more common and clinically relevant medication allergies, exemplified by a penicillin allergy. We clarified diagnostic workflows and aligned our example descriptions with clinical practice. We changed allergy diagnosis mockup screens to show a penicillin allergy diagnosis.

Third Iteration

A third iteration resulted in additional changes to the allergy diagnosis mockup screens to clarify the labels used for assessments. We refined the clinical concepts and corrected the role of medication for cross-reactivity in our penicillin allergy practical example.

5.4.7 Discussion of Complex Clinical Examples

We reviewed several ideas for how data quality trust could support more complex clinical examples that incorporate a patient's clinical history of diagnoses and different stages of clinical assessments, as exemplified with a food allergy and medication allergy diagnosis. An allergy diagnosis poses an interesting application area for data quality trust since it often relies on documented adverse reactions to substances that are asserted by healthcare practitioners with varying medical backgrounds and qualifications, or the patient. Boundaries between allergy and intolerance are not always well-defined or understood which raises the importance for trustworthiness of agent and data production method for data that forms the basis for an allergy diagnosis. Once a suspected allergy has been added to the EMR, it becomes imperative for clinicians to evaluate this risk through further verification activities, and to take necessary steps that avoid harmful exposure for the patient. The approach shown for extending FHIR AllergyIntolerance and FHIR Condition can be used with other FHIR resources that may be targeted in a data quality trust implementation.

Availability of Provenance Information for Diagnoses

A prerequisite for applying data quality trust to complex clinical examples with interoperable EMR systems is that the clinical domain concepts must have been represented with existing FHIR resources. A practical limitation is that FHIR resources targeted for data quality trust must already store or be linked with provenance information on their data producer, data production method and antecedents used to allow for the computation of data quality trust. In many EMR systems however, provenance with antecedent data entities used for a clinical diagnosis is not recorded in the EMR. This limits the data quality trust application of predictive or prescriptive functionality in complex clinical examples which rely on prior diagnoses in a patient's

clinical history.

5.5 Summary

This chapter introduced pediatric hypertension as a clinical example for data quality trust. We applied a prototype to a slice of life example by extending an existing SMART on FHIR app with data quality trust functionality. The prototype demonstrated how model and method can be integrated with interoperable EMR systems and applied to a clinical example. A detailed data example with the prototype demonstrated the data quality trust calculations in the context of the pediatric hypertension clinical example. We showed how FHIR Interface extensions allow for the transmission of data quality trust between a Trust Preferences Server and the extended BP Centiles app. The existing app user interface was extended to integrate data quality trust visualizations for users in heuristic or systematic cognitive processing modes.

The next chapter will evaluate our research against several validation criteria. It will discuss how threats to construct, internal, and external validity have been addressed. Several research challenges and limitations will be noted.

Chapter 6

Evaluation

This chapter evaluates our research results against validation criteria that were derived from research objectives in our validation framework (Chapter 1). It summarizes the validations undertaken as part of the justify/evaluate step in our iterative validation research process (Figure 1.2). We describe significance and expected impact of our research on patients, healthcare practitioners, and healthcare systems. We consider threats to validity from three distinct aspects: construct, internal, and external validity. We discuss several challenges and limitations of our research.

6.1 Validation

Evaluation of our research results requires validation against research objectives using the validation framework defined in Chapter 1. We derived four main validation criteria as summarized in Table 6.1 and described as follows.

Table 6.1: Summary of Validation Criteria for Research Objectives

Main Validation Criteria	**Validation Methods**	**Reference**
Existing methods cannot meet stated research objectives.	Systematic Literature Search, Scoping Review	Chapter 2
The method provides a contextual data quality assessment result for individual data items.	Simulation, Lab Testing	Chapter 3
The method can be integrated with the data models and user interfaces of interoperable and legacy EMR systems.	Prototyping, Lab Testing	Chapter 4
The method can be applied to a clinical example through integration with clinical workflow and application user interface.	Prototyping, Simulation, Lab Testing, Scenario Evaluation	Chapter 5

6.1.1 Literature Search

Our hypothesis that there are no existing solutions to our research objectives was validated through a systematic literature search which reviewed existing work in the EMR system domain.

6.1.2 Contextual Data Quality Assessment Method

Research objectives related to development of a novel data quality assessment method for individual data items were validated through simulation and lab testing. The assessment method was shown to generate contextual assessment results with labels for *high*, *medium*, *low* data quality. Simulated system response time meets Nielsen usability engineering guidelines [62].

6.1.3 Integration with EMR Systems

Feasibility of assessment method integration with interoperable EMR systems was demonstrated with a design pattern that extends an existing open healthcare interoperability standard with several FHIR resource extensions for data quality trust. Feasibility of integration with non-FHIR legacy EMR systems was demonstrated with several database schema extensions. We developed an archetype to show how OpenEHR-based systems can be extended for data quality trust. We discussed how our solution can be implemented in low network bandwidth environments to allow for a wider range of application settings. Prototyping and lab testing of FHIR resource extensions provided validation that the data model and FHIR interface extensions work in principle. Prototyping and lab testing of data model extensions in three legacy EMR systems identified suitable integration points and validation for the data model extensions.

6.1.4 Application to Clinical Example

Clinical workflow integration and visual user interface integration of the data quality assessment method with a clinical example was demonstrated through a prototype. This prototype consisted of functional descriptions of the workflows used by a clinical data user, and of a visual prototype of a data quality trust user interface. By prototyping extensions to an existing "SMART on FHIR" app with data quality trust, we demonstrated how workflow integration, data model integration and user interface integration can be brought together to apply data quality trust to a specific clinical example context. Simulation of the assessment method, workflow integration and user interface properties with a detailed data example validated feasibility of the novel data quality assessment approach. Expert opinion and usage scenario evaluation of the prototype with different usage scenarios provided further validation.

6.2 Significance of Research Results

Our data quality trust approach was developed to provide clinical data users with a quick way to assess whether the quality of a data item should be trusted for a clinical decision or not. The following aspects contribute to the novelty and significance of our approach:

- The method is fully automated once trust preferences have been set. Existing assessment methods require time-consuming manual interventions that make their use during clinical decision making impractical.

- The method provides a contextual data quality assessment that is relative to the individual clinical data user. Most existing methods assess data quality based on absolute measures that apply for all data users equally.

- The method allows for flexible configuration and simplified assessment result representations with labels that could be configurable separate from the application logic. Many existing methods require computational rules to be set and integrated with the application logic. Changes to definitions then require an application developer to make changes in the source code.

- The method allows for encoding fuzziness or uncertainty in data quality assessments which can be used to support advanced analytics applications. Existing data quality assessment methods are limited to traditional/crisp logic computations.

- The method can be integrated with modern interoperable EMR systems which allow for EMR-agnostic use cases. Existing methods may not support platform interoperability features yet.

6.3 Expected Impact of Research Results

If implemented in real-world projects, we hope that our research will have significant positive impact. We presented and discussed our partial research results at two conference workshops focused on software engineering in healthcare [108, 49]. In the few months that these two papers were made available on the Researchgate platform, they were downloaded and read over 200 times suggesting significant research interest in the topic and our solution. We expect patients, healthcare practitioners, and healthcare systems to benefit from our research.

6.3.1 Positive Impact to Patients

With data quality trust, patients are expected to benefit the most. Poor data quality leads to poor clinical decisions which in turn lead to poor patient outcomes. A patient who receives the correct diagnosis and the best treatment based on this diagnosis will have a better patient outcome than a patient who was misdiagnosed and received the wrong treatment due to poor diagnostic data quality. Ineffective or slow decision making and the inability to identify gaps in care should be reduced. If implemented, we also expect our approach to greatly reduce the number of patient safety incidents attributable to poor data quality. Allowing for automated data quality assessment during clinical decision making allows a healthcare practitioner to focus on the patient improving the patient experience during encounters.

6.3.2 Positive Impact to Healthcare Practitioners

With data quality trust, healthcare practitioners receive a new tool for providing better patient care. Clinical data users can quickly assess whether the quality of a data item can be trusted for a clinical decision or not. Access to interoperable

EMR systems can allow healthcare practitioners to work seamlessly across system contexts. Further development of the app economy for healthcare apps with data quality trust has the potential to give healthcare practitioners access to better app user experiences.

Our EMR user interface and visualization guidance for data quality trust considers the cognitive needs of healthcare practitioners. Professional burnout among physicians is a real problem which raises the importance of designing clinical user interfaces that consider usability and cognitive requirements of clinical data users at the point of care. Alerts that address the clinical data user's personal beliefs and contextual data quality information needs are expected to reduce cognitive load and improve system usability for those users.

6.3.3 Positive Impact to Healthcare Systems

With data quality trust, healthcare system participants are expected to have higher trust in their data for timely clinical decisions. Our approach in itself does not improve the data quality for clinical decision making but it increases awareness of potential data quality issues so those issues can be remediated. For example, if a data quality trust probe identifies a data quality issue and alerts a practitioner, that practitioner is likely to take remedial action and try to identify a root cause. Data quality trust can help improve data quality one data item at a time, leading to overall better data quality over time. Since data integration and interoperability are often seen as significant barriers to data analytics strategic priorities, our integration with the "SMART on FHIR" standard may enable further investment in data integration and interoperability initiatives.

6.4 Threats to Validity

Trustworthiness of the results, and to what extent the results were not biased by the researchers' subjective point of view, are important considerations when evaluating validity of research outcomes [109]. This section addresses several potential threats to the research validity and the extent to which they were mitigated. In this dissertation, we considered three aspects of validity: construct validity, internal validity, and external validity.

6.4.1 Construct Validity

Construct validity measures the degree to which the evaluation methods used can answer the research questions. Threats to construct validity were mitigated by integrating our research with established and previously validated domain constructs where possible. For example, fuzzy inference with the Mamdani method is a well researched construct that has been applied in many domains or problem contexts under different settings. Integration with the "SMART on FHIR" standard to achieve platform interoperability has been the subject of prior research. As part of future work, we recommend further research to validate our constructs with clinical users to ensure that systems based on our constructs can be integrated with clinical practice.

6.4.2 Internal Validity

Internal validity reflects the degree to which a resulting conclusion is justified as well as the selection bias and confounding factors.

Literature Review

The first threat to internal validity relates to our literature review. We relied on several search and inclusion criteria to systematically search the literature for existing solutions to the research problem. A threat to internal validity was potential researcher bias in the literature review as one researcher completed article selection, data extraction, and interpretation. The threat was partially mitigated through review and discussion of suitable search queries with other researchers.

Data Quality Assessment Method

A second threat to internal validity emerged from a lack of formal evaluation of the data quality assessment method with formal methods and mathematical proof. We selected a set of triangular membership functions and assumed certain properties for monotonicity of resulting fuzzy inference functions. To mitigate this threat, we simulated data quality trust inference for many combinations of these triangular and trapezoidal functions in Matlab and visually inspected the graphical result for monotonicity. There are an unlimited number of possible membership function combinations and resulting inference functions for data quality trust.

EMR System Integration

Integration of data quality trust with EMR systems introduces a third threat to internal validity. For the user interface extension, the lack of a user study in a naturalistic clinical setting poses a threat to internal validity. We mitigated this threat to some extent by validating proposed user interface screens with established cognitive engineering principles. Another threat to internal validity is that the FHIR extensions may not work as intended when applied to a real-world setting. We sought to mitigate this threat by performing limited lab testing of FHIR extensions. Data

model integration and schema extensions for legacy EMR systems were applied to a database with relational integrity and demo data.

Prototype

Validation of the prototype with a descriptive scenario evaluation poses a fourth threat to internal validity. The prototype was not evaluated through a user study. We mitigated this threat by seeking feedback on the evaluation scenarios from a domain expert and making revisions accordingly.

6.4.3 External Validity

External validity reflects the degree to which the research results can be generalized to other domains or problem contexts. In this dissertation, there are two main types of generalization possible.

Application to Other Clinical Examples

The first generalization relates to applying our EMR design pattern for data quality trust to other clinical examples. We started with pediatric hypertension as a simple and relatable clinical example where data quality trust was targeted at patient observations in retrospective. Application of the design pattern to additional clinical examples of higher complexity demonstrated feasibility for a broader range of clinical examples, including a forward-looking application with predictive and prescriptive functionality. The threat to external validity was mitigated with allergy clinical examples that relied on the data quality of several patient observations combined with existing diagnoses of underlying conditions.

Application to Other Domains

The second generalization relates to applying data quality trust to other domains. Our data quality trust model and assessment method are not limited to the HIS domain. To facilitate further generalization of our research to other domains, we separated the generic, domain-agnostic representation of model and method in Chapter 3 from the domain-specific integration with the EMR system domain in Chapter 4. As part of future work, further research is required to validate model and method externally with other domains.

6.5 Challenges and Limitations

We have encountered several challenges during our research and identified some limitations to its applicability in clinical environments.

6.5.1 Access to Clinical Environments and Domain Experts

Our research took place during the COVID-19 pandemic which severely impacted health systems around the world. We changed our earlier plans to evaluate aspects of this research in clinical settings since these clinical settings were overwhelmed and not available to us. Access to clinical domain experts was also difficult since the clinical decision making expertise of physicians required for our research was also needed to treat COVID-19 patients. At the time of this writing, the COVID-19 pandemic is still ongoing. A user study was not feasible within the time constraints of our research project.

6.5.2 Fuzzy Inference in Clinical Settings

To our knowledge, this research represents the first application of a fuzzy inference system to estimate data quality trust in a clinical setting. Review by a domain expert and a comparison of results between our data quality trust model and method with expert opinion would help strengthen our claim for validity of model and method. Another limitation lies in the nature of using fuzzy logic for trust inference. There are stringent regulatory requirements for medical devices in many jurisdictions. The inherent uncertainty in computations with fuzzy logic inference may make it difficult to assure and formally prove efficacy of a data quality trust model and inference for a particular medical context. This may limit the flexibility that a data user has in assigning individual trust settings, and for an organisation to define context specific trust membership functions if an app using data quality trust falls under a regulated medical device classification.

6.5.3 Availability of Provenance Information

Other limitations for practical use are that more advanced applications of provenance-based data quality trust require the availability of provenance information. For those EMR systems that have already fully implemented the FHIR interface standard, the FHIR Provenance resource can store provenance meta data for a range of other FHIR objects, including provenance chains of antecedent data items. While not widely used in healthcare settings yet, capture and storage of more comprehensive provenance meta data would allow more clinical use cases to benefit from data quality trust inference from data item provenance.

6.6 Summary

This chapter reviewed our research results against validation criteria that were derived from research objectives. We described significance and expected impact of our research on patients, healthcare practitioners, and healthcare systems. Several threats to validity in construct validity, internal validity, and external validity were reviewed. We also noted some challenges and limitations of our research.

The final chapter will summarize our research contributions. It will conclude the dissertation with a research outlook and suggestions for future work.

Chapter 7

Conclusions

This dissertation presented a new approach to assessing contextual data quality of clinical data. We have shown that it is possible to extend EMRs with algorithms and mechanisms to reason about data quality during clinical decision making. All stated research objectives from Chapter 1 have been met. We developed a novel data quality assessment method for individual data items that provides a contextual data quality assessment result that represents the relative degree of data quality trustworthiness for individual data items, provides assessment results in near real-time so they are available for reasoning about data quality during clinical decision making at the point of care, and provides those assessment results with a simplified contextual result representation. We developed integrations with interoperable and legacy EMR systems so that for interoperable EMR systems, data quality assessment results can be transmitted using open interface standards, and transmission of data quality assessment results is possible in low network bandwidth environments. For legacy EMR applications, data quality trust can function within the existing database models. Visualization of assessment results can be integrated with EMR system user interfaces. We demonstrated feasibility of the new approach with a visual prototype

that was applied to a clinical example in a lab setting, paving the way for further development and implementation with user evaluation in naturalistic system settings.

7.1 Contributions

This dissertation claims several direct contributions in original research:

1. A generic trust model and assessment method for contextual data quality of individual data items based on provenance (Chapter 3),

2. Two trust model and method extensions that provide additional verification and certification features beyond the core data quality trust model (Chapter 3),

3. A design pattern for integrating the assessment method with interoperable or legacy EMR systems, including data model integration and visualization with cognitive user interfaces (Chapter 4),

4. A prototype that demonstrates feasibility of applying assessment method and visualization in a pediatric hypertension clinical example context (Chapter 5), and

5. Guidance for applying assessment method and visualization to more complex clinical examples (Chapter 5).

7.2 Future Work

Several areas of our research require further investigation and offer opportunities for future research. Pediatric hypertension provided a simple and relatable real-world example to demonstrate our novel data quality assessment method. More complex clinical examples, such as allergy diagnosis, may provide additional value to patients

and healthcare practitioners. Model and method should now be applied and evaluated with more complex medical problems under increasingly realistic conditions to allow for scaling up to practice [23]. Since novel model and assessment method have only been subject to an ex-ante evaluation in an artificial environment, a user study with a more developed prototype implementation that is transferred to a naturalistic setting would be desirable to gain additional implementation knowledge, and feedback for further refinement of model and method. An important requirement for evaluating the success of information system implementations is engaging the actual system user to measure usage, intention to use, user satisfaction and net benefits [110, 111]. We recommend implementation and evaluation of model and method in other application domains that rely on data quality for decision making.

Appendix A

Literature Review Materials

A.1 Literature Review Queries

Database	**Search Query**	**Years**	**Hits**
Compendex / Inspec	(”Data Quality”) AND (EHR OR EMR OR electronic medical records OR electronic health records WN CV)		266
Compendex / Inspec	(”Data Quality Assessments” OR ”Data Quality Assessment” OR ”Data Quality Probes” OR ”Data Quality Probe”) WN KY AND (”Data Quality”) AND (EHR OR EMR OR electronic medical records OR electronic health records WN CV)		24
Compendex / Inspec	Automated AND (”Data Quality Assessments” OR ”Data Quality Assessment” OR ”Data Quality Probes” OR ”Data Quality Probe”) WN KY AND (”Data Quality”) AND (EHR OR EMR OR electronic medical records OR electronic health records WN CV)		1

Compendex / Inspec	("Data Quality") WN KY AND (electronic health records WN CV)		66
Compendex / Inspec	("Data Quality") WN KY AND (medical information systems WN CV)		303
Compendex / Inspec	(Auto*) WN KY AND ("Data Quality") WN KY AND (medical information systems WN CV)		45
Compendex / Inspec	("real-time") WN KY AND ("Data Quality") WN KY AND (medical information systems WN CV)		15
Compendex / Inspec	("monitoring") WN KY AND ("Data Quality") WN KY AND (medical information systems WN CV)		35
Compendex / Inspec	("Data Quality") WN KY AND (electronic health records WN CV)	2013-2020	66
Pubmed	(("Data Quality" OR "Data Accuracy"[Mh]) AND (EHR OR electronic medical record OR computerized medical record OR medical record systems, computerized [Mh]) AND English [lang]) AND (("2001"[Date - Publication] : "2020"[Date - Publication]))		986
Pubmed	("Data Quality Assessments" OR "Data Quality Assessment" OR "Data Quality Probes" OR "Data Quality Probe") AND (("Data Quality" OR "Data Accuracy"[Mh]) AND (EHR OR electronic medical record OR computerized medical record OR medical record systems, computerized [Mh]) AND English [lang]) AND (("2001"[Date - Publication] : "2020"[Date - Publication]))		66

Pubmed	Automated AND ("Data Quality Assessments" OR "Data Quality Assessment" OR "Data Quality Probes" OR "Data Quality Probe") AND (("Data Quality" OR "Data Accuracy"[Mh]) AND (EHR OR electronic medical record OR computerized medical record OR medical record systems, computerized [Mh]) AND English [lang]) AND (("2001"[Date - Publication] : "2020"[Date - Publication]))		4
Pubmed	(Data Accuracy[Mh]) AND (medical record systems, computerized [Mh])		236
Pubmed	("Data Quality" [TIAB]) AND (information systems [Mh])		1066
Pubmed	("Data Quality" [TIAB]) AND (Decision Support Systems, Clinical[Mh] OR medical record systems, computerized [Mh])		474
Pubmed	(Auto* [TIAB]) AND ("Data Quality" [TIAB]) AND (information systems [Mh])		139
Pubmed	("Data Quality" [TIAB]) AND (medical record systems, computerized [Mh])	2013-2020	293

A.2 Relevance Screening Form based on Abstract

Question	**Options**	**Exclusion If**
1. What type of source is the result?	Journal Article, Conference Paper, Other	Result of category "Other"
2. Is an abstract available? (Search for abstract using search engines)	Yes (go to question 3), No (go to question 2a)	

2a. Can it be concluded from the title that the article deals with data quality assessment in electronic health record systems?	Yes (go to question 3)	No
3. Does the article describe a data quality assessment method?	Yes (go to question 3a)	No
3a. Is the data quality assessment method described in title and abstract automated?	Automated (go to question 4), Unclear if automated or manual (go to question 4)	Manual
4. Does the article deal with Electronic Health Records?	Yes (article will remain for further review)	No

A.3 Eligibility Screening Form based on Full-Text

Question	**Options**	**Exclusion If**	**Additional Notes**
1. Is the full-text available?	Yes (go to question 2), No	No	
2. Is the full-text in English?	Yes (go to question 3), No	No	
3. Again: What type of source is the result?	Journal paper (go to question 4), conference paper (go to question 4), commentary, editorial, others	commentary, editorial, others	Article type may not always be identifiable based on title / abstract review only

4. Is the study primarily focused on data quality differences between electronic and paper-based workflows?	No	Yes	

A.4 Characterization Form based on Full-Text

Question	**Options**
1. What is the data quality assessment method?	
2. What data quality dimension does the article discuss?	
3. What is the temporal perspective of the data quality assessment method described?	Retrospective, Real-time, Prospective
4. How does the method perform data quality assessment?	

A.5 Articles Selected for Full-Text Review

Author(s)	Year	DQA Method(s) described	Data Quality Dimension	Temporal Perspective
Alvarez Sanchez et al. [52]	2019	Visual tool with missing value heatmap, correlation analysis (density plot) and outlier detection (MCD method)	Completeness, Accuracy, Redundancy, Readability	Retrospective
Botts et al. [112]	2014	NIST CDA validation tool to identify non-conformance of HIE messages with C32 standard	Conformance	Real-Time
Brown et al. [113]	2013	Data checks by distributed research data networks	Completeness, Accuracy, Correctness, Concordance	Retrospective
D'Amore et al. [114]	2018	XML Schematron validation rules to identify non-conformance of C-CDA documents, semantic rules testing with proprietary algorithm	Completeness, Conformance	Retrospective
Daymont et al. [41]	2017	Automated algorithm using standard deviation score and weighted moving average	Plausibility	Retrospective
Deng et al. [39]	2016	Validity check of temporal concordance between recorded hospital movement EHR timestamps	Concordance	Retrospective
Dungey et al. [115]	2014	automated calculation of data quality statistics and practice-based quality scores tailored to the intended data use of EMR data	Accuracy, Validity, Reliability, Timeliness, Relevance, Completeness, Integrity	Retrospective

Continued on next page

Table A.5 – *Continued from previous page*

Author(s)	**Year**	**DQA Method(s) described**	**Data Quality Dimension**	**Temporal Perspective**
Estiri and Murphy [42]	2018	Automatic encoding with neural networks for outlier detection, compared with silver standard	Plausibility	Retrospective
Estiri et al. [43]	2019	Unsupervised Clustering compared to standard deviation and anomaly detection (Mahalanobis distance)	Plausibility	Retrospective
Estiri et al. [116]	2018	webbased report showing results of completeness, conformance and missing data checks	Completeness, Conformance	Retrospective
Hart and Kuo [117]	2017	Metadata Rule-based data quality system for data validation with SQL queries	validity, accuracy, completeness, relevance, timeliness,availability, comparability, consistency, duplication, integrity and conformity	Real-time
Horth et al. [40]	2019	Record matching and concordance testing for structured and unstructured data between EHR and HIE databases	Concordance	Retrospective
Huser et al. [118]	2017	Several methods described from workshops, incl. PEDSnet Data Quality Assessment Toolkit (OMOP CDM)	Conformance	

Continued on next page

Table A.5 – *Continued from previous page*

Author(s)	**Year**	**DQA Method(s) described**	**Data Quality Dimension**	**Temporal Perspective**
Huser et al. [44]	2019	Rules-based data quality checks with OHDSI Achilles Heel tool to test model conformance, data completeness and data plausibility	Conformance, Completeness, Plausibility	Retrospective
Johnson et al. [119]	2015	Ontology-based rules for measurement of data quality dimensions	Correctness, Consistency, Completeness, Currency	Retrospective
Johnson et al. [120]	2016	Ontology-based rules for measurement of data quality task relevance and domain consistency		Retrospective
Kapsner et al. [45]	2019	MIRACUM DQA report for conformance, completeness and plausibility analysis of OMOP and i2b2 databases	Conformance, Completeness, Plausibility	Retrospective
Khare et al. [46]	2017	Semiautomated DQA of conformance, compleness and plausibility for PEDSnet (OMOP CDM)	Conformance, Completeness, Plausibility	Retrospective
Lettvin et al. [121]	2018	Automated DQA measures for completeness and accuracy in CancerLinQ	Completeness, Accuracy	Retrospective
Liaw et al. [122]	2014	DQA of ontology-based T2DM phenotyping algorithm	Accuracy	Retrospective
Lyle et al. [123]	2015	DQ scoring of HIE C32 documents for data richness, data correctness and semantic level	Completeness, Conformance	Retrospective
McCoy et al. [51]	2014	Automated link ratio as reputation metric for problem-medication pair accuracy	Accuracy, Provenance	Retrospective

Continued on next page

Table A.5 – *Continued from previous page*

Author(s)	**Year**	**DQA Method(s) described**	**Data Quality Dimension**	**Temporal Perspective**
Nobles et al. [124]	2015	Multi-site DQA of EHR for Completeness, Consistency, Availability for Use	Completeness, Consistency, Availability for Use	Retrospective
Noselli et al. [54]	2017	Visual tool to profile completeness and correctness DQ	Completeness, Correctness	Retrospective
Perimal-Lewis et al. [55]	2016	visual process-mining of patient control-flow data to highlight DQ issues	Completeness, Conformance	Retrospective
Pezoulas et al. [47]	2019	Automated method for detection and tracking of inconsistencies, missing values, outliers, and similarities	Completeness, Correctness, Consistency, Duplication, Plausibility	Retrospective
Rahimi et al. [125]	2014	Ontology-based algorithm to query ePBRN data repository for accuracy	Accuracy	Retrospective
Reimer et al. [126]	2016	DQA of Record linkage for longitudinal concordance in multi-site EHR setting	Completeness, Concordance	Retrospective
Singer et al. [127]	2016	Rule-based DQA queries to assess EMR data completeness	Completeness	Retrospective
Sirgo et al. [48]	2018	DQ comparison of automatic ICU-DaMa queries with gold standard for conformance, completeness and plausibility	Conformance, Completeness, Plausibility	Retrospective

Continued on next page

Table A.5 – *Continued from previous page*

Author(s)	**Year**	**DQA Method(s) described**	**Data Quality Dimension**	**Temporal Perspective**
Stoldt and Weber [49]	2020	DQ Probes for safety assurance with Data Quality by Contract	Conformance, Provenance, Concordance, Currency, Plausibility	Real-time
Sun et al. [30]	2017	Provenance-based DQA method	Provenance	Retrospective
Ta and Weng [50]	2019	DQA by normalizing/clustering concept frequencies with K-means	Temporal Plausibility, Consistency	Retrospective
Taggart et al. [128]	2015	Automated Structured Data Quality Report	Completeness, Correctness, Consistency, Duplication	Retrospective
Tate et al. [53]	2014	TrialViz visual tool to query for completeness and correlation of variables in EMR databases	Completeness, Correlation	Real-time
Tian et al. [129]	2019	DQA with Rules for Clinical Data Quality Assessment Based on OpenEHR Guideline Definition Language	Completeness	Retrospective
Tute et al. [130]	2019	DQA against OpenEHR Clinical Information Models for range-, format-, value set and cardinality constraints	Completeness, Conformance	Retrospective
Van der Bij et al. [131]	2017	Visual DQ feedback tool measuring data completeness	Completeness	Retrospective

Bibliography

[1] D. G. Gregg, U. R. Kulkarni, and A. S. Vinzé, "Understanding the Philosophical Underpinnings of Software Engineering Research in Information Systems," *Information Systems Frontiers*, vol. 3, no. 2, pp. 169–183, Jun. 2001. [Online]. Available: https://doi.org/10.1023/A:1011491322406

[2] "PROV Model Primer." [Online]. Available: https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

[3] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *International Journal of Medical Informatics*, vol. 114, pp. 57–65, Jun. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505618302466

[4] J.-P. Stoldt, M. Price, and J. Weber, "Towards a Clinical Analytics Adoption Maturity Framework for Primary Care," *Studies in Health Technology and Informatics*, vol. 257, pp. 399–403, 2019.

[5] J. C. Mandel, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records," *Journal of the American Medical Informatics Association*, vol. 23, no. 5, pp. 899–908, Sep. 2016. [Online]. Available: https://academic.oup.com/jamia/article/23/5/899/2379865

[6] J. H. Weber-Jahnke and F. Mason-Blakley, "The Safety of Electronic Medical Record (EMR) Systems: What Does EMR Safety Mean and How Can We Engineer Safer Systems?" *SIGHIT Rec.*, vol. 1, no. 2, pp. 13–22, Sep. 2011. [Online]. Available: http://doi.acm.org/10.1145/2047478.2047480

[7] W. R. Hogan and M. M. Wagner, "Accuracy of Data in Computer-based Patient Records," *Journal of the American Medical Informatics Association*, vol. 4, no. 5, pp. 342–355, 1997. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC61252/

[8] K. Thiru, A. Hassey, and F. Sullivan, "Systematic review of scope and quality of electronic patient record data in primary care," *BMJ : British Medical Journal*, vol. 326, no. 7398, p. 1070, May 2003. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC155692/

[9] K. S. Chan, J. B. Fowles, and J. P. Weiner, "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: A Review of the Literature," *Medical Care Research and Review*, vol. 67, no. 5, pp. 503–527, Oct. 2010. [Online]. Available: https://doi.org/10.1177/1077558709359007

[10] M. G. Kahn, T. J. Callahan, J. Barnard, A. E. Bauck, J. Brown, B. N. Davidson, H. Estiri, C. Goerg, E. Holve, S. G. Johnson, S.-T. Liaw, M. Hamilton-Lopez, D. Meeker, T. C. Ong, P. Ryan, N. Shang, N. G. Weiskopf, C. Weng, M. N. Zozus, and L. Schilling, "A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*, vol. 4, no. 1, p. 18, Sep. 2016. [Online]. Available: https://egems.academyhealth.org/article/10.13063/2327-9214.1244/

[11] N. G. Weiskopf and C. Weng, "Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research," *Journal of the American Medical Informatics Association : JAMIA*, vol. 20, no. 1, pp. 144–151, 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3555312/

[12] A. L. Terry, M. Stewart, S. Cejic, J. N. Marshall, S. de Lusignan, B. M. Chesworth, V. Chevendra, H. Maddocks, J. Shadd, F. Burge, and A. Thind, "A basic model for assessing primary health care electronic medical record data quality," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 30, Dec. 2019. [Online]. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0740-0

[13] J. Wider, "Survey: Only 20 Percent of Healthcare Executives Trust Their Data," Nov. 2021. [Online]. Available: https://www.hcinnovationgroup.com/analytics-ai/news/21244964/survey-only-20-percent-of-healthcare-executives-trust-their-data

[14] J. Horsky, G. J. Kuperman, and V. L. Patel, "Comprehensive Analysis of a Medication Dosing Error Related to CPOE," *Journal of the American Medical Informatics Association : JAMIA*, vol. 12, no. 4, pp. 377–382, 2005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1174881/

[15] E. Peters, N. F. Dieckmann, C. K. Mertz, P. Slovic, and J. H. Hibbard, "Bringing Meaning to Numbers: The Impact of Evaluative Categories on Decisions," *Journal of Experimental Psychology: Applied*, pp. 213–227, 2009.

[16] E. Peters, J. Hibbard, P. Slovic, and N. Dieckmann, "Numeracy skill and the communication, comprehension, and use of risk-benefit information," *Health Affairs (Project Hope)*, vol. 26, no. 3, pp. 741–748, Jun. 2007.

[17] T. D. Shanafelt, L. N. Dyrbye, C. Sinsky, O. Hasan, D. Satele, J. Sloan, and C. P. West, "Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction," *Mayo Clinic Proceedings*, vol. 91, no. 7, pp. 836–848, Jul. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0025619616302154

[18] P. J. Kroth, N. Morioka-Douglas, S. Veres, S. Babbott, S. Poplau, F. Qeadan, C. Parshall, K. Corrigan, and M. Linzer, "Association of Electronic Health Record Design and Use Factors With Clinician Stress and Burnout," *JAMA Network Open*, vol. 2, no. 8, p. e199609, Aug. 2019. [Online]. Available: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2748054

[19] E. R. Melnick, L. N. Dyrbye, C. A. Sinsky, M. Trockel, C. P. West, L. Nedelec, M. A. Tutty, and T. Shanafelt, "The Association Between Perceived Electronic Health Record Usability and Professional Burnout Among US Physicians," *Mayo Clinic Proceedings*, vol. 95, no. 3, pp. 476–487, Mar. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0025619619308365

[20] M. Shaw, "Writing good software engineering research papers," in *25th International Conference on Software Engineering, 2003. Proceedings.* Portland, OR, USA: IEEE, 2003, pp. 726–736. [Online]. Available: http://ieeexplore.ieee.org/document/1201262/

[21] M. Montesi and P. Lago, "Software engineering article types: An analysis of the literature. Journal of Systems and Software, 81(10), 1694-1714," *Journal of Systems and Software*, vol. 81, pp. 1694–1714, Oct. 2008.

[22] R. J. Wieringa, *Design Science Methodology for Information Systems and Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. [Online]. Available: http://link.springer.com/10.1007/978-3-662-43839-8

[23] R. Wieringa, "Empirical research methods for technology validation: Scaling up to practice," *Journal of Systems and Software*, vol. 95, pp. 19–31, Sep. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121213002793

[24] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.

[25] S. De Lusignan, "The optimum granularity for coding diagnostic data in primary care: report of a workshop of the EFMI Primary Care Informatics Working Group at MIE 2005," *Journal of Innovation in Health Informatics*, vol. 14, no. 2, pp. 133–137, Jun. 2006. [Online]. Available: http://hijournal.bcs.org/index.php/jhi/article/view/624

[26] H. Carr, S. de Lusignan, H. Liyanage, S.-T. Liaw, A. Terry, and I. Rafi, "Defining dimensions of research readiness: a conceptual model for primary care research networks," *BMC family practice*, vol. 15, p. 169, Nov. 2014.

[27] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996. [Online]. Available: http://www.jstor.org/stable/40398176

[28] J. Golbeck, B. Parsia, and J. Hendler, "Trust Networks on the Semantic Web," in *Cooperative Information Agents VII*, ser. Lecture Notes in Computer Science, M. Klusch, A. Omicini, S. Ossowski, and H. Laamanen, Eds. Berlin, Heidelberg: Springer, 2003, pp. 238–249.

[29] E. Bertino and H.-S. Lim, "Assuring Data Trustworthiness - Concepts and Research Challenges," in *Secure Data Management*, ser. Lecture Notes in Computer Science, W. Jonker and M. Petković, Eds. Berlin, Heidelberg: Springer, 2010, pp. 1–12.

[30] Y. Sun, T. Lu, and N. Gu, "A method of electronic health data quality assessment: Enabling data provenance," in *2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. Wellington, New Zealand: IEEE, Apr. 2017, pp. 233–238. [Online]. Available: http://ieeexplore.ieee.org/document/8066700/

[31] S. Rajbhandari, I. Wootten, A. S. Ali, and O. F. Rana, "Evaluating Provenance-based Trust for Scientific Workflows," in *Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)*, vol. 1, May 2006, pp. 365–372.

[32] S. Rajbhandari, O. F. Rana, and I. Wootten, "A fuzzy model for calculating workflow trust using provenance data," in *Proceedings of the 15th ACM Mardi Gras conference on From lightweight mash-ups to lambda grids: Understanding the spectrum of distributed computing requirements, applications, tools, infrastructures, interoperability, and the incremental adoption of key capabilities - MG '08*. Baton Rouge, Louisiana: ACM Press, 2008, p. 1. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1341811.1341823

[33] M. Richardson, R. Agrawal, and P. Domingos, "Trust Management for the Semantic Web," in *The Semantic Web - ISWC 2003*, ser. Lecture Notes in Computer Science, D. Fensel, K. Sycara, and J. Mylopoulos, Eds. Berlin, Heidelberg: Springer, 2003, pp. 351–368.

[34] N. Prat and S. Madnick, "Measuring Data Believability: A Provenance Approach," *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, 2008.

[35] H. Arksey and L. O'Malley, "Scoping studies: towards a methodological framework," *International Journal of Social Research Methodology*, vol. 8, no. 1, pp. 19–32, Feb. 2005. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/1364557032000119616

[36] B. Kitchenham and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Technical report, EBSE Technical Report EBSE-2007-01, 2007.

[37] D. Moher, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," *Annals of Internal Medicine*, vol. 151, no. 4, p. 264, Aug. 2009. [Online]. Available: http://annals.org/article.aspx?doi=10.7326/0003-4819-151-4-200908180-00135

[38] J. S. Brown, M. Kahn, and D. Toh, "Data Quality Assessment for Comparative Effectiveness Research in Distributed Data Networks:," *Medical Care*, vol. 51, pp. S22–S29, Aug. 2013. [Online]. Available: http://journals.lww.com/00005650-201308001-00006

[39] X. Deng, W. H. Lin, T. E-Shyong, K. Y. H. Eric, M. K. Salloway, and T. C. Seng, "From descriptive to diagnostic analytics for assessing data quality: An application to temporal data elements in electronic health records," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. Las Vegas, NV, USA: IEEE, Feb. 2016, pp. 236–239. [Online]. Available: http://ieeexplore.ieee.org/document/7455878/

[40] R. Z. Horth, S. Wagstaff, T. Jeppson, V. Patel, J. McClellan, N. Bissonette, M. Friedrichs, and A. C. Dunn, "Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension," *BMC Public Health*, vol. 19, no. 1, p. 1106, Dec. 2019. [Online]. Available: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7367-z

[41] C. Daymont, M. E. Ross, A. Russell Localio, A. G. Fiks, R. C. Wasserman, and R. W. Grundmeier, "Automated identification of implausible values in growth data from pediatric electronic health records," *Journal of the American Medical Informatics Association*, vol. 24, no. 6, pp. 1080–1087, Nov. 2017. [Online]. Available: http://academic.oup.com/jamia/article/24/6/1080/3767271/Automated-identification-of-implausible-values-in

[42] H. Estiri and S. N. Murphy, "Semi-supervised Encoding for Outlier Detection in Clinical Observation Data," Bioinformatics, preprint, May 2018. [Online]. Available: http://biorxiv.org/lookup/doi/10.1101/334771

[43] H. Estiri, J. G. Klann, and S. N. Murphy, "A clustering approach for detecting implausible observation values in electronic health records data," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 142, Dec. 2019. [Online]. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0852-6

[44] V. Huser, X. Li, Z. Zhang, S. Jung, R. W. Park, J. Banda, H. Razzaghi, A. Londhe, and K. Natarajan, "Extending Achilles Heel Data Quality Tool with New Rules Informed by Multi-Site Data Quality Comparison," *Studies in Health Technology and Informatics*, vol. 264, pp. 1488–1489, Aug. 2019.

[45] L. A. Kapsner, M. O. Kampf, S. A. Seuchter, G. Kamdje-Wabo, T. Gradinger, T. Ganslandt, S. Mate, J. Gruendner, D. Kraska, and H.-U. Prokosch, "Moving Towards an EHR Data Quality Framework: The MIRACUM Approach," *Studies in Health Technology and Informatics*, vol. 267, pp. 247–253, Sep. 2019.

[46] R. Khare, L. Utidjian, B. J. Ruth, M. G. Kahn, E. Burrows, K. Marsolo, N. Patibandla, H. Razzaghi, R. Colvin, D. Ranade, M. Kitzmiller, D. Eckrich, and L. C. Bailey, "A longitudinal analysis of data quality in a large pediatric data research network," *Journal of the American Medical Informatics Association*, vol. 24, no. 6, pp. 1072–1079, Nov. 2017. [Online]. Available: http://academic.oup.com/jamia/article/24/6/1072/3238563/A-longitudinal-analysis-of-data-quality-in-a-large

[47] V. C. Pezoulas, K. D. Kourou, F. Kalatzis, T. P. Exarchos, A. Venetsanopoulou, E. Zampeli, S. Gandolfo, F. Skopouli, S. De Vita, A. G. Tzioufas, and D. I. Fotiadis, "Medical data quality assessment: On the development of an automated framework for medical data curation," *Computers in Biology and Medicine*, vol. 107, pp. 270–283, Apr. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010482519300733

[48] G. Sirgo, F. Esteban, J. Gómez, G. Moreno, A. Rodríguez, L. Blanch, J. J. Guardiola, R. Gracia, L. De Haro, and M. Bodí, "Validation of the ICU-DaMa tool for automatically extracting variables for minimum dataset and quality indicators: The importance of data quality assessment," *International Journal of Medical Informatics*, vol. 112, pp. 166–172, Apr. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1386505618300443

[49] J.-P. Stoldt and J. H. Weber, "Safety Improvement for SMART on FHIR Apps with Data Quality by Contract," in *2020 IEEE International*

Conference on Software Architecture Companion (ICSA-C). Salvador, Brazil: IEEE, Mar. 2020, pp. 192–195. [Online]. Available: https://ieeexplore.ieee.org/document/9095642/

[50] C. N. Ta and C. Weng, "Detecting Systemic Data Quality Issues in Electronic Health Records," *Studies in health technology and informatics*, vol. 264, pp. 383–387, Aug. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6857180/

[51] A. B. McCoy, A. Wright, D. Rogith, S. Fathiamini, A. J. Ottenbacher, and D. F. Sittig, "Development of a clinician reputation metric to identify appropriate problem-medication pairs in a crowdsourced knowledge base," *Journal of Biomedical Informatics*, vol. 48, pp. 66–72, Apr. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1532046413001913

[52] R. Álvarez Sánchez, A. Beristain Iraola, G. Epelde Unanue, and P. Carlin, "TAQIH, a tool for tabular data quality assessment and improvement in the context of health data," *Computer Methods and Programs in Biomedicine*, vol. 181, p. 104824, Nov. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169260718304188

[53] A. R. Tate, N. Beloff, B. Al-Radwan, J. Wickson, S. Puri, T. Williams, T. Van Staa, and A. Bleach, "Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 292–298, Mar. 2014. [Online]. Available: https://academic.oup.com/jamia/article-lookup/doi/10.1136/amiajnl-2013-001847

[54] M. Noselli, D. Mason, M. A. Mohammed, and R. A. Ruddle, "MonAT: A Visual Web-based Tool to Profile Health Data Quality:," in *Proceedings of*

the 10th International Joint Conference on Biomedical Engineering Systems and Technologies. Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 26–34. [Online]. Available: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006114200260034

[55] L. Perimal-Lewis, D. Teubner, P. Hakendorf, and C. Horwood, "Application of process mining to assess the data quality of routinely collected time-based performance data sourced from electronic health records by validating process conformance," *Health Informatics Journal*, vol. 22, no. 4, pp. 1017–1029, Dec. 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1460458215604348

[56] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II," *Information Sciences*, vol. 8, no. 4, pp. 301–357, Jan. 1975. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0020025575900468

[57] L. Zadeh, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103–111, May 1996, conference Name: IEEE Transactions on Fuzzy Systems.

[58] S. Medasani, J. Kim, and R. Krishnapuram, "An overview of membership function generation techniques for pattern recognition," *International Journal of Approximate Reasoning*, vol. 19, no. 3-4, pp. 391–417, Oct. 1998. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888613X98100178

[59] J. Mendel, L. Zadeh, E. Trillas, R. Yager, J. Lawry, H. Hagras, and S. Guadarrama, "What Computing with Words Means to Me [Discussion Forum," *IEEE Computational Intelligence Magazine*, vol. 5, no. 1, pp. 20–26, Feb. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5386101/

[60] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, Jan. 1975. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020737375800022

[61] S. S. Izquierdo and L. R. Izquierdo, "Mamdani Fuzzy Systems for Modelling and Simulation: A Critical Assessment," *Journal of Artificial Societies and Social Simulation*, vol. 21, no. 3, p. 2, 2018.

[62] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, Oct. 1994, google-Books-ID: 95As2OF67f0C.

[63] "scikit-fuzzy: Scikit-Fuzzy version 0.4.2," Feb. 2021, original-date: 2013-03-19T04:27:23Z. [Online]. Available: https://github.com/scikit-fuzzy/scikit-fuzzy

[64] J. H. Weber, M. Price, and I. Davies, "Data Quality by Contract – Towards an Architectural View for Data Quality in Health Information Systems," in *Knowledge Representation for Health Care*, D. Riaño, R. Lenz, S. Miksch, M. Peleg, M. Reichert, and A. ten Teije, Eds. Cham: Springer International Publishing, 2015, vol. 9485, pp. 143–157. [Online]. Available: http://link.springer.com/10.1007/978-3-319-26585-8_10

[65] B. Meyer, "Applying 'design by contract'," *Computer*, vol. 25, no. 10, pp. 40–51, Oct. 1992.

[66] "Comparison - FHIR v4.0.1." [Online]. Available: https://www.hl7.org/fhir/comparison.html

[67] "HL7 Argonaut Project Wiki." [Online]. Available: https://argonautwiki.hl7.org/index.php?title=Main_Page

[68] "SMART Health IT." [Online]. Available: https://smarthealthit.org/

[69] "SMART App Launch Framework." [Online]. Available: http://www.hl7.org/fhir/smart-app-launch/

[70] D. Hardt, "The OAuth 2.0 Authorization Framework." [Online]. Available: https://tools.ietf.org/html/rfc6749

[71] "OpenID Connect | OpenID." [Online]. Available: https://openid.net/connect/

[72] "Overview - FHIR v4.0.1." [Online]. Available: https://www.hl7.org/fhir/overview.html

[73] "HAPI FHIR - The Open Source FHIR API for Java." [Online]. Available: https://hapifhir.io/

[74] Y. K. Rivera Sánchez, S. A. Demurjian, and M. S. Baihan, "A service-based RBAC & MAC approach incorporated into the FHIR standard," *Digital Communications and Networks*, vol. 5, no. 4, pp. 214–225, Nov. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352864817301803

[75] M. A. Hussain, S. G. Langer, and M. Kohli, "Learning HL7 FHIR Using the HAPI FHIR Server and Its Use in Medical Imaging with the SIIM Dataset," *Journal of Digital Imaging*, vol. 31, no. 3, pp. 334–340, Jun. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5959839/

[76] K. D. Mandl and I. S. Kohane, "No Small Change for the Health Information Economy," *New England Journal of Medicine*, vol. 360, no. 13, pp. 1278–1281, Mar. 2009, publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMp0900411. [Online]. Available: https://doi.org/10.1056/NEJMp0900411

[77] S. Purkayastha, R. Allam, P. Maity, and J. W. Gichoya, "Comparison of Open-Source Electronic Health Record Systems Based on Functional and User Performance Criteria," *Healthcare Informatics Research*, vol. 25, no. 2, pp. 89–98, Apr. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6517630/

[78] "Home - Documentation - OpenMRS Wiki." [Online]. Available: https://wiki.openmrs.org/

[79] "OpenMRS FHIR Module - Projects - OpenMRS Wiki." [Online]. Available: https://wiki.openmrs.org/display/projects/OpenMRS+FHIR+Module

[80] "openEHR Architecture Overview." [Online]. Available: https://specifications.openehr.org/releases/BASE/Release-1.0.3/architecture_overview.html

[81] J. L. Cardoso de Moraes, W. L. de Souza, L. F. Pires, and A. F. do Prado, "A methodology based on openEHR archetypes and software agents for developing e-health applications reusing legacy systems," *Computer Methods and Programs in Biomedicine*, vol. 134, pp. 267–287, Oct. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926071630298X

[82] L. Min, Q. Tian, X. Lu, and H. Duan, "Modeling EHR with the openEHR approach: an exploratory study in China," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 75, Aug. 2018. [Online]. Available: https://doi.org/10.1186/s12911-018-0650-6

[83] M. Zahabi, D. B. Kaber, and M. Swangnetr, "Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation," *Human Factors*, vol. 57, no. 5, pp.

805–834, Aug. 2015, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0018720815576827

[84] J. M. Borovina Josko and J. E. Ferreira, "Visualization properties for data quality visual assessment: An exploratory case study," *Information Visualization*, vol. 16, no. 2, pp. 93–112, Apr. 2017, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/1473871616629516

[85] J. Gerhardt-Powals, "Cognitive engineering principles for enhancing human-computer performance," *International Journal of Human–Computer Interaction*, vol. 8, no. 2, pp. 189–211, Apr. 1996, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447319609526147. [Online]. Available: https://doi.org/10.1080/10447319609526147

[86] D. Ripalda, C. Guevara, and A. Garrido, "Framework Based on Gestalt Principles to Design Mobile Interfaces for a Better User Experience," in *Advances in Usability, User Experience, Wearable and Assistive Technology*, ser. Advances in Intelligent Systems and Computing, T. Ahram and C. Falcão, Eds. Cham: Springer International Publishing, 2020, pp. 158–165.

[87] R. Marcilly, E. Ammenwerth, E. Roehrer, J. Niès, and M.-C. Beuscart-Zéphir, "Evidence-based usability design principles for medication alerting systems," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 69, Jul. 2018. [Online]. Available: https://doi.org/10.1186/s12911-018-0615-9

[88] F. Ricca, G. Scanniello, M. Torchiano, G. Reggio, and E. Astesiano, "On the effectiveness of screen mockups in requirements engineering: results from an internal replication," in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM

'10. New York, NY, USA: Association for Computing Machinery, Sep. 2010, pp. 1–10. [Online]. Available: http://doi.org/10.1145/1852786.1852809

[89] S. Watts, G. Shankaranarayanan, and A. Even, "Data quality assessment in context: A cognitive perspective," *Decision Support Systems*, vol. 48, no. 1, pp. 202–211, Dec. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923609001766

[90] T. Taft, C. Staes, S. Slager, and C. Weir, "Adapting Nielsen's Design Heuristics to Dual Processing for Clinical Decision Support," *AMIA Annual Symposium Proceedings*, vol. 2016, pp. 1179–1188, Feb. 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333283/

[91] S. L. Jackson, Z. Zhang, J. L. Wiltz, F. Loustalot, M. D. Ritchey, A. B. Goodman, and Q. Yang, "Hypertension Among Youths - United States, 2001-2016," *MMWR. Morbidity and mortality weekly report*, vol. 67, no. 27, pp. 758–762, Jul. 2018.

[92] M. L. Hansen, P. W. Gunn, and D. C. Kaelber, "Underdiagnosis of Hypertension in Children and Adolescents," *JAMA*, vol. 298, no. 8, p. 874, Aug. 2007. [Online]. Available: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.298.8.874

[93] D. Perloff, C. Grim, J. Flack, E. D. Frohlich, M. Hill, M. McDonald, and B. Z. Morgenstern, "Human blood pressure determination by sphygmomanometry," *Circulation*, vol. 88, no. 5 Pt 1, pp. 2460–2470, Nov. 1993.

[94] C. Selenta, B. E. Hogan, and W. Linden, "How Often Do Office Blood Pressure Measurements Fail to Identify True Hypertension?: An Exploration of White-

Coat Normotension," *Archives of Family Medicine*, vol. 9, no. 6, p. 533, Jan. 2000, publisher: JAMA.

[95] B. A. Rodrigues Filho, R. F. Farias, and W. dos Anjos, "Evaluating the impact of measurement uncertainty in blood pressure measurement on hypertension diagnosis," *Blood Pressure Monitoring*, vol. 23, no. 3, pp. 141–147, Jun. 2018.

[96] A. Podoll, M. Grenier, B. Croix, and D. I. Feig, "Inaccuracy in Pediatric Outpatient Blood Pressure Measurement," *Pediatrics*, vol. 119, no. 3, pp. e538–e543, Mar. 2007, publisher: American Academy of Pediatrics Section: Article. [Online]. Available: https://pediatrics.aappublications.org/content/119/3/e538

[97] Stergiou George S., Boubouchairopoulou Nadia, and Kollias Anastasios, "Accuracy of Automated Blood Pressure Measurement in Children," *Hypertension*, vol. 69, no. 6, pp. 1000–1006, Jun. 2017, publisher: American Heart Association. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/hypertensionaha.116.08553

[98] "smart-on-fhir/bp-centiles-app," Nov. 2020, original-date: 2015-06-24T14:41:16Z. [Online]. Available: https://github.com/smart-on-fhir/bp-centiles-app

[99] "SMART App Gallery." [Online]. Available: https://gallery.smarthealthit.org/app/cardiac-risk

[100] S. M. Salerno, J. L. Jackson, and E. P. Berbano, "Effect of Oral Pseudoephedrine on Blood Pressure and Heart Rate: A Meta-analysis," *Archives of Internal Medicine*, vol. 165, no. 15, p. 1686, Aug. 2005.

[Online]. Available: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.165.15.1686

[101] "Medical Device Recalls." [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfres/res.cfm

[102] J. T. Flynn, D. C. Kaelber, C. M. Baker-Smith, D. Blowey, A. E. Carroll, S. R. Daniels, S. D. d. Ferranti, J. M. Dionne, B. Falkner, S. K. Flinn, S. S. Gidding, C. Goodwin, M. G. Leu, M. E. Powers, C. Rea, J. Samuels, M. Simasek, V. V. Thaker, E. M. Urbina, and S. o. S. a. M. o. H. B. P. I. Children, "Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents," *Pediatrics*, vol. 140, no. 3, Sep. 2017, publisher: American Academy of Pediatrics Section: From the American Academy of Pediatrics. [Online]. Available: https://pediatrics.aappublications.org/content/140/3/e20171904

[103] C. Mailhol, F. Giordano-Labadie, V. Lauwers-Cances, A. Ammoury, C. Paul, and F. Rance, "Point prevalence and risk factors for food allergy in a cohort of 386 children with atopic dermatitis attending a multidisciplinary dermatology/paediatric allergy clinic," *European Journal of Dermatology*, vol. 24, no. 1, pp. 63–69, Jan. 2014.

[104] L. Lteif and L. S. Eiland, "The Basics of Penicillin Allergy: What A Clinician Should Know," *Pharmacy: Journal of Pharmacy Education and Practice*, vol. 7, no. 3, p. 94, Jul. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6789476/

[105] M. E. Pichichero, "Cephalosporins can be prescribed safely for penicillin-allergic patients," *The Journal of Family Practice*, vol. 55, no. 2, Feb. 2006.

[Online]. Available: https://www.mdedge.com/familymedicine/article/62044/cephalosporins-can-be-prescribed-safely-penicillin-allergic-patients

[106] "AllergyIntolerance - FHIR v4.0.1." [Online]. Available: https://www.hl7.org/fhir/allergyintolerance.html

[107] "Condition - FHIR v4.0.1." [Online]. Available: https://www.hl7.org/fhir/condition.html

[108] J.-P. Stoldt and J. H. Weber, "Provenance-based Trust Model for Assessing Data Quality during Clinical Decision Making," in *2021 IEEE/ACM 3rd International Workshop on Software Engineering for Healthcare (SEH)*, Jun. 2021, pp. 24–31.

[109] P. Runeson, M. Höst, R. Austen, and B. Regnell, *Case Study Research in Software Engineering – Guidelines and Examples*. John Wiley and Sons, 2012.

[110] W. H. DeLone and E. R. McLean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30, 2003. [Online]. Available: http://www.jstor.org/stable/40398604

[111] F. Lau, M. Price, and K. Keshavjee, "From Benefits Evaluation to Clinical Adoption: Making Sense of Health Information System Success in Canada," *Healthcare Quarterly*, Jan. 2011.

[112] N. Botts, O. Bouhaddou, J. Bennett, E. Pan, C. Byrne, L. Mercincavage, L. Olinger, E. Hunolt, and T. Cullen, "Data Quality and Interoperability Challenges for eHealth Exchange Participants: Observations from the Department of Veterans Affairs' Virtual Lifetime Electronic Record Health Pilot Phase," *AMIA Annual Symposium Proceedings*, vol. 2014, pp. 307–314,

Nov. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419918/

[113] J. S. Brown, M. Kahn, and S. Toh, "Data quality assessment for comparative effectiveness research in distributed data networks," *Medical Care*, vol. 51, no. 8 Suppl 3, pp. S22–29, Aug. 2013.

[114] J. D'Amore, O. Bouhaddou, S. Mitchell, C. Li, R. Leftwich, T. Turner, M. Rahn, M. Donahue, and J. Nebeker, "Interoperability Progress and Remaining Data Quality Barriers of Certified Health Information Technologies," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2018, pp. 358–367, 2018.

[115] S. Dungey, N. Beloff, S. Puri, R. Boggon, T. Williams, and A. R. Tate, "A pragmatic approach for measuring data quality in primary care databases," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. Valencia, Spain: IEEE, Jun. 2014, pp. 797–800. [Online]. Available: http://ieeexplore.ieee.org/document/6864484/

[116] H. Estiri, K. A. Stephens, J. G. Klann, and S. N. Murphy, "Exploring completeness in clinical data research networks with DQe-c," *Journal of the American Medical Informatics Association*, vol. 25, no. 1, pp. 17–24, Jan. 2018. [Online]. Available: https://academic.oup.com/jamia/article/25/1/17/4562678

[117] R. Hart and M.-H. Kuo, "Better Data Quality for Better Healthcare Research Results - A Case Study," *Studies in Health Technology and Informatics*, vol. 234, pp. 161–166, 2017.

[118] V. Huser, M. G. Kahn, J. S. Brown, and R. Gouripeddi, "Methods for examining data quality in healthcare integrated data repositories," in *Biocomputing 2018*. Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC,

Jan. 2018, pp. 628–633. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789813235533_0059

[119] S. G. Johnson, S. Speedie, G. Simon, V. Kumar, and B. L. Westra, "A Data Quality Ontology for the Secondary Use of EHR Data," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2015, pp. 1937–1946, 2015.

[120] S. Johnson, S. Speedie, G. Simon, V. Kumar, and B. Westra, "Application of an Ontology for Characterizing Data Quality for a Secondary Use of EHR Data," *Applied Clinical Informatics*, vol. 07, no. 01, pp. 69–88, Jan. 2016. [Online]. Available: http://www.thieme-connect.de/DOI/DOI?10.4338/ACI-2015-08-RA-0107

[121] R. J. Lettvin, A. Wayal, A. McNutt, R. S. Miller, and R. Hauser, "Assessment and Stratification of High-Impact Data Elements in Electronic Clinical Quality Measures: A Joint Data Quality Initiative Between CancerLinQ® and Cancer Treatment Centers of America," *JCO Clinical Cancer Informatics*, no. 2, pp. 1–10, Dec. 2018. [Online]. Available: https://ascopubs.org/doi/10.1200/CCI.17.00139

[122] S.-T. Liaw, J. Taggart, H. Yu, S. d. Lusignan, C. Kuziemsky, and A. Hayen, "Integrating electronic health record information to support integrated care: Practical application of ontologies to improve the accuracy of diabetes disease registers," *Journal of Biomedical Informatics*, vol. 52, pp. 364–372, Dec. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1532046414001798

[123] J. Lyle, O. Bouhaddou, N. Botts, M. Swall, E. Pan, T. Cullen, M. Donahue, and N. Hsing, "Veterans Health Administration Experience with Data

Quality Surveillance of Continuity of Care Documents: Interoperability Challenges for eHealth Exchange Participants," *AMIA Annual Symposium Proceedings*, vol. 2015, pp. 870–879, Nov. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765662/

[124] A. L. Nobles, K. Vilankar, H. Wu, and L. E. Barnes, "Evaluation of data quality of multisite electronic health record data for secondary analysis," in *2015 IEEE International Conference on Big Data (Big Data)*. Santa Clara, CA, USA: IEEE, Oct. 2015, pp. 2612–2620. [Online]. Available: http://ieeexplore.ieee.org/document/7364060/

[125] A. Rahimi, S.-T. Liaw, J. Taggart, P. Ray, and H. Yu, "Validating an ontology-based algorithm to identify patients with Type 2 Diabetes Mellitus in Electronic Health Records," *International Journal of Medical Informatics*, vol. 83, no. 10, pp. 768–778, Oct. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1386505614001038

[126] A. P. Reimer, A. Milinovich, and E. A. Madigan, "Data Quality Assessment Framework to Assess Electronic Medical Record Data for use in Research," *International journal of medical informatics*, vol. 90, pp. 40–47, Jun. 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4845906/

[127] A. Singer, S. Yakubovich, A. L. Kroeker, B. Dufault, R. Duarte, and A. Katz, "Data quality of electronic medical records in Manitoba: do problem lists accurately reflect chronic disease billing diagnoses?" *Journal of the American Medical Informatics Association*, vol. 23, no. 6, pp. 1107–1112, Nov. 2016. [Online]. Available: https://academic.oup.com/jamia/article-lookup/doi/10.1093/jamia/ocw013

[128] J. Taggart, S.-T. Liaw, and H. Yu, "Structured data quality reports to improve EHR data quality," *International Journal of Medical Informatics*, vol. 84, no. 12, pp. 1094–1098, Dec. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1386505615300435

[129] Q. Tian, Z. Han, J. An, X. Lu, and H. Duan, "Representing Rules for Clinical Data Quality Assessment Based on OpenEHR Guideline Definition Language," *Studies in Health Technology and Informatics*, vol. 264, pp. 1606–1607, Aug. 2019.

[130] E. Tute, A. Wulff, M. Marschollek, and M. Gietzelt, "Clinical Information Model Based Data Quality Checks: Theory and Example," *Studies in Health Technology and Informatics*, vol. 258, pp. 80–84, 2019.

[131] S. van der Bij, N. Khan, P. ten Veen, D. H. de Bakker, and R. A. Verheij, "Improving the quality of EHR recording in primary care: a data quality feedback tool," *Journal of the American Medical Informatics Association*, vol. 24, no. 1, pp. 81–87, Jan. 2017. [Online]. Available: https://academic.oup.com/jamia/article-lookup/doi/10.1093/jamia/ocw054