Breaking Data Silos in Healthcare: A Novel Framework for Standardizing and Integrating NHS Medical Data for Advanced Analytics

Daniel Thomas
Email: daniel.cromwel@gmail.com

***Abstract*—The rapid expansion of medical data within the National Health Service (NHS) presents both opportunities and challenges in leveraging healthcare analytics for improved patient outcomes and research. However, disparate data sources, inconsistent formats, and the lack of standardized integration mechanisms hinder effective data utilization. This study proposes a novel framework for standardizing and integrating NHS medical data by addressing structural heterogeneity, semantic inconsistencies, and interoperability gaps. The framework leverages machine learning techniques for data harmonization and Natural Language Processing (NLP) to extract insights from unstructured clinical notes. Additionally, I introduce a hybrid model that combines ontology-based mapping with federated learning to enhance data interoperability across healthcare institutions while ensuring data security and compliance with privacy regulations. The proposed approach is validated using real-world NHS datasets to assess its effectiveness in improving data accessibility and analytical performance. This research aims to bridge the gap between fragmented healthcare data and actionable insights, paving the way for more efficient, data-driven decision-making in clinical and research settings.**

I. RESEARCH PROBLEM

Medical data in the UK NHS comes in many different shapes and sizes and varies from highly structured and rigorously controlled data conforming to managed ontologies and datasets to photographed collections of handwritten and mostly illegible clinician notes.

Data interoperability in healthcare in the U.K. has only been partially achieved for a small number of specific use cases; overall, despite the heavy investment in new technologies, the majority of patient data and clinical trial results remain in local silos, unavailable for most medical research. Where specific funding is made available for particular projects, such as with Genomics England (REF), an effort is made to integrate research data and patient data.

However, most of the data from the UK's National Health Service (NHS) remains inaccessible for medical research. This is because it is used in a variety of commercial software systems that have been built at different times and are based on various standards. There are over 130 different data models and standards (ref: TRUD REF) currently being used in the NHS, and new ones are being introduced and implemented each year. At present, the UK NHS is preparing to implement SNOMED across all systems by 2020; this has largely been achieved in primary care but has hardly been started in secondary care systems. The main reason for the success of primary care is down to the fact that there are five main vendors of primary care systems, all of whom regard the UK primary care market as a major part of their business, and hence, all of them have agreed to implement the standard.

However, in secondary care, the larger hospital trusts will be running 2-300 different systems, many of these from vendors with bases outside the UK, who have little incentive to make major changes to systems that they are selling in a minority marketplace. Efforts by NHS England and NHS Digital to impose standards have met with partial success with specific standards in specific areas, but so far no one standard has been adopted across the board.

As well as the technical challenges, there are also legal issues imposed by the Health and Social Care Act 2012, NHS Act 2006, the Health and Social Care Act 2012, the Data Protection Act 1998, the Human Rights Act, and the shortly to be introduced General Data Protection Regulations. The complexity of these different acts results can result in data on a patient's other conditions (say liver disease) being withheld from Emergency Care Clinicians when they are performing, for instance, emergency heart surgery. Whilst such legislation clearly is intended to protect the patient's privacy, it can have unforeseen side effects not only on emergency medical care but also on the ability and ease with which medical data can be aggregated for further research purposes.

Medical data, when aggregated, is generally stored in relational databases, excel spreadsheets, or unstructured PDF and Word documents, and querying this data can be very difficult in all these cases. Reports are sent through to NHS Digital to satisfy reporting requirements in hundreds of different spread-sheet formats, with no real verification on the datasets contained. There is no question that the information could be valuable for research; however, extracting data for analysis is very often too costly to perform. When datasets are prepared for analysis, it is often discovered that the structure of the data, when stored, may not be the same as the structure it is required now, and there may be no exact record of how the data was originally recorded. For instance, gender may be stored in one system as an enumeration of *yes or no* and in another system as *yes, no, unknown*, if I write a query assuming the former, then I may arrive at incorrect conclusions if the data is actually collect in the format of the latter, but assumed to

be conforming to the former.

II. OUTLINE OF OBJECTIVES

This work started by examining ways in which semantic interoperability could be achieved within the Healthcare sector. *Semantic web* technology has been used to solve some of these interoperability concerns within the healthcare domain, and so have a number of other data standards technologies. Most of these, like the semantic web solutions, fail mostly because they seek to impose a *fixed-framework* type solution across the board. By fixed-framework solution, I mean one which assumes that a trust or group of trusts will impose one framework for everything; for instance, all the data will from time a be stored in triple-stores, or from time b, all suppliers of oncology-related systems will use ICD10 coding for their datasets. Such *fixed framework* solutions generally need a significant investment in ETL technology to make them work with legacy data, and so initially, they are no less expensive than just employing an ETL solution.

In this work, I have investigated ways in which a solution can be found that is practical enough to work with multiple dataset standards, multiple data stores, and multiple systems. My objectives are to explore ways in which semantic interoperability can be achieved, or at least partially achieved, with existing technology by using model-driven engineering principles and applying these to the way in which data is stored, transformed, and analyzed within the UK Healthcare sector.

Initially, I looked at applying the principles defined in the ISO11179 (0) standard to UK Healthcare data U.K. standards; however, having built a prototype metadata registry on the lines described, the standard was found to be flawed and very difficult to apply in practice. The standard itself was ambiguous and hard to understand for terminologists and clinicians, and as a result, many features were not used, and many meta-data points were simply ignored.

The metadata registry continued to be developed in partnership with Genomics England over 2 years, and practical results refined the meta-model. Soon after the start of this exercise, a separate piece of work was undertaken to define a formal metamodel in Z, and this is the basis for the third version of the metadata registry, which is currently only available as a prototype. Initially, the idea was to clarify the ideas outlined in the ISO11179 standard, but this exercise still has to be completed.

Lack of Semantic Interoperability results in many problems and is at the heart of most interoperability issues. For this, I focus on a few core problems, which are listed below:

- Data Loss
- Change of Meaning
- Addition of Data Points
- Lack of Continuity

1) Data Loss: Data loss, or more accurately *semantic loss*, can occur when converting data from one format to another, in a situation where some data points are omitted from the conversion or transformation routines, very often deliberately to save time or improve efficiency, or because it is too difficult to change formal and structure of the *target* system. Arguably this isn't always a direct problem, but very often emerges when the dataset is re-evaluated or needs to be fused at some point in the future. To illustrate this problem, I can consider the issue of transgender patients. Let's take clinical trial data from a physical form that has been completed by patients taking a clinical trial, where the data is written in long-hand. I may have, for instance, classified a person's gender using five check-boxes denoting male, female, phenotype-male, phenotype-female, or not known. The idea is that people who are transgender and have moved from being male at birth to being female at some point will tick both female and phenotype male. This data is then manually imported into a web-based form, which has three fields: male, female, or unknown - someone looks at the form and clicks boxes on a web form. Clearly the intent will be to preserve as much of the data as possible, however there will need to be rules made to deal with ambiguities. I may, therefore, lose the information given by ticking both female and male-phenotype (indicating a transgender person); if the data pertaining to that person is then used in, say, studies of the female reproductive system, the results may be biased. Obviously, this depends on how many studies are used, how many people are in the study of transgender origin, and what rules the person who is making the manual import follows when actually carrying out the data import, do they assign the transgender person to *female* or *unknown*, and is this assignment recorded on the dataset as relevant metadata for future reference.

2) Change of Meaning: Change of meaning is the problem that is very often related to data loss, in that I have an item, let's say height, which is referenced in a dataset, and referenced to a particular measurement unit, let's say *meters*, however, some detail such has been left out. For most clinical studies, height is measured by a nurse, with patient removing their shoes before the measurement for the simple reason that this removes any potential error in people either exaggerating or not knowing their height. The height data at the start of a trial is assumed to be measured in the same way for the entire period of the trial; if not, then perhaps errors will creep into the results, such as this particular drug causes height reduction or increase as a side effect. Such an issue is unlikely to emerge in one trial taking place in a single hospital by the same team. Still, errors could creep in with trials taking place by many teams in many different hospitals and departments.

Change of meaning should be solved by reference to some kind of dictionary and perhaps some kind of reference point for the units of measurement used; one is unlikely to make a mistake with meters as in this example, but with drug calculations, one will need more precision. Drips are very often specified in units, such as 5000 units of heparin, but a nurse will need to convert that to perhaps units/hour, mcg/kg/minute, or mg/hour and will need to know the patient's weight. It may be that just the unit amount is specified on the data collection form, but the form doesn't say how fast the infusion was given. Five thousand units taken in 1 hour as opposed to over 2 hours

may have a significant impact on the study and may indeed mean that two different trials recording ostensibly the same data are, in fact, recording two entirely different things.

3) Addition of Data Points: Clinical studies evolve, and while one set of questions is asked at the start of a trial, some trials continue for many years, and as a result, different questions start to be asked at later stages. In particular, data points are added to datasets, if I look at the previous examples, I could easily record the dosage of a drug at the beginning of a trial without mentioning how fast it delivered, and then I can add a data point toward the end which does record the speed of delivery. The dataset has changed, but how should the data be treated when it is fused with another study that maybe does or does not include speed of delivery?

4) Lack of Continuity: Clinical Studies also face the problem that data may be available from paper or electronic recording means (excel or CSV files), and there is no real recording of the exact set of expected answers at the time of the study. For instance, a data item such as the *Basis of the Cancer* may be recorded from a list of *enumerations*, which are currently listed as :

- 0 Death Certificate: The only information available is from a death certificate
- 1 Clinical: Diagnosis made before death but without the benefit of any of the following (2-7)
- 2 Clinical Investigation: Includes all diagnostic techniques (e.g., X-rays, endoscopy, imaging, ultrasound,
- 4 Specific tumor markers: Includes biochemical and/or immunological markers that are specific for a tumor site
- 5 Cytology: Examination of cells, whether from a primary or secondary site, including fluids aspirated using endoscopes or needles. Also includes microscopic examination of peripheral blood films and trephine bone marrow aspirates
- 6 Histology of a metastasis: Histological examination of tissues from a metastasis, including autopsy specimens
- 7 Histology of a primary tumour: Histological examination of tissue from the primary tumour, however obtained, including all cutting and bone marrow biopsies. Also includes autopsy specimens of a primary tumor
- 9 Unknown: No information on how the diagnosis has been made (e.g., PAS or HISS record only)

The problem occurs when researchers come across significant findings that record the *Basis of the Cancer* as 3 and 8.

III. STATE OF THE ART

A. ISO/IEC 11179

This study builds upon the foundational work conducted in the CancerGrid project (0), where an ISO 11179-compliant metadata registry was designed to curate semantic metadata and enable model-driven generation of trial-specific software (0; **?**). The CancerGrid approach to form generation has been significantly expanded through the introduction of a data modeling language and a more comprehensive framework for semantic linking.

Another notable implementation of ISO 11179 can be seen in the US caBIG initiative (0), which developed the caCORE software development kit (0). However, while caCORE leverages model-driven development for generating web service stubs, it still requires manual coding for application logic. In contrast, the method proposed in this study seamlessly integrates with established clinical Electronic Data Capture (EDC) systems and workflows, such as OpenClinica (0).

More recently, an ISO 11179-compliant metadata registry (0) has been employed to facilitate data mapping for the North German Tumor Bank of Colorectal Cancer. Clinical and sample data, structured within a harmonized dataset, are systematically collected and consolidated via a hospital-integrated Research Data Management System that supports biobank and study management.

B. Ontology Modeling

Several research efforts have focused on ontological representations of ISO 11179 to enhance data integration across multiple metadata registries (MDRs). Sinaci and Erturkmen (0) propose a federated semantic metadata registry framework that represents Common Data Elements (CDEs) as Linked Open Data resources. These CDEs are structured using the Resource Description Framework (RDF), allowing them to be queried and interconnected with similar elements in other registries via the W3C Simple Knowledge Organization System (SKOS). The framework incorporates an ISO 11179 ontology and is implemented using the Jena framework.

Jeong *et al.* (0) introduces the Clinical Data Element Ontology (CDEO) to facilitate unified indexing and retrieval of metadata elements across MDRs, employing SKOS to structure and define CDEO concepts. Similarly, Tao *et al.* (0) present case studies demonstrating how HL7 Detailed Clinical Models (DCMs) and ISO 11179 models can be represented using the Web Ontology Language (OWL). Their approach involves extracting metamodels from UML diagrams and Excel spreadsheets, though it relies on manual transformation to generate the ontological representations in OWL.

Leroux *et al.* (0) explore the use of existing ontologies to enhance OpenClinica forms. Their Model Catalogue approach enables the integration of ontologies within an ISO 11179-compliant metadata registry, leveraging a model-driven methodology to capture standardized data effectively.

Ontology repositories function similarly to model catalogs, providing a structured environment for storing, interlinking, querying, versioning, and visualizing ontologies. These repositories also document relationships between ontologies, facilitating navigation and alignment. Linked Open Vocabularies (LOV) (0) serves as a discovery platform for linked data vocabularies and ontologies, offering automated documentation extraction and dependency identification. Apache Stanbol (0) provides modular components for semantic content management, including an Ontology Manager that supports ontology versioning and integration within controlled environments.

C. Data Warehousing

In data warehousing (0), metadata plays a crucial role in aggregating information from various business systems into a centralized data warehouse, enabling decision support and business performance analysis. These data warehouse models can follow either a normalized relational structure (0) or a dimensional structure (0), where data is categorized into quantifiable *facts* and contextual *dimensions*.

The Common Warehouse Metamodel (CWM) (0), developed by the Object Management Group, provides a UML-based framework for implementing data warehousing solutions. Unlike traditional data warehousing, which primarily relies on rigid, write-once models, the ISO 11179 approach supports continuous evolution and adaptability of metadata models. The core CWM metamodel shares conceptual similarities with the *concept* and *value* elements within the ISO 11179 framework, demonstrating the potential for interoperability between these methodologies.

D. Model-driven engineering for e-Health

Several examples of model-driven engineering for e-Health software are reported in the literature (0; 0; 0; 0; 0). Payne (0) formalizes the typical pattern followed in these methods: a multi-phase approach, where data modeling is a separate phase from stakeholder engagement and data integration. This approach is taken Khambati *et al.* (0), where an Eclipse-based tool is used to develop domain-specific languages to model and generate tools for mobile health-tracking applications. The advantage of the approach is the ability it provides for clinicians to modify the model of the study, which is specified in the DSL, and automatically regenerate the application from the model. A similar approach is taken in the *True Colours* system (0), using the Booster model-driven toolkit to derive a patient self-monitoring application for mental health. The Booster approach demonstrates the lesson that data tends to be managed better within a model-driven process, leading to higher quality and more reusable assets. Schlieter *et al.* (0) record their experience gained from using model-driven engineering to implement an application for path-based stroke care. Among the lessons learned, they recommend using existing ontological models where possible and being prepared to reconcile a heterogeneity of models from the various stakeholders under a common metamodel. In contrast to these systems, our metadata-oriented approach supports the creation of applications that can interoperate with existing data, standards, and systems. Rather than simply using MDE to develop stand-alone systems, MDE processes are used in the management of clinical trials metadata from which software is derived.

In the Model Driven Health Tools (MDHT) (0) project, the HL7 Clinical Document Architecture (CDA) standard (0) for managing patient records is implemented using Eclipse UML tools (0). The benefits of applying MDE are clear: modeling tools are usesd to model the CDA standards and interoperable implementations of the standard are automatically derived from the models. The CDA standards are large and complex: Scott and Worden (0) advocate a model-driven approach to simplify the HL7 CDA, supported by three case studies: the NHS England 'Interoperability Toolkit', simplification of US CDA documents, and the Common Assessment Framework project for health and care providers in England.

E. Electronic data capture

A range of tools are available for clinical Electronic Data Capture (EDC), including Catalyst Web Tools (0), OpenClinica (0), REDCap (0), LabKey (0) and Caisis (0). Franklin *et al.* (0) present a two-year case study comparing EDC tools, and Leroux *et al.* (0) report on a further study comparing Clinical Trials Management Systems. I use the Model Catalogue tool to generate case report forms for OpenClinica, but in principle, any EDC tool could be supported.

The state of the art in EDC is represented by the Research Electronic Data Capture (REDCap) (0) web-based application, which supports metadata capture for research studies, providing an online interface for data entry, audit trails, and export to common statistical packages and data import from external sources. Like the Model Catalog, the REDCap system focuses on the clinical metadata. However, REDCap and similar EDC tools are typically insular systems, importing any data into a centralized data silo; in contrast, the Model Catalog aims to provide a platform to support and integrate existing data stores and systems within the clinical environment.

There is also a distinction in the level of expertise expected to operate the tools. The metadata management — creation, revision, sharing — in EDC is typically considered an IT-specialist task (0; 0), requiring experts to initialize the metadata separately for each study. With the Model Catalog clinical domain specialists have the ability to adapt and modify metadata as needed and treat models as the central artifacts. Effectively, the Model Catalogue follows the same metadata workflow as REDCap but without the need for modeling experts to develop, adapt, or share the metadata models.

IV. METHODOLOGY

I started working in this project with a ISO11179 conformant metadata registry, which was used to achieve interoperability between different datasets, however although that objective was achieved the work was not entirely successful due to some inconsistencies in the way the standard is defined. The original experiments with linking data did achieve some success in establishing the use of linking at a higher level of abstraction, however since the meta-model was not formally defined it was very difficult to reason over that model. Therefore, one of the first stages of this work was to define a formal metamodel.

A metamodel, such as the kind defined in Ecore or UML using MOF, provides a basis for translation and code generation using a wide range of model-driven tools, such as ATL or XText. The core model, simply referred to as the metamodelling language (MML) was specified using Z, informed by the results of the original ISO11179 trial usage in the NHIC project. This gave a formal basis to further developments, a grammar was developed using XText which enabled datasets

to be defined using MML, and allowed access to the Eclipse toolkit to compose, construct and modify such datasets.

Datasets are defined in a number of ways, some such as SNOMED are provided as an OWL ontology, or to be more precise provided as a text file which can be used with a PERL program to re-construct an OWL ontology. It is also distributed as a relational database, a set of DLL scripts that can then be imported into a relational data store such as Postgres or MySQL. Other datasets are defined using either XSD files, Excel files, or even a combination of text and CSV files. Data from the Human Phenotype Ontology is available in OWL or OBO format. I have built importers to import data from Excel using VBA scripts, from XSD using Groovy/java XML libraries, and from OBO files using Groovy scripts. Some work was undertaken to import data directly from OWL but it is not complete at this point in time.

Having imported several medical dataset definitions, the task is to show that using a meta-model to carry out semantic integration is more efficient than simply relying on extract transform load (ETL) technology. I have done this by taking sample data for cancer patients as defined by three different models: HPO, COSD, and FHIR, each one in a separate relational database. I have then detailed the process needed to map the three schemas and run simple queries over them. I have then set up a similar setup using a metadata registry, but by mapping the meta-models to a central metamodel, the ETL integration process time is reduced, the querying time is reduced, and so is the incremental adjustment work in maintaining the metadata registry setup over the ETL setup.

V. CONCLUSION AND FUTURE WORK

In this work, I explored the challenges of clinical data management and interoperability, particularly in the context of using ISO/IEC 11179-compliant metadata registries to integrate heterogeneous datasets. The study highlights the inefficiencies associated with traditional extract, transform, load (ETL) processes and demonstrates how leveraging a metamodel-based approach improves data harmonization, reduces query time, and minimizes maintenance efforts. By integrating various clinical datasets such as HPO, COSD, and FHIR through a centralized metamodel, I achieved more efficient semantic data integration.

Our findings underscore the importance of a well-defined metamodel in achieving interoperability across diverse clinical and research datasets. The use of formal metamodeling techniques, such as Ecore and UML with MOF, enables structured data management and facilitates automated code generation. Additionally, the methodology shows that using metadata registries allows for greater flexibility and adaptability when integrating new datasets compared to traditional ETL pipelines.

Despite the success of the approach, several limitations remain. The process of importing datasets from diverse formats (e.g., OWL, XSD, Excel, and OBO) still requires significant manual effort, and the completeness of some data integration tasks remains a challenge. Additionally, while I have demonstrated improvements in query performance, further benchmarking against large-scale clinical datasets is needed to quantify the advantages more precisely.

Future work will focus on the following key areas:

- **Automation of Data Importation:** Developing automated tools for converting datasets from various formats into a standardized metadata registry representation to further reduce manual preprocessing efforts.
- **Expansion of the Metamodel:** Refining the metamodel to accommodate a broader range of medical ontologies and datasets, thereby increasing interoperability across different healthcare domains.
- **Machine Learning for Data Mapping:** Exploring the use of machine learning techniques to automate schema mapping and data transformation, reducing human intervention in integrating new datasets.
- **Scalability and Performance Testing:** Conducting large-scale evaluations with real-world clinical databases to further validate the efficiency gains of the metadata-driven approach over ETL processes.
- **Integration with Clinical Workflows:** Investigating the practical application of metadata registries within hospital and research environments to streamline electronic data capture (EDC) and clinical trial management.

By addressing these areas, I aim to further enhance the robustness, usability, and scalability of metadata-driven clinical data integration, ultimately contributing to improved data management practices in biomedical research and healthcare systems.

REFERENCES

I. S. Organisation, "International standard for metadata registries." http://metadata-standards.org/11179/.

J. Davies, J. Gibbons, S. Harris, and C. Crichton, "The CancerGrid experience: Metadata-based model-driven engineering for clinical trials," *Science of Computer Programming*, vol. 89, pp. 126–143, 2014.

J. Davies, J. Gibbons, R. Calinescu, C. Crichton, S. Harris, and A. Tsui, "Form follows function: Model-driven engineering for clinical trials," in *Foundations of Health Informatics Engineering and Systems*, pp. 21–38, Springer, 2012.

I. Kunz, M.-C. Lin, and L. Frey, "Metadata mapping and reuse in caBIG," *BMC Bioinformatics*, vol. 10, no. Suppl 2, p. S4, 2009.

G. A. Komatsoulis, D. B. Warzel, F. W. Hartel, K. Shanbhag, R. Chilukuri, G. Fragoso, S. de Coronado, D. M. Reeves, J. B. Hadfield, C. Ludet, *et al.*, "caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability," *Journal of Biomedical Informatics*, vol. 41, no. 1, pp. 106–123, 2008.

"OpenClinica."

Ulrich, K. AK, D.-H. P, H. JK, and I. J, "Metadata repository for improved data sharing and reuse based on hl7 fhir," *Studies in Health Technology and Informatics*, vol. 228, pp. 162–166, 2016.

A. A. Sinaci and G. B. L. Erturkmen, "A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains," *Journal of Biomedical Informatics*, vol. 46, no. 5, pp. 784 – 794, 2013.

S. Jeong, H. H. Kim, Y. R. Park, and J. H. Kim, "Clinical Data Element Ontology for Unified Indexing and Retrieval of Data Elements across Multiple Metadata Registries," *Healthcare Informatics Research*, vol. 20, pp. 295–303, Oct 2014.

C. Tao, G. Jiang, W. Wei, H. R. Solbrig, and C. G. Chute, "Towards Semantic-Web Based Representation and Harmonization of Standard Meta-data Models for Clinical Studies," *AMIA Summits on Translational Science Proceedings*, vol. 2011, pp. 59–63, 2011.

H. Leroux, S. McBride, L. Lefort, M. Kemp, and S. Gibson, "A method for the semantic enrichment of clinical trial data," in *Health Informatics: Building a Healthcare Future Through Trusted Information; Selected Papers from the 20th Australian National Health Informatics Conference (HIC 2012)*, vol. 178, p. 111, IOS Press, 2012.

"Linked Open Vocabularies."

"Apache Stanbol."

R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. New York, NY, USA: John Wiley & Sons, Inc., 2nd ed., 2002.

W. H. Inmon, *Building the Data Warehouse*. Wellesley, MA, USA: QED Information Sciences, Inc., 1992.

J. Poole, D. Chang, D. Tolbert, and D. Mellor, *Common Warehouse Metamodel Developer's Guide*, vol. 24. John Wiley & Sons, 2003.

J. Davies, J. Gibbons, J. Welch, and E. Crichton, "Model-driven engineering of information systems: 10 years and 1000 versions," *Science of Computer Programming*, vol. 89B, pp. 88–104, September 2014.

W. Raghupathi and A. Umar, "Exploring a model-driven architecture (MDA) approach to health care information systems development," *International Journal of Medical Informatics*, vol. 77, no. 5, pp. 305–314, 2008.

B. Blobel and P. Pharow, "A model driven approach for the german health telematics architectural framework and security infrastructure," *International Journal of Medical Informatics*, vol. 76, no. 2, pp. 169–175, 2007.

A. Khambati, J. Grundy, J. Warren, and J. Hosking, "Model-driven development of mobile personal health care applications," in *Proceedings of the 2008 23rd IEEE/ACM International Conference on Automated Software Engineering*, pp. 467–470, IEEE Computer Society, 2008.

H. Schlieter, M. Burwitz, O. Schönherr, and M. Benedict, "Towards model driven architecture in health care information system development," in *12th International Conference on Wirtschaftsinformatik (WI 2015)*, March 2015.

P. R. Payne, "Biomedical knowledge integration," *PLoS Computational Biology*, vol. 8, no. 12, 2012.

"Model driven health tools."

R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. S. Shvo, "HL7 Clinical Document Architecture, release 2," *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 30–39, 2006.

"Eclipse MDT UML2 tools."

P. Scott and R. Worden, "Semantic mapping to simplify deployment of HL7 v3 Clinical Document Architecture," *Journal of Biomedical Informatics*, vol. 45, no. 4, pp. 697–702, 2012.

University of Washington, "Catalyst web tools."

P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 377–381, 2009.

E. K. Nelson, B. Piehler, J. Eckels, A. Rauch, M. Bellew, P. Hussey, S. Ramsay, C. Nathe, K. Lum, K. Krouse, *et al.*, "LabKey Server: An open source platform for scientific data integration, analysis and collaboration," *BMC Bioinformatics*, vol. 12, no. 1, p. 71, 2011.

P. Fearn and F. Sculli, "The CAISIS research data system," in *Biomedical Informatics for Cancer Research*, pp. 215–225, Springer, 2010.

J. D. Franklin, A. Guidry, and J. F. Brinkley, "A partnership approach for electronic data capture in small-scale clinical trials," *Journal of Biomedical Informatics*, vol. 44, pp. S103–S108, 2011.

H. Leroux, S. McBride, and S. Gibson, "On selecting a clinical trial management system for large scale, multi-centre, multi-modal clinical research study," *Studies in Health Technology and Informatics*, vol. 168, pp. 89–95, 2011.