BIRTH OF IDENTITY: UNDERSTANDING THE VALUE AND POLICY CONSIDERATIONS OF USING BIRTH CERTIFICATES FOR IDENTITY RESOLUTION

by

Jeffrey Dean Duncan

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biomedical Informatics

The University of Utah

December 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Jeffrey Dean Duncan**

has been approved by the following supervisory committee members:

Catherine Staes	, Chair	**8/04/2015** Date Approved
Sidney Thornton	, Member	**8/04/2015** Date Approved
Karen Eilbeck	, Member	**8/04/2015** Date Approved
Stephen Clyde	, Member	**8/04/2015** Date Approved
Scott Narus	, Member	**8/04/2015** Date Approved

and by **Wendy Chapman** , Chair of

the Department of **Biomedical Informatics**

and by David B. Kieda, Dean of The Graduate School

ABSTRACT

Exchanging patient-specific information across heterogeneous information systems is a critical but increasingly complex and expensive challenge. Lacking a universal unique identifier for healthcare, patient records must be linked using combinations of identity attributes such as name, date of birth, and sex. A state's birth certificate registry contains demographic information that is potentially very valuable for identity resolution, but its use for that purpose presents numerous problems. The objectives of this research were to: (1) assess the frequency, extent, reasons, and types of changes on birth certificates; (2) develop and evaluate an ontology describing information used in identity resolution; and (3) use a logical framework to model identity transactions and assess the impact of policy decisions in a cross-jurisdictional master person index.

To understand birth certificate changes, we obtained de-identified datasets from the Utah birth certificate registry, including history and reasons for changes from 2000 to 2012. We conducted cohort analyses, examining the number, reason, and extent of changes over time, and cross-sectional analyses to assess patterns of changes. We evaluated an ontological approach to overcome heterogeneity between systems exchanging identity information and demonstrated the use of two existing ontologies, the Simple Event Model (SEM) and the Clinical Element Model (CEM), to capture an individual's identity history. We used Discrete Event Calculus to model identity events

across domains and over time. Models were used to develop contextual rules for releasing minimal information from birth certificate registries for sensitive cases such as adoptions.

Our findings demonstrate that the mutability of birth certificates makes them a valuable resource for identity resolution, provided that changes can be captured and modeled in a usable form. An ontology can effectively model identity attributes and the events that cause them to change over time, as well as to overcome syntactic and semantic heterogeneity. Finally, we show that dynamic, contextual rules can be used to govern the flow of identity information between systems, allowing entities to link records in the most difficult cases, avoid costly human review, and avoid the threats to privacy that come from such review.

To my family

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my gratitude, at the end of this long journey, to those who helped along the way. First and foremost, Catherine Staes, PhD, the chair of my supervisory committee, who helped me from the start to focus on this research, and kept me motivated all the times I stalled or wandered off course. I'm also very grateful to the rest of the committee: Scott Narus, PhD; Stephen Clyde, PhD; Karen Eilbeck, PhD; and Sid Thornton, PhD. Each of them provided unique expertise, timely advice and continuous encouragement.

I am indebted to all my colleagues at the Utah Department of Health. In particular, I'm indebted to Wu Xu, PhD, my supervisor, who was always supportive of my efforts. I could not have completed this research without her support. I also appreciate the help of Kailah Davis, PhD, who was always willing to read a paper or serve as a sounding board for my crazy ideas.

To my wife, Wendy, who thinks by now that a laptop computer is a natural extension of my fingers, I am grateful for the patience, support, and encouragement throughout this journey.

CHAPTER 1

INTRODUCTION

In a 1946 paper, Dr. Halbert L. Dunn, chief of the U.S. Office of National Vital Statistics, used the analogy of a "Book of Life" to describe record linkage. Each person, Dunn wrote, creates a book beginning with birth and ending with death, with individual life events as pages of the book. The goal of record linkage, Dunn wrote, was to assemble the pages of an individual's book into a volume.[1]

Fast forward nearly seven decades to a world where birth certificates, death certificates, immunizations, clinic visits, healthcare encounters and imaging studies are all electronic, pages of an individual's "e-book" stored in different information systems in various healthcare and public health entities. Advances in information technology have created significant opportunities to improve care for individuals and populations through the sharing of patient-specific information. Patient-specific information sharing examples include health information exchange (HIE),[2] comparative effectiveness research (CER),[3] public health reporting,[4] Informatics for Integrating Biology and the Bedside (i2b2),[5] the cancer Biomedical Informatics Grid (caBIG)[6] and the Federated Utah Research and Translational Health eRepository (FURTHeR).[7] Each of these applications requires the ability to uniquely identify patients across heterogeneous information systems—to assemble the relevant pages in Dunn's analogous book for individuals and groups.

Dunn envisioned a unique number assigned at birth that could be used to identify and link life events from birth to death into a longitudinal record. While Dunn's vision of a unique identifier assigned at birth has never been realized, his concept of a longitudinal record beginning at birth and linking clinical and public health information across healthcare domains is increasingly important to biomedical research and healthcare. In contrast to Dunn's time, today's birth and death registries are maintained by state public health departments using electronic birth registration systems (EBRS)[8] and electronic death registration systems (EDRS),[9] with information submitted by healthcare providers. These registries are potentially very valuable for identity resolution in healthcare, but their use for that purpose presents unique problems and challenges, including quality, heterogeneity, and policy considerations.

Birth certificates in the United States serve two distinct and sometimes incongruous roles: (1) they are an identity record used to establish future identification, and (2) they are a primary source of public health data for maternal and newborn child health. Birth certificates are typically submitted by hospital medical records staff reporting information contained on two worksheets that correspond to these two roles. The parental worksheet contains demographic information that establishes, among other things, the name of the child and the names and birthplaces of parents. A medical worksheet contains information abstracted from the mother's prenatal and hospital records as well as the newborn baby's medical record. These data are used to compile statewide and national data on maternal and child health that are used for public policy, epidemiology, and research. Both worksheets are based on the 2003 National Standard Birth Certificate.[10]

The goal of this dissertation research was to investigate identity resolution in healthcare and public health, with a particular focus on the role, value, quality, and barriers to the use of birth certificate identity information. We began with the following hypotheses: 1) Identity information included in a birth certificate is dynamic and can be used to improve identity resolution in healthcare, particularly for children; 2) An ontology could be used to represent changes in identities over time; and 3) A logical formalism such as Event Calculus could be used to reason about policy decisions and their effect on exchanging identity information between healthcare and public health for identity resolution.

Chapter 2 provides a background of identity resolution challenges, the current state of research in the field, and problems specific to using birth certificate information for identity resolution. Chapter 3 presents a heuristic analysis of privacy concerns regarding the sharing of birth certificate information. Chapter 4 presents an evaluation of birth certificate data quality, focusing on specific events that change identities recorded on birth certificates. Chapter 5 describes the development and evaluation of an ontology for identity resolution and its potential to overcome issues of semantic and syntactic heterogeneity in implementing cross-enterprise identity resolution. Chapter 6 describes the use of Event Calculus, based on first-order predicate calculus, to model identity events and policy decisions, and their impact on identity resolution. Chapter 7 summarizes the findings of this project and presents a roadmap of potential future research in this subject area.

References

1. Dunn HL. Record Linkage. *Am J Public Health Nations Health*. 1946;36(12):1412-6.

2. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011;4:47-55.

3. Brixner DI, Watkins JB. Can CER Be an Effective Tool for Change in the Development and Assessment of New Drugs and Technologies? *J Manag Care Pharm*. 2012;18(5 Supp A):S06.

4. Nangle B, Xu W, Sundwall DN. Mission-driven priorities: public health in health information exchange. *Proc AMIA Symp*. 2009;2009:468.

5. Natter MD, Quan J, Ortiz DM, Bousvaros A, Ilowite NT, Inman CJ, et al. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. *J Am Med Inform Assoc*. 2013;20(1):172-9.

6. Wang H, Yatawara M, Huang SC, Dudley K, Szekely C, Holden S, et al. The integrated proactive surveillance system for prostate cancer. *Open Med Inform J*. 2012;6:1-8.

7. Livne OE, Schultz ND, Narus SP. Federated querying architecture with clinical & translational health IT application. *Med Syst*. 2011;35(5):1211-24.

8. Brumberg HL, Dozor D, Golombek SG. History of the birth certificate: from inception to the future of electronic data. *J Perinatol*. 2012;32(6):407-11.

9. Davis K, Staes C, Duncan J, Igo S, Facelli JC. Identification of pneumonia and influenza deaths using the Death Certificate Pipeline. *BMC Med Inform Decis Mak*. 2012;12:37.

10. NVSS - National Vital Statistics System Homepage: National Center for Health Statistics; 2013 [cited 3/15/2015]. Available from: http://www.cdc.gov/nchs/nvss.htm.

CHAPTER 2

BACKGROUND

Record linkage

The value of linking records across administrative domains predates the computer era: Dr. Halbert Dunn first published on the value of record linkage in 1946.[1] In 1959, a geneticist named Howard Newcombe introduced the use of odds ratios and the use of computers to perform what he termed, "automatic record linkage."[2] Perhaps the most famous early work in record linkage was published by Ivan Fellegi and Alan Sunter in 1969.[3] Their eponymous methodology, known as the Fellegi-Sunter method, forms the foundation for many probabilistic record linkage technologies in use today. Much of the current research in record linkage focuses on improving match scores, blocking, and advanced statistical estimates of link parameters.[4-7]

Within single healthcare facilities, record linking and matching are routinely performed to identify duplicate patient records and create a database of unique identities known as a Master Patient Index (MPI).[8] The use of an MPI to uniquely identify patients has been shown to improve continuity of care, decrease medical errors, reduce unnecessary procedures and reduce costs.[9]

The growth of integrated delivery systems (networks of providers and facilities providing a continuum of care under a single administrative umbrella) has paralleled the advances in distributed information systems and information exchange. As information systems have matured with standardized interfaces to enable the sharing of data across providers and facilities within health networks, so too has the need to uniquely identify patients beyond facility boundaries. Healthcare networks have responded to this need by integrating facility MPIs into a single Enterprise Master Person Index (EMPI).[10] As a unique identifier across an entire integrated delivery system, EMPIs accrue the same benefits as MPIs, but on a much larger scale.

While the technical challenges of data linkage and real-time data exchange have received considerable research,[2, 3, 8-12] there remain significant concerns about the ability of information systems to protect the privacy, security and confidentiality of protected health information.[13]

Birth certificates as a potential authoritative source

Despite advances in linkage methodologies, there remains a residual of identities that are difficult to link and require manual review. The use of an "authoritative source" of demographic information, such as a government registry, has been proposed as a possible solution for these difficult records.[14] Birth certificate registries, managed in state public health departments, have been seen as a particularly valuable source for name and date of birth information. Identity information from birth certificates is often used to create or de-duplicate identifiers in other public health systems such as immunization information systems (IIS)[15] and in child health integrated information systems such as Utah's Child Health Advance Record Management (CHARM).[16]

Birth certificate registries may also be very valuable for identity resolution in healthcare delivery settings, particularly for children, but their use for that purpose presents problems due to the mutability of birth certificates and concomitant policy considerations.

Problem 1: Mutability

Birth certificates reflect the facts of birth at the time of the birth event, but many of those "facts" are subject to change over time for a variety of reasons. Information on a birth certificate may be administratively corrected or amended because of errors that occur as a result of the manual processes used to report birth certificate information. These errors may include misspellings of names, or incorrectly recording other fields such as date of birth or sex. Paternity establishments, two-parent adoptions, and stepchild adoptions can result in changes to names on birth certificates for many years after the birth.[17] In 2010, according to a special report from the US Bureau of the Census, 2.3% of children of all ages were adopted and an additional 4.7% were stepchildren.[18] For birth certificates to be a valuable resource for identity resolution, the frequency, types and time distributions of changes to birth certificate data need to be understood.

Problem 2: Policy considerations

Birth and death certificates are excluded from the Health Insurance Portability and Accountability Act (HIPAA) restrictions on electronic data exchange, but their release is subject to state laws governing the release and use of vital records data. In addition, adoptions in Utah are private, meaning that when a child is adopted, the original birth certificate is sealed and can only be opened by court order. Attempts to manually link data between birth certificates and other data sources, such as immunization registries, potentially reveal adoption details in violation of privacy laws. However, excluding adoptees from identity resolution entirely can result in disjoint immunization and EHR records for these individuals. As a result, the current manual processes for resolving disjoint records results in clerks identifying preadoption identifiers, inadvertently violating privacy laws. Moreover, children who are relinquished for adoption at birth may be delivered by mothers with inadequate prenatal care and thus are at higher risk for adverse outcomes, a risk that is compounded by the inability to link pre- and post-adoption records.

Motivation for this research

Significant advances have been made enabling the real-time exchange of health information to support clinical needs, public health, and translational research. At the same time, much research has focused on improving the probabilistic and deterministic algorithms used to link identities, enabling such exchanges. Despite improvements in the accuracy of record linkage, none are able to achieve perfect sensitivity and specificity, and there is always some residual of records that cannot be linked automatically using probabilistic or deterministic methods. These records often require time-consuming and costly manual review.

According to a report issued by the U.S. Office of the National Coordinator for Health Information Technology (ONC), patient safety concerns dictate that matching algorithms be adjusted to produce duplicates rather than overlays (false positives), because wrong care could be provided based on an incorrect match.[19] The resulting possible matches require costly human resolution. The ONC report cited Intermountain Healthcare in Utah as reporting a cost of $60 per record while CurrentCare, a Rhode Island-based healthcare network, reported spending over $70,000 per year in staff costs for manual identity resolution.

This research explores the flow of identity information typical in cross-enterprise identity exchanges in early childhood, focusing specifically on the unique contributions, and limitations, of using birth certificate data for identity resolution. By understanding the strengths and limitations of birth certificate data and how policy decisions affect the flow of information in a hierarchical MPI, data stewards can make effective policy decisions in a cross-enterprise, hierarchical master person index. Operating at the highest levels of maturity and being fully standardized, information exchanges between providers, laboratories, public health, and others can potentially save the healthcare industry billions of dollars.[12]

References

1. Dunn HL. Record Linkage. *Am J Public Health Nations Health*. 1946;36(12):1412-6.

2. Newcombe HB, Kennedy JM, Axford S, James AP. Automatic Linkage of Vital Records Computers can be used to extract" follow-up" statistics of families from files of routine records. *Science*. 1959;130(3381):954-9.

3. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183-210.

4. Winkler WE, editor Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage. *Proceedings of the Section on Survey Research Methods, American Statistical Association*; 1988.

5. DuVall SL, Kerber RA, Thomas A. Extending the Fellegi–Sunter probabilistic record linkage method for approximate field comparators. *J Biomed Inform*. 2010;43(1):24-30.

6. Sadinle M, Fienberg SE. A generalized Fellegi–Sunter framework for multiple record linkage with application to homicide record systems. *J Am Stat Assoc*. 2013;108(502):385-97.

7. Sureka A, Lal S, Agarwal L, editors. Applying fellegi-sunter (fs) model for traceability link recovery between bug databases and version archives. *Software Engineering Conference (APSEC), 2011 18th Asia Pacific*; 2011: IEEE.

8. Hillestad RJ, Bigelow JH, Chaudhry B, Dreyer P, Greenberg M, Meili R, et al. Identity crisis: An examination of the costs and benefits of a unique patient identifier for the US health care system. *RAND Health*, 2008, pp 1-97 .

9. Bittle MJ, Charache P, Wassilchalk DM. Performance Improvement: Registration-Associated Patient Misidentification in an Academic Medical Center: Causes and Corrections. *Joint Commission Journal on Quality and Patient Safety*. 2007;33(1):25-33.

10. Mills ME. Linkage of patient records to support continuity of care: Issues and future directions. *Stud Health Technol Inform*. 2006;122:320-4.

11. Newcombe HB. Record linking: the design of efficient systems for linking records into individual and family histories. *Am J Hum Genet*. 1967;19(3 Pt 1):335.

12. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. *Health Aff (Millwood)*. 2005;24:W5.

13. Ancker JS, Silver M, Miller MC, Kaushal R. Consumer experience with and attitudes toward health information technology: a nationwide survey. *J Am Med Inform Assoc*. 2013;20(1):152-6.

14. Winkler WE. *Overview of Record Linkage and Current Research Directions*. Research Report Series: Statistical Research Division, U.S. Census Bureau; 2006.

15. Centers for Disease Control and Prevention. Immunization Information Systems Functional Standards 2015 [cited 2015 04/01/2015]. Available from: http://www.cdc.gov/vaccines/programs/iis/func-stds.html.

16. Hinman AR, Eichwald J, Linzer D, Saarlas KN. Integrating child health information systems. *Am J Public Health*. 2005;95(11):1923-7.

17. Duncan J, Narus SP, Clyde S, Eilbeck K, Thornton S, Staes C. Birth of identity: understanding changes to birth certificates and their value for identity resolution. *J Am Med Inform Assoc*. 2015;22(e1):e120-9.

18. Kreider RM, Lofquist DA. Adopted Children and Stepchildren: 2010. *Adoption Quarterly*. 2010;13:268-91.

19. Audacious Inquiry LLC. Patient Identification and Matching Final Report Baltimore, MD2014 [cited 2015 May 22, 2015]. Available from: http://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.

CHAPTER 3

BIRTH CERTIFICATES AND PRIVACY

Congress passed the Health Insurance Portability and Accountability Act (HIPAA) in 1996, partially in response to the need for a framework to protect the privacy of electronic medical information being exchanged in the emerging networked healthcare environment.[1] The HIPAA privacy rule, a regulation implemented by the Department of Health and Human Services in 2003, created the legal concept of Protected Health Information (PHI) as information that is individually identifiable. The privacy rule protects individuals from unauthorized releases of their personal, identifiable health information by restricting the situations in which healthcare organizations can release records that contain PHI.[2] Because identifiable health information is critically important for public health practice, the privacy rule specifically excludes releases of records that contain PHI from healthcare to public health agencies from its restrictions.[3] The effect of this exclusion is to balance an individual's right to privacy with the government's need to protect the public's health. Once records are released to a public health agency, public health authorities are not at liberty to release records containing PHI indiscriminately.

Access to information on birth certificates, which fall under HIPAA's public health exclusion, is governed by state statutes. The National Center for Health Statistics (NCHS) has periodically collaborated with state authorities to create model vital statistics statutes and regulations for use by state governments.[4, 5] The latest version of the model law and regulations was released in 2011.[6] As a result, laws governing release of birth certificates are somewhat similar across the U.S.

The Utah Vital Statistics Act[7] permits disclosure of information in vital records to those with a "direct, tangible, and legitimate interest." The statute and its associated administrative rule[8] both define what constitutes a direct, tangible, and legitimate interest. The statute gives the state registrar of vital records great latitude in determining the circumstances under which such an interest exists, and although not specifically mentioned in statute or rule, the provision of information for identity resolution in healthcare and public health has typically been determined to be permissible. Because of policy considerations, however, registrars have balked at releasing information for children who are flagged as pending adoption or whose records have been sealed by an adoption. In Chapter 6, we present an approach using contextual rules and minimal disclosures that could enable the use of birth certificates, including those for children pending or post-adoption, to be used for identity resolution. Provided that no specific details are disclosed, it may be reasonable to expect that the envisioned exchanges would be allowable under existing laws and regulations in Utah. What must be established is whether such exchanges would be ethical under existing privacy norms.

To address this question, we applied the *Contextual Integrity* (CI) framework,developed byHelen Nissenbaum, for assessing privacy issues in information exchange in emerging technologies.[9] CI does not conflate the concept of an individual's right to privacy with a right to secrecy, but rather views the right to privacy in terms of limiting flows of information to those that are appropriate, where appropriateness ultimately depends on the context of the information flow and individual expectations of privacy. In the CI framework, those privacy expectations are called *information norms*, and those norms depend not on the content of the information, but entirely on the context of an information exchange. For example, a physician may share a patient's confidential information with another physician in the course of developing a treatment plan, but not at a social function.

Nissenbaum articulated a decision heuristic to enhance the ability of the original framework to apply to emerging technologies where entrenched norms may not exist.[10] We used the contextual integrity decision heuristic, which consists of the following nine steps, to evaluate exchanges of birth certificate information to healthcare for identity resolution purposes:

1. Describe the new practice in terms of information flows.

   The proposed new practice is to allow third party healthcare provider organizations to query an enhanced birth certificate registry for qualitative information regarding the existence and match status of records. Healthcare providers will submit identifying items such as name, date of birth, and sex. Birth certificate registries will return limited qualitative information regarding the existence, match status, and adoption status of the records.

2. Identify the prevailing context and identify potential impacts from contexts nested in it.

   The prevailing context for this proposed exchange is healthcare. Healthcare providers and healthcare information exchanges have a critical need to identify duplicate records. As duplicate records may result in unnecessary care or even wrong care being provided, there are both patient safety and financial considerations.

3. Identify information subjects, senders, and recipients.

Information subjects include patients in healthcare facilities with potentially duplicated records. Healthcare providers act as senders, sending personal identifying information (PII) in queries to the state public health department. The health department receives the information and uses it to search the birth registry. The health department returns the match status and adoption status of the records to the requesting provider.

4. Identify transmission principles.

In CI, transmission principles are defined as constraints that govern the flow of information within a given context. We identified three transmission principles. When a hospital registers a birth certificate there is a transmission principle of *obligation*, because hospitals are required by law to report birth certificates for children born in their facility. Parents, in turn, are obligated to obtain birth certificates from the state for purposes such as school registration or to obtain a passport. The parental obligation includes the requirement for parents to amend or update birth certificates when the information is not correct. A second transmission principle of *reciprocity* may be considered in the case where a birth registry returns birth certificate information to the facility that originally provided that information. However, reciprocity does not necessarily apply to information that is amended after a record has been submitted to a state public health entity. Finally, the CI framework identifies a *fiduciary* transmission principle as the trust placed in an organization to safeguard private information and to use it only to the benefit, and not the harm, of the subject of the information. The fiduciary transmission principle applies to state vital statistics agencies, particularly given that state statutes do not require consent of the subject of a vital record to release information, but only that a requestor show a *direct, tangible, and legitimate* interest in the information. Therefore, under the fiduciary transmission principle, then, vital statistics agencies only need to be assured that a healthcare entity has such an interest and that a release of identity information is in the best interest of the subject of the record.

5. Locate applicable entrenched informational norms and identify significant points of departure.

As stated previously, identifiable information on birth certificates is protected by state law and limited to those who demonstrate a direct, tangible and legitimate interest in obtaining the records. De-identified information is often furnished for statistical purposes. Oftentimes, vital statistics staff must use identifiers to link birth certificate information with external sources prior to de-identifying it for statistical analysis.

Within the Utah Department of Health (UDOH), birth certificate identifiers are routinely shared with other information systems, such as the Utah Statewide Immunization Information System (USIIS) or Child Health Advanced Records Management System (CHARM), to assist with identification and deduplication of records in UDOH systems that contain child-specific public health information.

An entity external to UDOH, such as a healthcare organization, may be determined to have a direct, tangible, and legitimate interest in birth certificate information by virtue of a treatment relationship with a patient, and thus a state registrar could reasonably release information subsequent to the verification of that relationship. This determination would extend to records that are flagged as pending adoption, but legally cannot extend to records that are sealed subsequent to an adoption.

6. *Prima facie* assessment

The release of limited information such as whether a record exists or does not exist does not violate any information norms. In the scenarios described, hospitals are querying the birth registry with identity information that is already known to them and the birth registry is responding as to whether or not the records exist, and if they do exist whether they identify the same person. In addition, birth registries may release information regarding the adoption status of a record to assist the hospital with merging pre- and post-adoption identities in an automated way. Automating the manual resolution process, and thus avoiding human inspection of possible matches with records in a healthcare MPI system, may strengthen the privacy of these records.

7. Evaluation 1: Consider moral and political factors affected by the practice in question.

State vital statistics agencies have a legal and ethical responsibility to safeguard the identity information provided by citizens and their agents when registering vital events such as birth. The occurrence of a name change or paternity event, the identity of a birth parent, even a child's date of birth, are all considered sensitive and confidential. Healthcare institutions, on the other hand, have a legal and moral imperative to combine disjoint health records to improve the quality of care, reduce redundant care, and to prevent mistaken care. In this context, one may reasonably conclude that healthcare's need for identifying information constitutes a direct, tangible, and legitimate interest under the statute governing the release of birth certificate information.

8. Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context.

In the scenario described later in Chapter 6, contextual rules enable healthcare entities to obtain minimal, yet valuable, information regarding the identities of patients identified in their systems, while birth registries are able to avoid the release of specific information regarding sensitive and legally protected events.

9. Conclusions

On the basis of these findings, we conclude that a limited release of birth certificate information to healthcare EMPIs for the purpose of identity resolution does not violate information norms for the contexts we have described. Further, the ability to automatically resolve identities in sensitive situations involving adoptions may prove to enhance individual privacy.

References

1. Dwyer III SJ, Weaver AC, Hughes KK. Health Insurance Portability and Accountability Act. *Security Issues in the Digital Medical Enterprise*. 2004;72(2):9-18.

2. Centers for Disease Control and Prevention. HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services. *MMWR: Morbidity and Mortality Weekly Report*. 2003;52(Suppl. 1):1-17, 9.

3. Hodge JG. Health information privacy and public health. *The Journal of Law, Medicine & Ethics*. 2003;31(4):663-71.

4. National Center for Health Statistics. Model State Vital Statistics Act and Model State Vital Statistics Regulations. Hyattsville, MD: Dept. of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics; May, 1978.

5. National Center for Health Statistics. Model State Vital Statistics Act and Regulations, 1992 Revision. *DHHS Publication No(PHS)*. Hyattsville, MD: National Center for Health Statistics; 1995;94:1115.

6. National Center for Health Statistics. Model State Vital Statistics Act and Regulations, 2011 Revision. Hyattsville, MD: National Center for Health Statistics; 2011.

7. Utah Vital Statistics Act, Utah Code, Title 26, Chapter 2, (1995).

8. Utah Administrative Code, Rule R46-14: Copies of Data from Vital Records, R436-14 (1995).

9. Nissenbaum H. Privacy as contextual integrity. *Washington law review*. 2004;79(1).

10. Nissenbaum H. Privacy in context: Technology, policy, and the integrity of social life: Stanford University Press; 2009.

CHAPTER 4

UNDERSTANDING BIRTH CERTIFICATES AND THEIR VALUE FOR IDENTITY RESOLUTION

This is a pre-copy-editing, author-produced copy of an article accepted for publication in the Journal of the American Medical Informatics Association following peer review. The definitive publisher-authenticated version, Duncan J, Narus SP, Clyde S, Eilbeck K, Thornton S, Staes C. Birth of identity: understanding changes to birth certificates and their value for identity resolution. J Am Med Inform Assoc. 2014,is available online at: http://jamia.oxfordjournals.org/content/early/2014/12/17/amiajnl-2014-002774.long

Background and significance

In a 2013 recommendation to the US Congress, the Health Information and Management Systems Society (HIMSS) wrote, "One of the largest unresolved issues in the safe and secure electronic exchange of health information is the need for a nationwide patient data matching strategy to ensure the accurate, timely, and efficient matching of patients with their healthcare data across different systems and settings of care.[1] Despite considerable research in record linkage[2-9] and the creation of enterprise master person indices,[10-13] problems with identity resolution continue to challenge efforts to improve the delivery of quality healthcare.

With the growth of data-sharing initiatives such as health information exchange[14] and comparative effectiveness research,[15] the need to link patient records across institutions and organizations presents a growing challenge that is exacerbated by the lack of a unique national identifier for healthcare in the United States.

The relevance of birth certificates

For the 95% of births occurring in hospitals in Utah,[16] birth certificates are reported to the state by hospital medical records staff. Each birth certificate includes demographic information obtained from parents who indicate the desired name of the child and other information such as race and ethnicity. This birth certificate identity often propagates to other public health information systems such as immunization registries, early hearing detection and intervention registries, and metabolic and other newborn screening systems. While the identity information submitted at birth identifies a newborn child's name, date of birth, and sex, that information is not necessarily permanent. In its role as a civil registry, and because birth certificates are a foundational identity document, vital statistics offices routinely correct birth certificate information to reflect changes in real identity and to correct mistakes on original records. Understanding these changes and their implications for the use of birth certificates in identity resolution is the goal of this analysis.

As a fundamental source of identity information, there are at least two possible roles birth registries may play in identity resolution for healthcare and public health. First, birth registries may be used as a source in a hierarchical master person index, such as Utah's statewide master person index (MPI), incorporating clinical and public health sources.[17] Second, automated queries to birth registries may be used to facilitate the resolution of potential record matches.

While we know that birth certificate information is subject to change at any age and for multiple reasons, to date there has been no assessment of the number, frequency, reasons and age distribution of changes on birth certificates. We propose that understanding changes made to the assigned identities on birth certificates could improve record matching strategies. The goal of this project was to document and understand the frequency and types of changes to birth certificates and to assess the value of this information for improving identity resolution across healthcare and public health. As the original source for and a registry of changes to an individual's legal name, date of birth, and sex, birth certificates have great potential for identity resolution. However, to fully utilize this potential it is important to understand the stability of birth certificate information. Therefore the objectives of this analysis were to describe: (1) the frequency and cause of changes to birth certificate identifiers as children get older, and (2) the frequency of events (i.e., adoptions, paternities, and amendments) that may trigger changes and the impact of the different types of events on a name.

Methods

Study population

Utah's birth registry system, maintained by the Utah Department of Health, includes information about births in Utah since statewide registration first began in 1905. In 2000, the system started tracking changes (i.e., updates) to the information included in the registry. In 2009, the system started tracking more detailed information about adoptions and paternities. The types of change events identified on birth certificates are described in Table 4.1. Each change event may result in changes to the facts recorded on a birth certificate.

Identity classification

Three tiers of identity have been described by Durand (Figure 4.1).[18] Tier 1 is a person's *real* identity: the identity that is owned by and completely under the control of the individual or the individual's agent (e.g. parent). Tier 2 is an *assigned* identity: an identity that is created by some entity for a specific context or purpose. Tier 3 is an *aggregate* identity: an identity that is assigned based on inclusion in a group because of a specific attribute of the individual.

Real (Tier 1) identities are described by nonunique identifiers, a subset of which are recorded at a point in time to form an assigned (Tier 2) identity. For example, a person's real identity is described by identifiers such as name, sex, date of birth, address, social security number, and so on. Many of these identifiers, including name and address, are subject to change over time at the direction of the individual who owns the identity. Although rare, persons may change their gender identity and the sex recorded in a record, creating challenges for electronic health records.[19]

Tier 2 identities are records or snapshots of Tier 1 identity taken at specific points in time, for a specific purpose. A birth certificate is thus an assigned identity documenting a child's first legal name, sex, date of birth, place of birth and parentage. For most newborn children born in the United States since 1989, in a process known as Enumeration at Birth,[20] state birth certificate registrars electronically apply for social security numbers by providing identity information from the birth certificate. A driver license is an assigned identity typically created when an individual is around 16 years of age using the birth certificate and social security card as the source of identity, especially the name, date of birth, and sex.

A record in a hospital information system is a Tier 2 identity. As part of the intake process for clinical visits, a person presents identifying information that may include a driver license and an insurance card. Either a new assigned identity is created in the hospital's information system, or information from an existing record is verified and updated if necessary.

Record linkage attempts to associate records identifying the same individual in different electronic systems. De-duplication refers to linkage techniques used to identify multiple instances of the same individual in a single system. Using Durand's model as a framework for understanding, it can be said that record linkage attempts to locate and link two or more different *assigned identities* for a single *real identity*. If two or more assigned identities in the same database refer to the same real identity, they are duplicate records. If assigned identities in two or more separate systems are equivalent, they are usually assumed to refer to the same real identity. Integrated health delivery systems[21] typically use an enterprise master person index (EMPI) to continuously search for, de-duplicate, and link identities across the multiple information systems within the organization. Most individuals can be identified and linked by matching identifiers. Problems arise when assigned identities for the same real identity do not match.

When performing record linkage, two identities assigned to one real individual at different times, t1 and t2, may fail to be identical for two reasons:

Identity Problem 1: The individual's identifiers changed between t1 and t2. A person may change names, address, phone number, and other identifiers.

Identity Problem 2: Identifiers listed for the assigned identity were recorded incorrectly at t1, t2, or both. This may be caused by a data entry error, or incomplete or incorrect submission of information from the source. Data entry errors during inpatient registration are a common source of duplicate records in electronic health record systems.[22]

A third problem that confounds linkage methods, which we will call Identity Problem 3, occurs when two assigned identities that appear to belong to the same person in fact do not. This situation may occur when two or more newborns (e.g. as a result of multiple births) have the same last name, date of birth, and sex, and different but similar first names, or when a child shares a first and last name, and even a month and day of birth, with a parent.

Of the events identified in Table 4.1, adoptions and paternities reflect changes to a person's Tier 1 identity, resulting in Identity Problem 1. Amendments reflect corrections to the Tier 2 identity recorded on the birth certificate resulting from Identity Problem 2.

Records and fields abstracted for analysis

In November 2013, we abstracted information from Utah's birth registry system, which includes information about all births registered since 1905. We obtained detailed change history information, including the date of change, field changed, and reason for changes processed from January 1, 2000 to December 31, 2012. To preserve the confidentiality of persons identified in the records, we obtained information about the fact of change but not the actual pre- or post-change values. The following fields in the record were considered to be ‘identify fields’: first name, middle name, last name, date of birth, and sex.

We also obtained de-identified information about the dates and occurrence of an adoption, paternity or any amendments for two birth cohorts (1987 and 2000). This information was linked to change information using each birth certificate’s unique file number.

Cohort analysis

We analyzed two cohorts of births (babies born in 1987 and in 2000) to identify the distribution and sequence of identity change events over time. Detailed change history from the Utah birth record system was only available from 2000 onward, therefore, we reviewed records for the cohort born in 2000 to understand changes documented during the first 12 years of life, i.e., between 2000 and 2012. Because we were limited to detailed change history after 2000, we reviewed records for the cohort born in 1987 to document changes during the second 12 years of life (i.e., when the cohort born in 1987 were 13 to 25 years of age between 2000 and 2012). We conducted nonparametric univariate survival analysis using SAS software’s PROC LIFETEST, using each individual’s year of age, not calendar year, as the time variable.[23] For the survival model, we calculated time to the first change to identity fields on the birth certificate for any reason. In the event that individuals experience multiple changes to identity fields at different times, we only used the time to the first change in our analysis. Understanding multiple changes to an individual requires more complex statistical models and is beyond the scope of this analysis. Birth records are flagged upon death to limit illegal use, thus death was used as a censoring event in our analysis.

Cross-sectional analysis

We performed a cross-sectional analysis of changes made to birth certificates between 2000 to 2012 for births in any prior year to identify the frequency, distribution, and reasons for changes to identity information. We summarized the changes recorded to any existing birth certificate during each year. We aimed to understand whether the rate of change by type varied over time. We stratified the findings by age groups 0-2, 3-5, and 6 years and older for each year analyzed. Additional detail captured after upgrades to Utah's birth registry in 2009 allowed us to analyze the frequency of changes to first, middle, and last names by event type for all adoptions, paternities, and amendments to births occurring in 2010 in which a name change occurred. Table 4.2 summarizes the types of analyses, birth records used, and time periods addressed in this study.

This study was approved by the University of Utah's Institutional Review Board.

Results

Study population

A total of 2,589,265 births have been registered in Utah since 1905. The annual number of births registered in Utah has steadily increased each year from 1,406 in 1905 to 51,439 in 2012. Between 2000 and 2012, a total of 685,984 birth records were added to the Utah birth certificate registry. For the cohort analysis, we used the 35,285 births that occurred in 1987 and the 48,350 births that occurred in 2000. For the cross-sectional analysis, we used all birth records never marked as deceased for the denominator.

<u>Cohort analysis: Distribution of changes as children age</u>

During the first 13 years of life, 3,147 (6.5%) of the 48,350 children born in 2000 had changes to their birth certificate (Table 4.3). Using nonparametric univariate survival analysis, children in their first year of life faced the highest likelihood of change (5%); the probability decreased markedly after the first year to 0.2% during the 13th year of life. The second cohort born in 1987 demonstrated a similar rate of change events (0.1%) during the 13th year of life. The rates of changes remained low with the exception of increases observed for those between 15 and 19 years of age, which are the years when persons usually apply for their first driver's license. Amendments and Adoptions triggered over 90% of identity changes for nearly every age group, but the proportions varied by age (Table 4.3).

<u>Cross-sectional analysis: Frequency of change events over time</u>

Amendment and adoption events in the first two years of life caused the greatest rates of name changes on birth certificates, with changes due to amendments showing the greatest variation: from 1,110 per 100,000 births in 2001 to a high of 2,736 per 100,000 births in 2010 (Figure 4.2A). The rate of name changes due to adoptions in the first two years of life were more consistent, ranging between 350 and 450 per 100,000 births (Figure 4.2B). In contrast, the rate of name changes due to paternity acknowledgment, though relatively low for all age groups, showed considerable variation over time for all age groups (Figure 4.2C).

The likelihood and extent of a name change varied considerably depending on the type of change event. Figure 4.3 illustrates changes due to amendments, adoptions and paternities finalized in 2010 to births in any prior year. Among all records that changed, 55% (n=6459) were due to amendments, with changes most frequently occurring to middle or last name.

Of the 5,341 paternity and adoption events that were processed in 2010 to births in any prior year, 2750 (51%) were voluntary/administrative paternities (Table 4.4). Despite being the most frequently occurring event, voluntary paternities cause the fewest name changes compared to other types (Figure 4.3). Among adoption events, two-new-parent adoptions occur more frequently than other types. Step-parent adoptions most frequently result in changes to a child's last name, but not first or middle names. In contrast, nearly two thirds of two-new-parent adoptions result in changes to both first and last names and over half include changes to a child's first name (Figure 4.3).

Discussion

Our findings indicate that birth certificate identities change over time, particularly in the first years of life, but also that the patterns of change fluctuate temporally, potentially due to societal factors. The birth record is not static and is subject to change for multiple reasons. Updated information in a birth certificate may reflect corrections or changes in Tier 1 (real) identity. This information may be useful for identity resolution in electronic healthcare and public health record systems.

The timing of changes to birth certificates may be impacted by a variety of processes. Minor typographical errors are often corrected by parents when they request a certified copy of the certificate from the state. In the case of births to unmarried mothers, most jurisdictions will subsequently add a biological father to the birth certificate only after paternity establishment, either voluntary/administrative, or court-ordered. Adoption decrees do not automatically trigger immediate changes to birth certificates in Utah, and likely in many other states that follow standard practices under the Model Vital Statistics Act. Adoption decrees are issued to adoptive parents who then must request a *supplementary* birth certificate from the state vital records office and pay a fee to have the new birth certificate issued and the original record sealed. In practice this often may happen years after the adoption takes place, for example when a child turns 16 and wants to obtain a driver license. Thus, to interpret identity changes due to adoptions, it is important to understand that the year of change due to adoption is the year that the adoptive parents present a court's decree of adoption to obtain a new birth certificate, not necessarily the year when the court issues the decree. In other words, there may be a lag of several years between the time when a court changes a child's Tier 1 identity and the time when his or her parents change the birth certificate (Tier 2) identity. In addition, child adoption proceedings in most states are 'closed', meaning that a new birth certificate is created and the original is sealed when an adoption occurs.[24] In Utah's birth registry, this policy is implemented by changing information in the electronic record, recording a change event, and severing links to the record's previous history.

Birth records are mutable, but the changes to identity fields occur for a limited number of reasons that reflect changes to a person's real identity or corrections to the assigned birth certificate identity. Changes to the date of birth or sex represented in the birth record likely reflect errors (Identify Problem 2) that are corrected after errors are noticed when birth certificates are obtained by parents or teenagers for other reasons. The slight increase in the number of changes for 5-year-olds and 16-year-olds shown in Table 4.3 are likely due to this phenomenon: birth certificates are obtained at these ages when entering elementary school or obtaining a driver license.

While the overall rate of change events to persons over 5 years of age is very low, the rate of change events for children in the first two years of life may be substantial. Our analysis also showed that the rate of events that cause identity changes fluctuates considerably for children in the first two years of life. The two peaks in name changes associated with adoptions for infants in 2001 and 2009, shown in Figure 4.2B, correspond to increases in adoptions finalized in Utah courts for the same years.[25] In general, the rate of adoptions of children in foster care in the United States is increasing,[26] and paternities are increasing due to the increase in extramarital births and Federal welfare reform laws encouraging state child support enforcement agencies to increase paternity establishment.[27] The sharp increase in name changes due to paternities reflected in Figure 4.3C can be attributed to such efforts in Utah beginning in 2004. Since paternities only result in name changes when filed after birth registration, the subsequent decline beginning in 2009 reflects the implementation of a new electronic birth registration system and a fax-to-image system that facilitated in-hospital paternity establishment prior to birth registration. Finally, the sharp increase in amendments beginning in 2009, shown in Figure 4.2A, may be attributed to hospital staff learning the new birth registry software implemented in Utah in January, 2009.

Adoptions present numerous challenges for data linkage, not only because names change, but also because those changes are often obscured by adoption privacy laws. In today's electronic exchanges, however, adoption privacy is often breeched when program staff investigate mismatched records and research potential matches, inadvertently discovering the identities of birth mothers. Our results show that the frequency and type of name change varies with the type of adoption. Currently the records of adopted children must be linked manually, often resulting in inadvertent identification of birth mothers, in violation of adoption privacy laws. Knowledge of the occurrence, type, and date of adoption or paternity, and probability of changes to name can aid in the development of automated strategies to improve identity resolution for adoptees while preserving confidentiality.

Record linkage methods can be deterministic or probabilistic.[28, 29] Regardless of which methods are used, none achieve 100% sensitivity while maintaining the 100% specificity required for medical records, meaning there is always a nonzero number of real matches that fail for any number of reasons. Many of these records fail to match due to Identity Problems 1 and 2 described above, yet the problems preventing a match could be resolved in an automated way with queries to, or updates from, a birth registry. Currently, Integrating the Healthcare Enterprise (IHE) is working to develop standards and profiles for the collection of birth certificate information from EHRs in order to automate the birth registration process.[30] Given that 95% of Utah births occur in hospitals, standards for reciprocal exchange may also make sense, except in cases such as adoptions where a record is sealed by law. Enabling hospitals to receive allowed updates when birth records change is not currently considered in the standards being developed.

A major limitation of using an electronic birth registry as a data source for identity resolution arises from the fact that any state's birth registry only includes births occurring in the state, not necessarily residents of the state. Very often large segments of any state's residents were likely born in other states or countries, given the mobile nature of today's population. Even so, migration is likely lower during the first five years of life, when changes to identity are more frequent and there are increased needs to link to the multitude of child health-related systems in public health, including registries for immunizations, newborn hearing screening, metabolic screening, and others.

This analysis has limitations. First, the findings reflect the experience in Utah, which may differ from other states. However, Utah has adopted the Model Vital Statistics Law, so the findings should be similar in other states that have also adopted the vital statistics procedures defined in the Law. Second, the data used for this study was limited to years after 2000 because of changes to Utah's vital statistics systems. Even so, the rates are likely valid due to the high number of records analyzed, and recent patterns are more relevant than previous patterns for addressing current identity management issues.

From a practical standpoint, these findings can influence practice in two ways. First, practitioners of data linkage in public health who currently use birth certificate data, such as immunization registries, can use knowledge of the age-dependent quality of birth identifiers to adjust blocking strategies and weight calculations. Second, birth certificate data is a potentially invaluable resource for informing identity resolution in healthcare and health information exchange settings, particularly in situations that involve identity changes due to adoptions. As societal trends such as gestational surrogacy and same-sex marriage may further contribute to fluctuations in patterns of identity change, more research will be needed to develop policies and technologies so that birth certificate information may be used to inform identity resolution while protecting the sensitive information of people identified.

Conclusions

Birth certificate identities change over time, particularly in the first years of life, but also patterns of change fluctuate temporally due to societal factors. The fact that changes to a birth certificate are overwhelmingly tied to changes in a person's real identity enhances the value of birth certificates for identity resolution in healthcare and public health information systems. Currently, system users struggle to link records that represent distinct snapshots of identity over time. Understanding the timing, frequency and scope of these changes is an important first step in incorporating birth registries into data linkage strategies in healthcare and public health.

Table 4.1 Descriptions of change events that impact birth certificates

Event	Description	Effect on birth certificate
Amendment		
Amendment	A change to correct minor errors or omissions on birth certificates. An amendment requires a signed affidavit and may require documentary evidence.	Identity information is changed and amendment histories are documented on printed birth certificates.
Adoption		
Two-new-parent adoption	A court awards parental rights to two new parents, neither of which is a biological parent of a child.	The original birth certificate is sealed and a new certificate is issued reflecting the names of the adoptive parents.
Stepparent adoption	A court awards parental rights to a stepparent.	The name of the stepparent is entered on the birth certificate, replacing a biological parent. Original certificate is sealed.
Family adoption	A court awards parental rights to a family member such as an older sibling, aunt, uncle, etc.	The name of the adoptive family member and spouse (if applicable) replace the names of the biological parents. Original certificate is sealed.
Single parent adoption	A court awards parental right to a single person, either male or female.	The birth certificate is amended with the name of the single parent as father or mother, as appropriate, and the original certificate is sealed.
Paternity		
Court-ordered paternity	A court determines biological fatherhood and orders a male's name entered as father on a birth certificate.	The father is listed as the court order decrees. The decree may also change the name of the child
Voluntary/ Administrative Declaration of Paternity	A male voluntarily acknowledges paternity of a child or is administratively determined to be the father by the state's child support enforcement agency.	The child's name on the birth certificate may change at the discretion of the parents.

Table 4.2. Summary of analyses and data used in this study

Analysis strategy (Results)	Birth certificates in Utah's birth registry included in analysis	Years analyzed for changes to a birth record	Purpose
Cohort (Table 4.3)	All births in 1987 and 2000	2000-2012	To describe the frequency of changes in birth certificate data as children grow older.
Cross-sectional (Figure 4.2)	All births in any year prior to the year of analysis that were never marked as deceased	2000-2012	To describe the patterns of changes to birth certificates between 2000 and 2012 for all births with no record of having died in in any prior year to the year of analysis.
Cross-sectional (Figure 4.3)	All births prior to 2010 that were never marked as deceased.	2010	To analyze changes related to adoptions and paternities enabled by changes to Utah's birth registry implemented in 2009.

Table 4.3 Results of nonparametric survival analysis to describe identity change events* that occurred between 2000-2012 for two birth cohorts.

Age in years (Lower, Upper)	Number of changes to the following fields:			Number of records censored due to death (d)	Effective Sample Size**	Conditional Probability of Change (p)	proportion of reasons for changes observed		
	Name	DOB	Sex				Amendments (%)	Paternity (%)	Adoptions (%)
Cohort born during 2000, with changes recorded January 1, 2000, through December 31, 2012									
0 - <1	1831	0	2	264	48350	0.05	88%	10%	2%
1 - <2	346	0	3	18	45818	0.014	87%	11%	2%
2 - <3	159	4	2	14	46146	0.007	74%	20%	6%
3 - <4	131	2	2	11	45907	0.006	53%	11%	35%
4 - <5	135	0	5	11	45690	0.005	46%	8%	46%
5 - <6	207	3	4	6	45493	0.006	43%	12%	45%
6 - <7	105	0	0	9	45236	0.004	30%	3%	67%
7 - <8	98	0	0	8	45081	0.003	21%	3%	76%
8 - <9	54	0	3	7	44972	0.003	33%	5%	61%
9 - <10	18	1	2	5	44846	0.003	95%	5%	0%
10 - <11	17	1	0	8	44748	0.002	94%	6%	0%
11 - <12	10	2	0	2	44660	0.003	80%	20%	0%
12 - <13	16	0	1	2	44558	0.002	94%	6%	0%
Cohort born in 1987, with changes recorded January 1, 2000, through December 31, 2012									
12 - <13	52	0	0	4	36301	0.001	94%	2%	4%
13 - <14	39	0	0	9	36243	0.001	97%	0%	3%
14 - <15	48	0	0	11	36194	0.001	92%	2%	6%
15 - <16	69	0	0	15	36133	0.002	83%	1%	16%
16 - <17	106	1	0	20	36046	0.003	75%	2%	23%
17 - <18	53	1	0	15	35922	0.002	54%	2%	44%
18 - <19	64	1	0	21	35850	0.002	69%	0%	31%
19 - <20	45	0	0	26	35761	0.001	67%	0%	33%
20 - <21	34	1	0	27	35690	0.001	74%	3%	23%
21 - <22	18	1	0	24	35629	0.001	74%	0%	26%
22 - <23	4	0	0	20	35588	0.000	100%	0%	0%
23 - <24	11	0	0	27	35561	0.000	100%	0%	0%
24 - <25	13	0	1	26	35523	0.000	100%	0%	0%

* Records were censored after the first change event or when death was recorded.
**Effective sample size is the number of persons entering each interval minus half the number censored during the interval.

Table 4.4 Number and percentage of adoption and paternity events processed in Utah in 2010 for births in any prior year.

Event	Number	Percentage of events
Two-new-parent adoption	1069	20%
Stepparent adoption	845	16%
Family adoption	263	5%
Other adoption*	203	4%
Voluntary/administrative paternity	2750	51%
Court order paternity	211	4%
Total	5341	100%

*Other adoption includes legitimations, single-parent adoptions, and foreign adoptions.

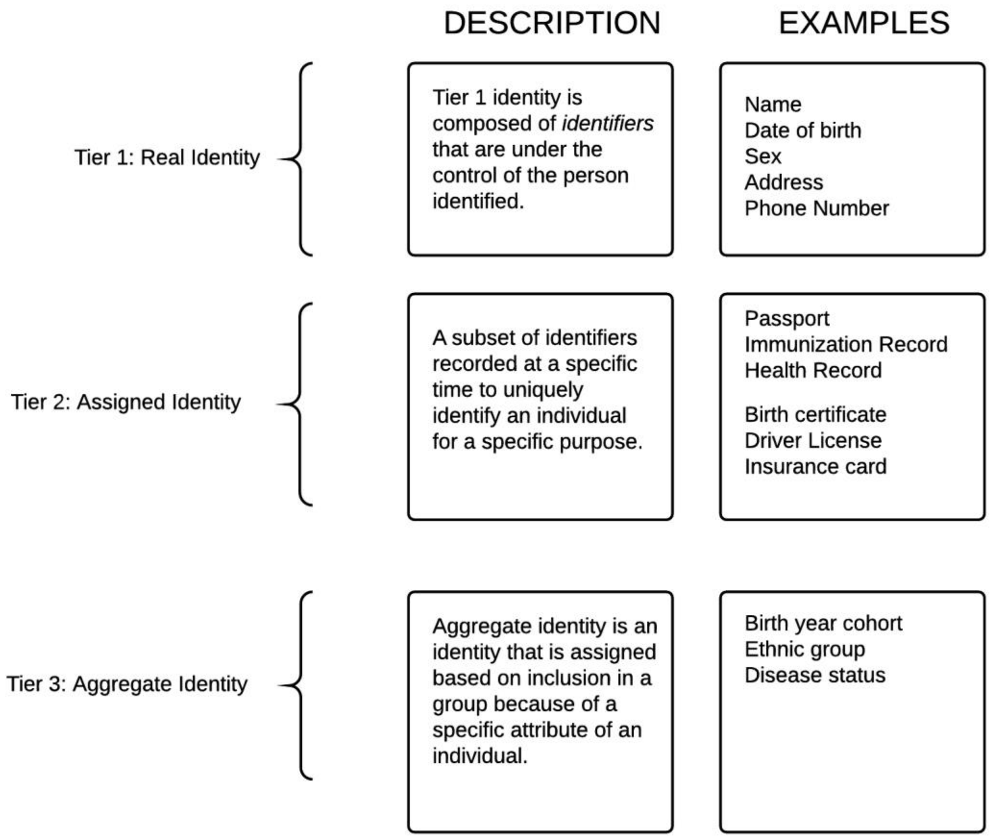

Figure 4.1 Three tiers of identity with examples

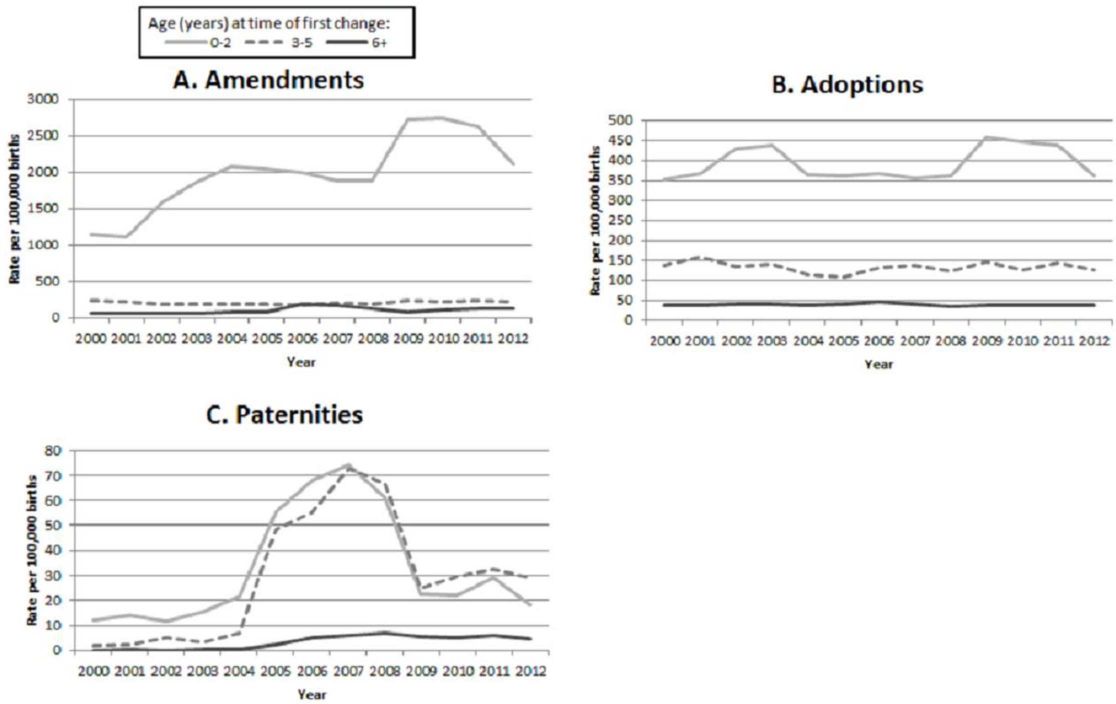

Figure 4.2 Rate of name changes due to amendments, adoptions, and paternities by year of change and age group of the child, recorded on birth certificates between 2000 and 2012 for births occurring in any year (n=2,589,265 births from 1905 to 2012).

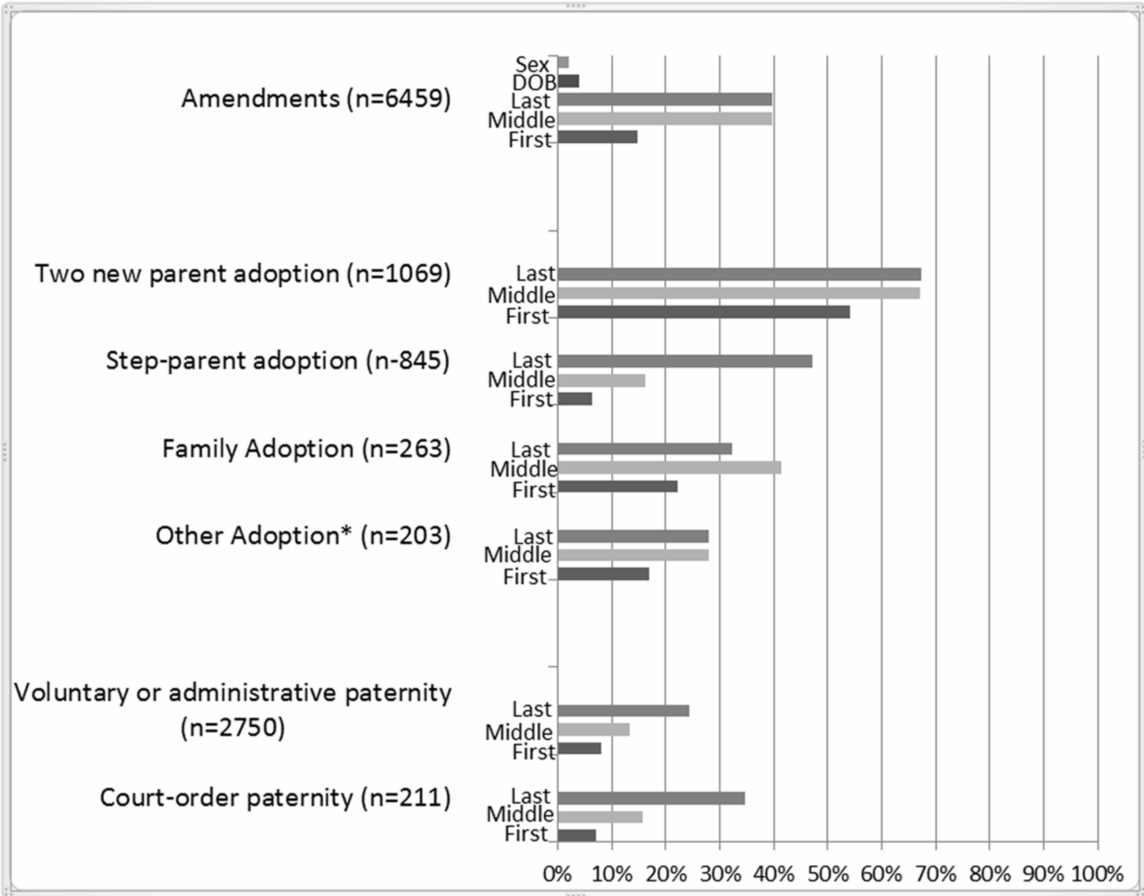

Figure 4.3 Frequency of changes to identity information, by field changed and type of change event for changes to a Utah birth certificate in 2010 for births in any prior year (n=2 589 265 births between 1905 and 2012). DOB, date of birth. Last, middle, first indicate last, middle, and first names.

References

1. HIMSS. Recommendation to Congress 2012 [cited 2013 10-11-2013]. Available from: http://www.himss.org/files/HIMSSorg/policy/d/2012_Ask1_PatientDataMatchingStrategy.pdf.

2. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183-210.

3. Newcombe HB. *Handbook of record linkage: methods for health and statistical studies, administration, and business*: Oxford University Press, Inc.; 1988.

4. Newcombe HB, Kennedy JM. Record linkage: making maximum use of the discriminating power of identifying information. *Communications of the ACM*. 1962;5(11):563-6.

5. Winkler WE. Overview of record linkage and current research directions. *US Bureau of the Census Research Report Series*; 2006.

6. Acheson ED. Medical record linkage. *Medical record linkage*. 1967.

7. DuVall SL, Fraser AM, Kerber RA, Mineau GP, Thomas A. The impact of a growing minority population on identification of duplicate records in an enterprise data warehouse. *Studies in Health Technology and Informatics*. 2010;160(Pt 2):1122.

8. DuVall SL, Kerber RA, Thomas A. Extending the Fellegi–Sunter probabilistic record linkage method for approximate field comparators. *J Biomed Inform*. 2010;43(1):24-30.

9. Sariyar M, Borg A, Pommerening K. Controlling false match rates in record linkage using extreme value theory. *J Biomed Inform*. 2011;44(4):648-54.

10. Hillestad RJ, Bigelow JH, Chaudhry B, Dreyer P, Greenberg M, Meili R, et al. *Identity crisis: An examination of the costs and benefits of a unique patient identifier for the US health care system*: RAND; 2008.

11. Bittle MJ, Charache P, Wassilchalk DM. Performance Improvement: Registration-Associated Patient Misidentification in an Academic Medical Center: Causes and Corrections. *Jt Comm J Qual Patient Saf*. 2007;33(1):25-33.

12. Aung E, Rao C, Walker S. Teaching cause-of-death certification: lessons from international experience. *Postgrad Med J*. 2010;86(1013):143-52.

13. Mills ME. Linkage of patient records to support continuity of care: Issues and future directions. *Stud Health Technol Inform*. 2006;122:320-4.

14. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk management and healthcare policy*. 2011;4.

15. Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, et al. Design of a National Distributed Health Data Network. *Annals of Internal Medicine*. 2009;151(5):341-4.

16. Utah Office of Vital Records and Statistics. Utah's Vital Statistics: Births and Deaths, Updated 2012 [cited 2013 09/28/2013]. Available from: http://health.utah.gov/vitalrecords/2011bx_UpdatedFinal_03292013.pdf.

17. Duncan J, Xu W, Narus SP, Nangle B, Thornton S, Facelli J. A Focus Area Maturity Model for a Statewide Master Person Index. *Online J Public Health Inform*. 2013;5(2).

18. Durand A. Three Tiers of Identity. *Digital Identity World*. 2002:19-69.

19. Deutsch MB, Green J, Keatley J, Mayer G, Hastings J, Hall AM. Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group. *J Am Med Inform Assoc*. 2013;20(4):700-3.

20. Long WS. Social Security numbers issued: A 20-year review. *Soc Sec Bull*. 1993;56:83.

21. Arellano MG, Weber GI. Issues in identification and linkage of patient records across an integrated delivery system. *J Healthc Inf Manag*. 1998;12(3):43-52.

22. McClellan MA. Duplicate medical records: a survey of twin cities healthcare organizations. *Proc AMIA Symp*. 2009: American Medical Informatics Association; 2009. p. 421.

23. SAS Software 2014 [cited 2014 01-27-2014]. Available from: http://www.sas.com/en_us/home.html.

24. Herman E. *The adoption history project*: Ellen Herman, Department of History, University of Oregon; 2003.

25. Courts US. Utah State Courts Statistics 2014 [cited 2014 02/24/2014]. Available from: https://www.utcourts.gov/stats/.

26. U.S. Department of Health and Human Services. *The Adoption and Foster Care Analysis and Reporting System (AFCARS) Report*; 2013 [cited 02-03-2014]. Available from: http://www.acf.hhs.gov/sites/default/files/cb/afcarsreport20.pdf.

27. Solomon-Fears C. *Paternity establishment: Child support and Beyond*: Nova Publishers; 2003.

28. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. *Proc AMIA Symp*. 2003: American Medical Informatics Association; 2003. p. 259.

29. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011;64(5):565-72.

30. Birth and Fetal Death Reporting: Integrating the Healthcare Enterprise (IHE); 2013 [cited 2013 10/11/2013]. Available from: http://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_Suppl_BFDR.pdf.

CHAPTER 5

BUILDING AN ONTOLOGY FOR IDENTITY RESOLUTION IN HEALTHCARE AND PUBLIC HEALTH

This is a pre-copy-editing, author-produced copy of an article accepted for publication in the Online Journal of Public Health Informatics following peer review. The definitive publisher-authenticated version, Duncan J, Eilbeck K, Narus SP, Clyde S, Thornton S, Staes C. Building an Ontology for Identity Resolution in Healthcare and Public Health Online Journal of Public Health Informatics.7(2):219, is available online at http://journals.uic.edu/ojs/index.php/ojphi/article/view/6010.

Background and significance

Many strategies for healthcare improvement rely on integrating patient clinical data from multiple encounters and from multiple provider organizations. The ability to correctly match patient-specific records within and across organizations in healthcare and public health to support Health Information Exchange (HIE) has become such a critical need that the U.S. Office of the National Coordinator for Health Information Technology (ONC) launched the Patient Identification and Matching Initiative in September, 2013. The goal of this collaborative initiative was to conduct environmental scans and in-depth literature reviews across stakeholder organizations to identify problems in patient matching and to develop recommendations for improvement. The Initiative's final report cited, among other things, the need to standardize both the structure and content of patient identity attributes used to link records to realize improvements in patient matching across the many disparate organizational boundaries.[1]

Without standards for personal identity attributes, record linkage is complicated by issues of both structural and semantic heterogeneity.[2] Structural heterogeneity arises because different information systems vary in quality, completeness, and formats for storing identifying information. Semantic heterogeneity arises from differences in the content and meaning of demographic identity fields in disparate information systems.

Past research has focused on developing and improving methods for record linkage.[3-8] These methods are constrained by the need to attain extremely high degrees of sensitivity while maintaining almost perfect specificity. According to the ONC report, patient safety concerns dictate that matching algorithms be adjusted to produce duplicates rather than overlays (false positives), because wrong care could be provided based on an incorrect match.[1] In practice, both probabilistic and deterministic linkage methods typically divide records being linked into three groups: matches, non-matches, and possible matches. Possible matches, which are records that match in many but not all respects, require costly human resolution, estimated to be as much as $60 per record.[1]

Possible matches often arise from the fact that demographic attributes used to link records, such as names and addresses, may be recorded incorrectly[9] or may change over time. Previously, we showed that events such as adoptions, paternity acknowledgments, and amendments result in changes to birth certificate identities for over 6% of children, particularly in their first two years of life.[10] Following the birth of a child in a hospital, these events, combined with numerous reports from hospitals to public health, creates unique challenges for integrating information.

A hospital birth drives the creation of electronic records in multiple healthcare and public health information systems. The hospital creates administrative and electronic medical records for the newborn child. Hospital staff administer a hepatitis B immunization, details of which are sent to an immunization registry in a public health department.[11, 12] Universal newborn hearing screening (UNHS) test results are reported to the public health department,[13-15] as are newborn metabolic screening (NBS) (heelstick) test results.[16] Integrated child health information systems,[17] such as Utah's Child Health Advanced Record Management (CHARM),[18] attempt to link these records using combinations of nonunique demographic identifiers such as name, date of birth, sex, address, and telephone number, and locally unique identifiers such as newborn screening kit numbers and birth certificate state file numbers. In addition, efforts such as Utah's statewide master person index have attempted to link persons across public health and healthcare master person indices (MPIs).[19]

Ontologies are formal descriptions of the terms in a domain and the relationships between terms. They have proven useful in overcoming challenges in integrating information due to semantic and structural limitations.[2, 20] For example, OntoGrate[21] is an ontology-based framework that demonstrates the utility of converting relational database schemas to ontologies to solve query translation and data translation problems across heterogeneous relational databases. Ontologies have been used in diverse applications such as semantic integration in biomedical experimental protocols,[22] and integrating clinical information for oncology research.[23]

In addition to promoting data integration, ontologies modeled in languages such as the W3C standard Web Ontology Language (OWL) demonstrate the ability to employ description-logic-based reasoning.[24] OWL's reasoning capability has been demonstrated in genomics,[25] developing clinical practice guidelines,[26] and for studying relationships among biological entities.[27]

Despite the growing use of ontologies for data integration, we were unable to find literature describing their use for identity resolution or record linkage. The goal of this project was to investigate existing ontologies, or to develop a new one, to facilitate linking birth and early-childhood records in both clinical and public health information systems. Our specific objectives were to develop and validate an ontology to: 1) identify concepts in the domain of identity, including the components of identity and the events subsequent to birth that result in creation or change of identity; 2) develop an ontology to facilitate the integration of data from multiple sources such as an electronic health record (EHR), birth certificate registry, immunization registry, and other public health sources; and 3) validate our ontology's ability to model identity-changing events over time and their resulting changes to individual identity components.

Methods

We adopted the methods of Uschold and Gruninger,[28] progressing along a continuum of formality from informal domain descriptions to rigorously formal structured ontology language. The basic methodology includes: identify the ontology's purpose and scope; build the ontology through knowledge acquisition, coding, and integration of existing ontologies; and evaluation.

Identify ontology purpose and scope

We defined our ontology's purpose as describing: a) the sources of identity information, b) events that result in the creation, change, or sharing of identity information, and c) the components of identity that are created, changed or shared among healthcare and public health entities. Because our interest is in the integration of early childhood identities, we restricted the ontology's scope to the events surrounding the birth of a child in a hospital and the subsequent reports to public health. Ultimately, however, this ontology of identity may be extended to cover the continuum of life events.

Knowledge acquisition

We conducted interviews with administrative domain experts at three Salt Lake City-area hospitals, including University of Utah Health Sciences Center, Intermountain Healthcare, and St. Mark's Hospital. We also interviewed public health domain experts within the Utah Department of Health, from the Office of Vital Records and Statistics, Utah Statewide Immunization Information System (USIIS), Early Hearing Detection and Intervention Program, and Newborn Screening Program. These interviews resulted in the development of process models describing the creation and transmission of identity information between healthcare and public health entities for postbirth activities. We created process models using Business Process Modeling Notation (BPMN),[29] with the goal of documenting specific postbirth events and the identity artifacts created and transmitted among various information systems.

Integration of existing ontologies

To promote interoperability and reuse of domain knowledge, Uschold and Grueninger recommend integration of existing ontologies. We searched for existing ontologies that describe events and their timing, as well as ontologies for identity information, using various online sources, including: National Center for Biomedical Ontologies (NCBO) Bioportal (http://bioportal.bioontology.org/); Protégé Ontology Library (http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library); OBO Foundry (http://www.obofoundry.org/); and Google Scholar (https://scholar.google.com/).

Ontology coding

We represented our ontology using the Web Ontology Language (OWL)[24] using the Protégé OWL Editor.[30] We chose Protégé because of its status as an open-source application with a significant user community, availability of plug-ins to extend its functionality, support of automated reasoning and consistency checking, and its ability to both create and instantiate our ontology using the same tool.

Evaluation

We evaluated both the content of our ontology and its potential utility for tasks in identity resolution. One author (JD) mapped identifiers from public health databases, including birth certificates, death certificates, and immunization information system (IIS) to ontology classes to validate the ontology's content and coverage. Independently, a domain expert from USIIS mapped IIS identity fields to ontology classes, and a vital statistics domain expert did the same for birth and death certificates. We compared the independent mappings and demonstrated concurrence between them. We then simulated identity events and their corresponding attributes in Protégé and used SPARQL queries to demonstrate ontology use cases. We also explored additional benefits of using an ontological approach for storing and searching identity information.

Results

Interviews with domain experts within UDOH and in various area hospitals revealed marked similarities, with some interesting differences, in administrative events following the birth of a child. Figure 5.1 depicts a high-level process model derived from these interviews. All of the process models created are included as supplemental materials.

Childbirth results in the creation of a unique record for the child in the hospital's information system and enterprise master person index (EMPI). In some facilities, this new record creation may take place as preregistration, while in other facilities the newborn child's record is only created after a live birth. Regardless of its timing, the name in the new record is usually a placeholder name consisting of a combination of the mother's first and last names and the sex of the child, such as 'Baby Boy Jane Doe' as the newborn son of Jane Doe. Before discharge, a newborn child typically undergoes metabolic screening, hearing screening, and a hepatitis B vaccination, each resulting in a report to the state public health department. These records may be transmitted individually or in batches, electronically or on paper, and may contain the child's real or placeholder name. Before the child is discharged, parents of the newborn complete a worksheet that documents parent and child demographic information, including the name of the newborn child. (An example of the national standard birth certificate worksheet can be found at http://www.cdc.gov/nchs/data/dvs/momswkstf_improv.pdf). Hospital birth certificate clerks abstract health information for mother and child using another standardized worksheet, called the facility worksheet, which is based on the 2003 U.S. national standard birth certificate.[31] The contents of both the parental and facility worksheets constitute the child's birth certificate. In Utah, this information is submitted to public health using a web-based form. At some point, typically after discharge, hospital staff will replace the placeholder name in the child's hospital EMPI record with the birth certificate name. The timing of this update, and the source of the birth certificate name, varied for each of the three hospitals we interviewed.

Integration of existing ontologies

Analysis of the birth events and process models suggested that we focus ontology development on two broad categories: events and their associated timing, and the components of personal identity.

Event ontologies have been used in distributed event-based systems to integrate temporal information from various sources.[32] Eventory, which Wang X-j et al. developed as an event-based repository of multimedia artifacts, uses an ontological approach that defines an event as an occurrence that unfolds over time.[33] The ontology behind Eventory identifies who, what, when, and where as the characteristics used to describe events. The Event Ontology,[34] developed to describe the domain of music, combines an event ontology with the reasoning capabilities of OWL to create a *semantic workspace* in which new knowledge added to the repository gains semantic value from existing knowledge in the repository. Event Model F is a comprehensive event model based on the foundational ontology DOLCE[35] that provides support for representing mereological and causal relationships. The Simple Event Model (SEM) was designed as a general-purpose event model with the ability to integrate domain-specific vocabularies.[36]

After a review of event models and their characteristics, we chose SEM as our event model because of its simplicity and ability to integrate existing domain-specific ontologies. SEM allows for different viewpoints of a single event, resulting in the ability to define event-bounded roles, time-bounded validity of facts, and attribution of the authoritative source of a statement. Each of these characteristics is potentially important in a cross-enterprise exchange for identity resolution. Event-bounded roles are useful for modeling situations where a person may be a child in one event and a parent in another, for example. Time-bounded validity of facts can be used to model changes in specific identifiers over time, while attributing a fact to an authoritative source can be used to create a “golden record” of identity facts based on the most current facts from the most authoritative sources.

Components of personal identity

Much work has been completed attempting to standardize both the storage and exchange of patient clinical information to support interoperability and clinical decision support, including the HL7 Reference Information Model (RIM),[37] OpenEHR archetypes,[38] and Clinical Element Models (CEM).[39] Each of these implements its own language for representation: Clinical Document Architecture (CDA) for the HL7 RIM, Archetype Definition Language (ADL) for OpenEHR, and Clinical Element Modeling Language (CEML) for CEM. Because the personal identifiers are similar across all three, and because the CEM has been implemented and validated in OWL,[40] we chose to adopt CEM's to represent identifiers.

We integrated the OWL representation of the CEM Core Patient class as a domain-specific representation of the SEM sem:ActorType property. A high-level overview of the relationship between the two ontologies and a subset of classes and relationships is shown in Figure 5.2. We manually mapped public health source database fields to CEM attributes for birth certificates, death certificates, and the immunization registry. In Protégé, we mapped individual data elements from contributing systems to our ontology using rdf:sameAs relationships. The complete CEM Core Patient model and typical value sets for coded values may be obtained at http://clinicalelement.com. Our combined SEM-CEM ontology contains 92 classes, 32 object properties, 4 data properties, and 1404 axioms.

Figure 5.2 shows a high-level overview of the combined SEM-CEM ontologies. Each event in SEM-CEM can be described with multiple actors, places, and times. SEM implements a constraint class named *Role* that is used to modify the *actor*(s) in an event. This feature allows the same *actor* to appear in multiple *events*, as is the case in a database such as the birth certificate registry. We used the *Role* class to indicate an actor's *role* in an event record. We added an additional property, *recordType*, as a link to the CEM Core Patient class, thus providing event-specific identity information.

Time is one of the core classes in SEM. The advantage of modeling time as an OWL class as opposed to a simple data property is that numerous property assertions may be made about a time instance. For example, a sem:Time class may have a data property pointing to a timestamp indicating the time of an event. Additionally, an instance of time can be described by a sem:TimeType which may be used to classify a time as actual, estimated, or observed.

To validate our ontology as a SPARQL endpoint for queries, we created simulated events and identities in a test birth certificate repository using Protégé and the SEM-CEM ontology. Our repository simulated various birth certificate events, including change events such as paternity registration, amendment, and adoption events that we identified in a previous paper.[10] We then developed SPARQL queries to search for a combination of identifiers and extract all of the resulting information for the given person, including names and associated events.

To validate SEM-CEM as a central integration agent, we implemented SEM-CEM in a simulated central repository of identity, integrating events from various public health and healthcare sources including hospital, birth certificate, immunization information systems (IIS), early hearing detection and intervention, and newborn metabolic screening. We then used SPARQL to query and assemble identity history across time for our simulated persons.

We created instances of identity events using the combined SEM-CEM ontology in Protégé. Table 5.1 describes the events, actors and places that were modeled.

We created a simulated birth-certificate knowledgebase in Protégé using the SEM-CEM ontology. For example, we created a child, John Richard Doe, born on 11/28/2014 to an unmarried mother, Jane Doe. A voluntary declaration of paternity filed a few days later changes the child's last name to Stagg in the birth-certificate registry. Figure 5.3 illustrates the SPARQL query and results for the simulated child. The query returns two events, a birth registration event and a paternity event. It is important to note that the actor class, in this case JohnDoeActorNode, is the URI that refers to the same person involved in both events.

A subsequent SPARQL query was used to drill down into the CEM identity items associated with each role returned above. That query and its results are shown in Figure 5.4.

<u>Additional strengths of model</u>

The approach used to model names in CEM, as depicted in Figure 5.5, can be effectively used to enable unstructured searches of proper names in our triple-store. In the CEM ontology, each component of a person's name, including names with multiple values such as Mary Jane, can be modeled as the object of a cem:item property of a CEInstance. Each object has a corresponding rdf:type.

This model enables unstructured name queries using SPARQL against our identity triple-store, resulting in the ability to search on any combination of first, middle, and last name, given in any order. For example a SPARQL query for the Mary Jane Doe in Figure 5.5 would return the individual record regardless of whether Jane is classified as a first or middle name. This is very advantageous when searching for names, which may often be reversed, missing, or incorrectly split between first, middle, and last name fields in a traditional database. This can also be useful for modeling informal variations or nicknames used in place of canonical names, such as Jim for James or Marge for Margaret, or for names encoded phonetically using algorithms such as soundex or metaphone.[41]

Discussion

Identity resolution and record linkage strategies are able to achieve high degrees of accuracy; however there are always possible matches that must be manually linked.[42] Manual linkage, in fact, is typically the "gold standard," as a human judge is able to review a record pair and infer the occurrence of a typographical error or an event such as a name change or marriage. Human review is time-consuming and costly, but also essential for some records. A semantic repository that models events and their corresponding identities can be valuable in the resolution of questionable identities.

The CEM ontology by itself, with its comprehensive list of identifiers, is sufficient to solve the issues of semantic heterogeneity in a record-linkage system. The SEM model adds context that can be used to automate the manual linkage of questionable identities by reasoning about changes due to specific events. We did not incorporate contextual reasoning into this project. Following are two distinct scenarios for using the SEM-CEM ontology for identity resolution.

Scenario 1: Integration of distributed events. Clinical events such as birth, immunization, and clinic visits result in administrative events such as creating a new patient record, modifying or verifying an existing patient record, or merging or unmerging records in an MPI. The diffuse nature of these events across healthcare organizations or registries within a public health department suggests the need for a distributed event-based architecture to manage and coordinate identity. For example, MPIs in an MPI cluster may subscribe to events and receive notifications when they occur. Thus, any MPI in the cluster may be able to keep up to date when an identity is verified, when a name is changed, or when records are merged or unmerged in any other MPI in the cluster. In this example, the ontology can provide semantic information with respect to the source, quality, and provenance of the identity record.

Scenario 2: Ontology as a query model. When an identifier such as a name is changed in an information system, a master record is typically updated, while the previous information may be stored in a relational table as a part of change history. A database query typically searches against what is in the master record for a person, not what previously was in the record. Querying for 'what was' requires an understanding of the relational structure of the database. Using an ontology and storing identity information as triples facilitates the use of SPARQL, allowing users to query against what is and what was without understanding the underlying structure of the data. If the record is for a child and the difference is in surname, the MPI may initiate a query to the birth database and determine if a name change has been registered. Similarly, if surnames and dates of birth are the same but the first names are different, the MPI may initiate a query to determine if a child was part of a multiple birth event. This automated function may be particularly useful in the sensitive context of linking records involving children who are adopted, where human review reveals the link between pre- and post-adoption identities.

Limitations

The primary limitation of this work is that the events and activities we observed and modeled were in three Salt Lake City facilities and the Utah Department of Health and may not correspond to other settings. However, national standards and routine practices for in-patient registration and other events in early childhood likely result in similar workflows in other facilities and jurisdictions and our model allows for variation. A second limitation is that we used simulated identity events to test common scenarios that occur during hospital birth and early childhood. More formal testing with real data and scenarios, for a variety of facilities and public health jurisdictions, is needed to thoroughly validate this model.

Conclusions

The SEM-CEM ontology can be used to overcome structural and semantic heterogeneity issues when linking disparate data sources. The ontology also may be used to create a semantic repository that can be used to provide a view of how an individual's identity evolves over time, or to provide a more complete view of identity when integrating incomplete or partial records. This view can be useful for both manual and automated resolution of possible matches in the record linkage process. Further research is needed to explore the potential of the description-logic-based reasoning capabilities of OWL in identity resolution.

Acknowledgments

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United

States National Institutes of Health.

The terminology and CEM models referenced in this work are modified versions of terminology and CEM models originally created by General Electric Company and Intermountain Healthcare.

We appreciate the contributions of the following domain experts who provided information for our knowledge acquisition and ontology mapping: Chris Pratt, Terry Lucherini, Mary Staub, Robyn Hansen, Stephanie Robinson, Jullie McBratney, Christine Perfili, Michelle Wilde, and Lynne Barrett.

Table 5.1. Information system events, actors, and places modeled in SEM-CEM Ontology

Event Name	Place	Actors	Comments
BirthRegistrationEvent	Birth Registry	Child Mother Father	A birth certificate records information about a child, mother and, optionally, a father
AddNewPatientEvent	Hospital EMPI or EHR	Child	
Immunization RecordEvent	EHR or IIS	Child	Immunization may be recorded in the EHR or directly entered by hospital staff into IIS
Immunization ReportEvent		Child EHR IIS	Immunization recorded in EHR are reported to IIS in real-time messages or in batches
NewbornScreening ReportEvent[1]		Child Birth Facility Laboratory	Birth facility submits blood spot and identifying information to laboratory for analysis. This is typically a manual process.
NewbornScreening Results ReportEvent[1]		Child Laboratory EHR	Reporting results back to the source hospital may be done electronically or manually with a fax
HearingScreening RecordEvent	EHR	Child	
HearingScreening ReportEvent[1]		Child EHR EHDI	EHDI = Early Hearing Detection and Intervention system
PaternityEvent	Birth registry	Child	
AdoptionEvent	Birth registry	Child Child 2	The original record is sealed A new child record is created, using the original child's State File Number (unique identifier)
DeathRegistrationEvent	Death Registry	Decedent	
DeathReportEvent[1]		Death Registry External system(s)	Fact of death information, including date, transmitted from death registry to an external system
BirthCertificate AmendmentEvent	Birth registry	Child	Amendment, may need to only model fields that change

Table 5.1 Continued

Event Name	Place	Actors	Comments
DataUpdateEvent	All	Information System	Incorrect or missing information is updated in an existing record
PostDischarge NameUpdateEvent	Hospital EMPI	Child	Change event--hospital updates the placeholder name to the legal name on birth certificate
BirthCertUpdateEvent		Child	A child's name may be updated in IIS or other system
RecordMergeEvent		Person1 Person2	A record repository such as an EMPI may merge multiple identities into one, or may split one into multiple
RecordSplitEvent		Person1 Person2	

[1]In Report events, information systems are modeled as actors, not places.

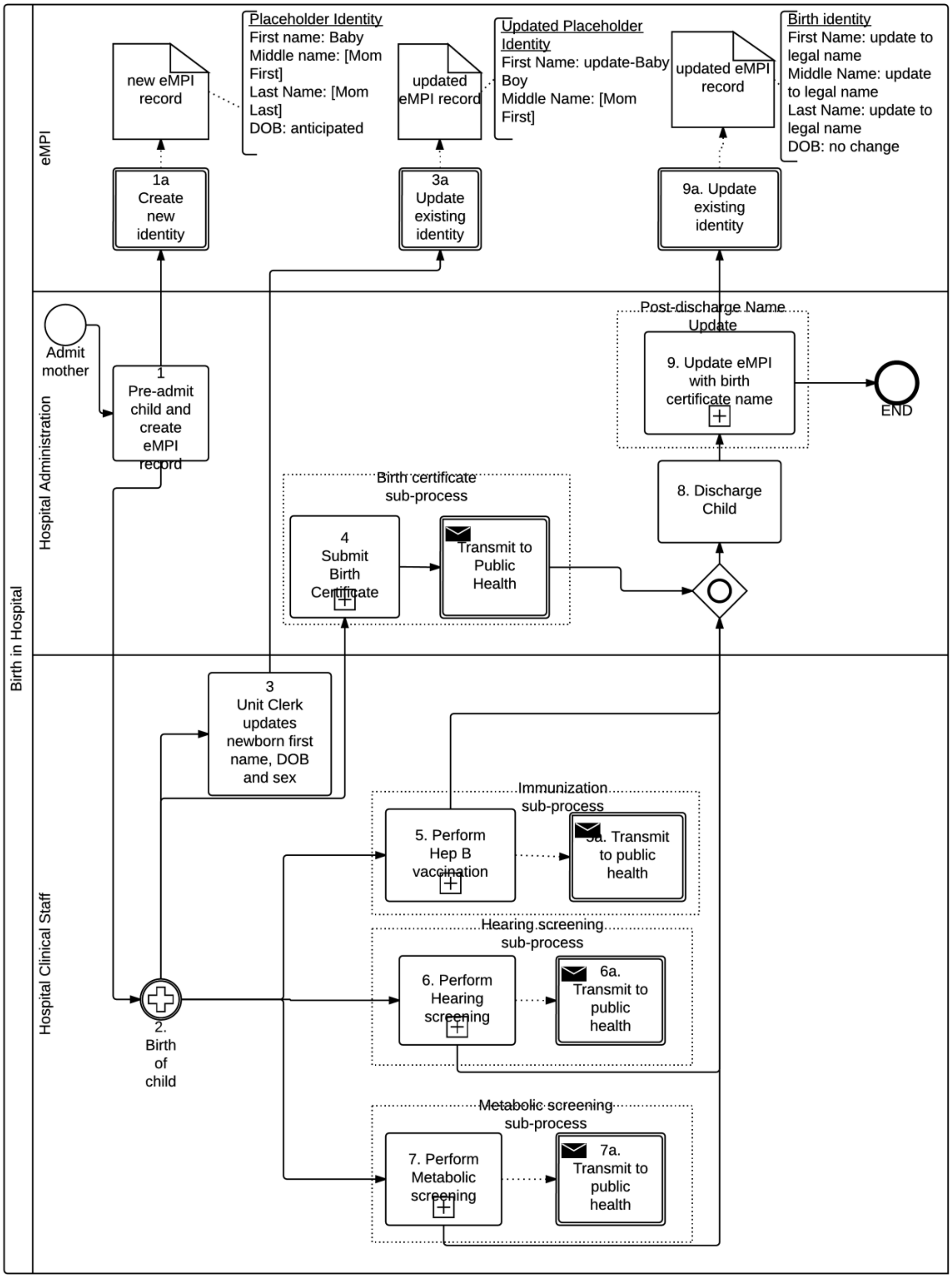

Figure 5.1. High-level process model for birth-related events in a hospital using BPMN

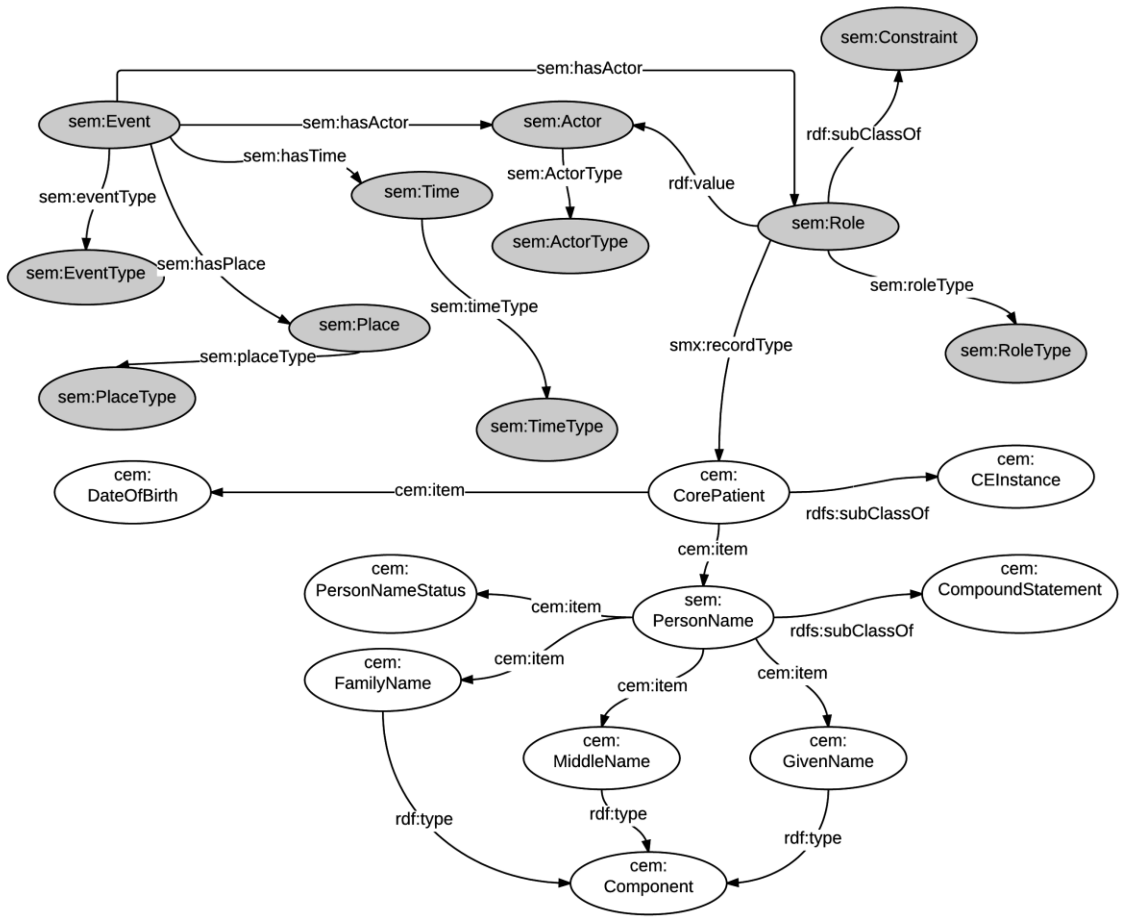

Figure 5.2. High-level overview of the combined SEM-CEM ontologies. (Classes are represented by ovals and relationships are represented by arrows.)

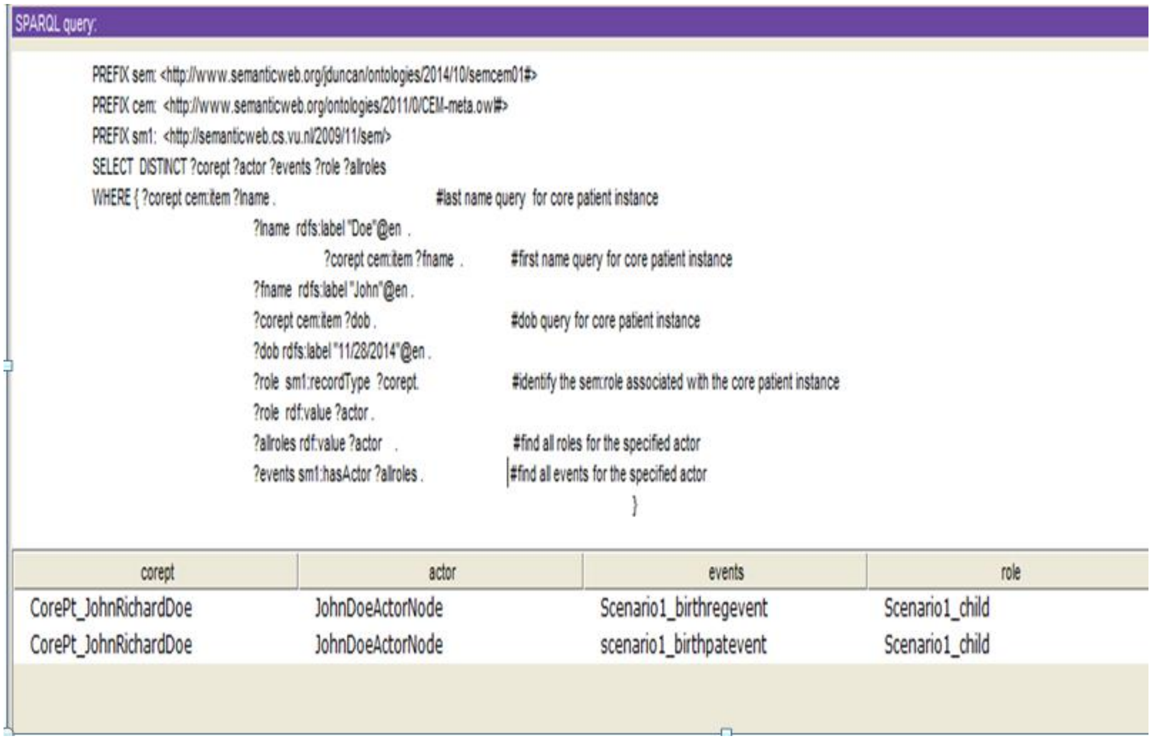

Figure 5.3. SPARQL query returns associated actors, roles, and events for an individual named John Doe, born 11/28/2014.

SPARQL query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sem: <http://www.semanticweb.org/jduncan/ontologies/2014/10/semcem01#>
PREFIX cem:  <http://www.semanticweb.org/ontologies/2011/0/CEM-meta.owl#>
PREFIX sm1:  <http://semanticweb.cs.vu.nl/2009/11/sem/>
SELECT  DISTINCT   ?ceinstance ?items ?type
                WHERE {

                   { ?role sm1:recordType ?ceinstance .
                                        ?role rdfs:label "Scenario1_child"@en }
                                        UNION {
                                        ?role sm1:recordType ?ceinstance .
                                        ?role rdfs:label "Scenario1_child1"@en  . }
                                        ?ceinstance cem:item ?items .
                                        ?items a ?type.
```

ceinstance	items	type
CorePt_JohnRichardDoe	Male	AdministrativeGender
CorePt_JohnRichardStagg	Male	AdministrativeGender
CorePt_JohnRichardDoe	11/28/2014	BirthDate
CorePt_JohnRichardStagg	11/28/2014	BirthDate
CorePt_JohnRichardDoe	Doe	FamilyName
CorePt_JohnRichardStagg	Stagg	FamilyName
CorePt_JohnRichardDoe	John	GivenName
CorePt_JohnRichardStagg	John	GivenName
CorePt_JohnRichardDoe	Richard	MiddleName
CorePt_JohnRichardStagg	Richard	MiddleName
CorePt_JohnRichardDoe	11/28/2014	StartTime
CorePt_JohnRichardStagg	11/28/2014	StartTime

Execute

Figure 5.4. SPARQL query returns identity items and their corresponding types for the two CEM instances identified in Figure 5.3.

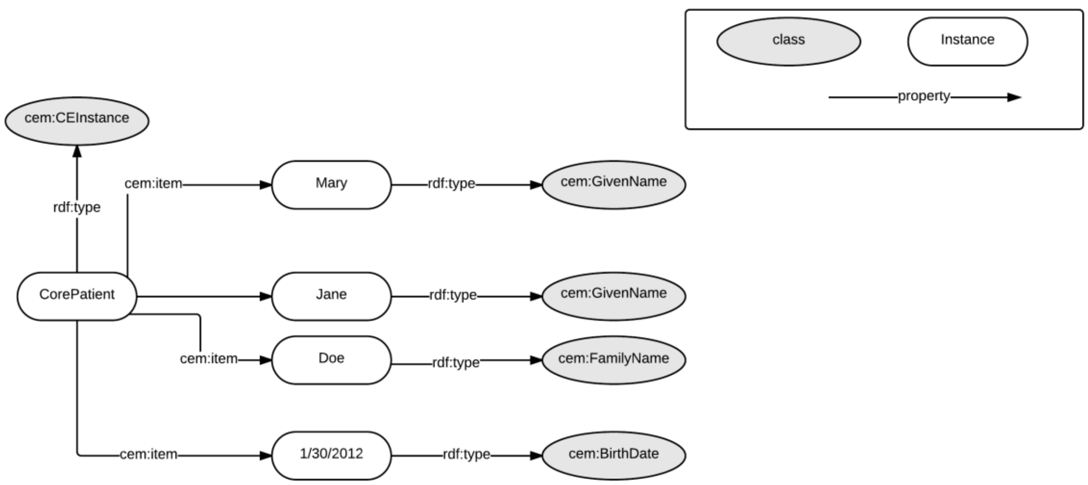

Figure 5.5. Example of the modeling of identity properties in SEM-CEM.

References

1. Audacious Inquiry LLC. Patient Identification and Matching Final Report Baltimore, MD2014 [cited 2015 May 22, 2015]. Available from: http://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.

2. Wache H, Voegele T, Visser U, Stuckenschmidt H, Schuster G, Neumann H, et al., editors. Ontology-based integration of information-a survey of existing approaches. *IJCAI-01 workshop: ontologies and information sharing*; 2001: Citeseer.

3. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183-210.

4. Newcombe HB. *Handbook of record linkage: methods for health and statistical studies, administration, and business*: Oxford University Press, Inc.; 1988.

5. Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. *Proc AMIA Symp*. 2002:305-9.

6. Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. *Proceedings of the annual symposium on computer application in medical care: American Medical Informatics Association*; 1995. p. 397.

7. Campbell KM, Deck D, Krupski A. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and abasic'deterministic algorithm. *Health Inform J*. 2008;14(1):5-15.

8. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011;64(5):565-72.

9. McClellan MA. Duplicate medical records: a survey of twin cities healthcare organizations. *Proc AMIA Symp*. 2009: American Medical Informatics Association; 2009. p. 421.

10. Duncan J, Narus SP, Clyde S, Eilbeck K, Thornton S, Staes C. Birth of identity: understanding changes to birth certificates and their value for identity resolution. *J Am Med Inform Assoc*. 2014.

11. Martin DW, Lowery NE, Brand B, Gold R, Horlick G. Immunization Information Systems: A Decade of Progress in Law and Policy. *J Public Health Manag Pract*. 2014.

12. Yasuda K. Immunization information systems. *Pediatrics*. 2006;118(3):1293-5.

13. Halpin KS, Smith KY, Widen JE, Chertoff ME. Effects of universal newborn hearing screening on an early intervention program for children with hearing loss, birth to 3 yr of age. *J Am Acad Audiol*. 2010;21(3):169-75.

14. Wrightson AS. Universal newborn hearing screening. *Am Fam Physician*. 2007;75(9).

15. Yoshinaga-Itano C. Levels of evidence: universal newborn hearing screening (UNHS) and early hearing detection and intervention systems (EHDI). *J Comm Disorders*. 2004;37:451-65.

16. Kim S, Lloyd-Puryear MA, Tonniges TF. Examination of the communication practices between state newborn screening programs and the medical home. *Pediatrics*. 2003;111(2):E120-6.

17. Hinman AR, Atkinson D, Diehn TN, Eichwald J, Heberer J, Hoyle T, et al. Principles and core functions of integrated child health information systems. *J Public Health Manag Pract*. 2004;10:S52-S6.

18. Hinman AR, Eichwald J, Linzer D, Saarlas KN. Integrating child health information systems. *Am J Public Health.* 2005;95(11):1923.

19. Duncan J, Xu W, Narus SP, Nangle B, Thornton S, Facelli J. A Focus Area Maturity Model for a Statewide Master Person Index. *Online J Public Health Inform*. 2013;5(2).

20. Gagnon M. Ontology-based integration of data sources. *Information Fusion*, 2007 10th International Conference on: IEEE; 2007. p. 1-8.

21. Dou D, LePendu P. Ontology-based integration for relational databases. *Proceedings of the 2006 ACM symposium on Applied computing*: ACM; 2006. p. 461-6.

22. Soldatova LN, Nadis D, King RD, Basu PS, Haddi E, Baumle V, et al. EXACT2: the semantics of biomedical protocols. *BMC Bioinformatics*. 2014;15 Suppl 14:S5.

23. Segagni D, Tibollo V, Dagliati A, Perinati L, Zambelli A, Priori S, et al. The ONCO-I2b2 project: integrating biobank information and clinical data to support translational research in oncology. *Stud Health Technol Inform*. 2010;169:887-91.

24. McGuinness DL, Van Harmelen F. *OWL web ontology language overview*. *W3C recommendation*. 2004;10(10):2004.

25. Wolstencroft K, Stevens R, Haarslev V. *Applying OWL reasoning to genomic data. Semantic Web*: Springer; 2007. p. 225-48.

26. Abidi S. Exploiting Semantic Web Technologies to Develop OWL-Based Clinical Practice Guideline Execution Engines. *IEEE J Biomed Health Inform*. 2014 Dec 18.

27. Chen H, Chen X, Gu P, Wu Z, Yu T. OWL Reasoning Framework over Big Biological Knowledge Network. *BioMed research international*. 2014;2014.

28. Uschold M, Gruninger M. Ontologies: Principles, methods and applications. *The knowledge engineering review*. 1996;11(02):93-136.

29. Mendling J, Weidlich M, Weske M, editors. Business Process Modeling Notation. *Second International Workshop, BPMN 2010, Potsdam, Germany*; 2010.

30. Knublauch H, Fergerson RW, Noy NF, Musen MA. The Protégé OWL plugin: An open development environment for semantic web applications. *The Semantic Web–ISWC 2004*: Springer; 2004. p. 229-43.

31. NCHS. National Vital Statistics System 2013 [cited 2013 10-11-2013]. Available from: http://www.cdc.gov/nchs/nvss.htm.

32. Mühl G, Fiege L, Pietzuch P. *Distributed event-based systems*: Springer; 2006.

33. Wang X-j, Mamadgi S, Thekdi A, Kelliher A, Sundaram H, editors. Eventory--An Event Based Media Repository. *Semantic Computing, 2007 ICSC 2007 International Conference on*; 2007: IEEE.

34. Raimond Y, Abdallah S. The event ontology. Technical report, 2007 [cited 10-11-2013]. Available from: http://motools.sourceforge.net/event, 2007.

35. Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L. Sweetening ontologies with DOLCE. Knowledge engineering and knowledge management: *Ontologies and the semantic Web*: Springer; 2002. p. 166-81.

36. Van Hage WR, Malaisé V, Segers R, Hollink L, Schreiber G. Design and use of the Simple Event Model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*. 2011;9(2):128-36.

37. Health Level Seven I. HL7 Reference Information Model Ann Arbor, MI: Health Level Seven; 1994 [cited 2014 12/10/2014]. Available from: http://www.hl7.org/library/data-model/RIM/modelpage_non.htm.

38. Welcome to openEHR: openEHR Foundation; [cited 2015 03-07-2015]. Available from: http://www.openehr.org/.

39. Coyle JF, Mori AR, Huff SM. Standards for detailed clinical models as the basis for medical data exchange and decision support. *Int J Med Inform*. 2003;69(2):157-74.

40. Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. *J Am Med Inform Assoc*. 2012:amiajnl-2012-001326.

41. Karakasidis A, Verykios VS. Privacy preserving record linkage using phonetic codes. *Informatics*, 2009 BCI'09 Fourth Balkan Conference in: IEEE; 2009. p. 101-6.

42. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. *Proc AMIA Symp*. 2003: American Medical Informatics Association; 2003. p. 259.

CHAPTER 6

USING CONTEXT-BASED PRIVACY POLICIES AND COMMONSENSE REASONING TO ENABLE IDENTITY RESOLUTION

Background and significance

Birth certificates in the United States are important identity documents used to establish a person's name, parentage, date of birth, and citizenship. Identity information from birth certificates is often used to create or deduplicate identifiers in other public health systems, such as immunization information systems (IIS),[1] and child health integrated information systems such as Utah's Child Health Advance Record Management (CHARM).[2]

Birth registration, a state responsibility in the United States, has evolved from a paper process to an electronic one, such that by 2010 all states had electronic systems for collecting, registering, and issuing birth certificates.[3] Birth certificate identifiers may be seen as a potential source for operational identity resolution in healthcare enterprise master person indices (EMPIs), but their use for this purpose is complicated by the fact that they may change to reflect changes in a person's real identity due to events such as adoptions.[4] Adoptions may complicate identity resolution for a significant number of children. According to a special report from the US Bureau of the Census, in 2010 2.3% of children of all ages in the United States were adopted and an additional 4.7% were stepchildren.[5]

In most U.S. states, adoptions are private, and original legal birth certificates are sealed by law.[6] In practice, this means that an adopted child's original birth certificate is replaced with a record showing the adoptive parents as birth parents, while the original legal birth certificate is prohibited from release or inspection except when authorized by law. When birth certificates were primarily paper documents, sealing the original birth certificate was as simple as placing the document in an envelope, sealing it, and preparing a new paper certificate with the adoptive parents listed. Utah's administrative rule regarding birth certificates and adoption (Utah Administrative Code R436-5-7) still reflects this paper process. This process of sealing and creating a new certificate only occurs when the adoptive parents present a decree of adoption signed by a judge to the Vital Statistics office in the state of the child's birth and request a new birth certificate, in some cases months or years after the judge has signed the decree. In our previous work we described the difference between Tier I identity change events that occur to people and Tier II events that occur to records about people.[4] Figure 6.1 illustrates a typical timeline of Tier I and II events in an adoption process.

There are two important variations in the timing of events in Figure 6.1. First, the birth certificate may be flagged at the time the birth certificate is registered for a child relinquished at birth, or later when an adoption agency or state child services agency notifies Vital Statistics that an adoption is pending, or in many cases not at all. Second, there is considerable variation between the time a court issues an adoption decree (t=6) and when parents request a new birth certificate (t=7). Some parents may do this immediately, while others will wait until there is a need, such as to obtain a passport or register the child for school.

This variation in the timing of events creates challenges for linking birth certificates for adopted children. For example, if a birth record is not promptly flagged, it may be used to incorrectly update or populate other public health information systems. And since updating and merging pre- and post-adoption records typically depends on a new birth certificate being issued, delays in that process result in the inability of systems to merge known duplicate records. Linkages for statistical purposes may intentionally omit,[7] or fail to link[8] records where an adoption has been completed or is pending. Moreover, many linkages to support statistical analyses involving maternal characteristics must link to a child's biological, not adoptive, mother.

The current practice in Utah's Vital Statistics office is to suppress providing birth certificate information to the Utah Statewide IIS or CHARM when a record is flagged as pending adoption. This static approach is relatively easy to implement in an operational data exchange, but restricts potentially valuable information from being shared in contexts where it might be allowable or even advantageous to do so. For example, a report published in 2010 by the American Academy of Pediatrics recommended a comprehensive review of medical history for newly adopted children, when possible, to include much of the information maintained in public health registries, including birth certificate data, immunization, newborn hearing and metabolic screening results.[9] This information is readily available, but its integration is prevented by static policies that restrict linking records that are flagged as pending adoption.

The sharing of electronic records between healthcare and public health entities effectively relegates the sealing of an adoptee's original paper certificate to an anachronism. Oftentimes, identity teams in healthcare or public health organizations performing manual linkage of possible duplicate records are able to identify pre- and post-adoption records, effectively *breaching* the sealed record. Thus, not only do static policies arbitrarily limit information sharing, they may actually enable and encourage privacy breaches because records must be manually reviewed.

In the emerging interoperable environment between healthcare and public health, where information exchange depends critically on patient identification, replacing static policies with a contextual approach to sharing will enable the use of birth certificates for identity resolution and may even enhance privacy protection for sensitive adoption records.

Because the timing of the events shown in Figure 6.1 has a substantial effect on how a person's identity is recorded and shared in multiple information systems, we sought to investigate the use of computerized logic methods to reason about identity events and policy decisions over time. For this purpose, we chose to focus on a logical formalism known as Event Calculus.[10]

Event calculus

Event Calculus, based on first-order predicate logic, is used to reason about events and their consequences over time. It shows potential for both modeling and reasoning about identity change events over time and for characterizing policy decisions based on contextual factors. First introduced in 1989 as a method to reason about events over time,[11] event calculus has been used to analyze the effects of policy and system behavior in systems management,[12] for tracking contracts,[13] and for describing workflows in information systems.[14] Because we are interested in the sequence of events and not the absolute times between them, we focused on the discrete event calculus (DEC).[15] The three basic components of DEC are *events, fluents* (time-varying properties), and *time*.

Table 6.1 shows some of the simple constructs of DEC.

Commonsense law of inertia

The logic of event calculus includes the notion that domain properties do not change unless they are directly affected by events. This idea is known as the commonsense law of inertia. In identity resolution, the commonsense law of inertia means that, for example, a child's name in a database will not change spontaneously, but only as a result of observable events.

Objective

Our goal was to assess the potential of DEC to model and reason about identity-changing events and policy considerations when releasing birth certificate information, particularly for complicated cases involving adoptions. Additionally, we sought to validate the potential of event calculus to reason about changes to identity. Our specific objectives were to: 1) Use DEC to model identity events and their effects following the birth of a child; 2) Assess the potential of contextual policies to enable the sharing of information between a birth certificate registry and external systems for identity resolution; and 3) Explore the use of abductive reasoning from the perspective of a healthcare EMPI.

Methods

Formalization of the Problem

Let R_H be a database of records in a healthcare enterprise master person index (EMPI) and R_B a database of records in a birth certificate registry in public health. Other databases such as IIS may be designated as R_I, for example. Further, uppercase letters A, B, C and so on will refer to unique, individual Tier I persons, while a_i, b_i, c_i refer to collections of identity attributes (Tier II identities) recorded at some discrete time point i, where a_i corresponds to A, b_i corresponds to B, and so on. Let r_H represent a record in R_H such that $r_H(a_i)$ identifies person A as a unique patient in R_H with a collection of identifiers, a_i. Using this notation, $r_H(b_i)$ would uniquely identify person B. We define r_B as a unique record in R_B that is an individual's birth certificate such that $r_B(a_0)$ identifies person A's original birth identity. Following this scheme, records in other systems may be designated as $r_X(a_i)$ for a collection of identifiers a_i in some database R_X. To further clarify differences in birth certificates, we will use the following conventions:

$r_B(a_0)$ is person A's original birth certificate

$r_B(a_1)$ is an amended birth certificate for person A

$r_{B,F}(a_i)$ means A's birth certificate was flagged at discrete time i. We will further stipulate that if a birth certificate is flagged at i, it will be flagged for all previous identities a_j where $0 \leq j < i$. (Rule 1)

$r_{B,S}(a_i)$ indicates that A's birth certificate is sealed at discrete time i. We further stipulate that if a birth certificate is sealed at i, it will be sealed for all previous identities a_j where $0 \leq j < i$. (Rule 2)

$r_{B,N}(a_i)$ is a new postadoption birth certificate created for A at discrete time i. We stipulate that when a new birth certificate $r_{B,N}(a_i)$ is created, all birth certificates are sealed for previous identities a_j where $0 \leq j < i$. (Rule 3)

Suppose a hospital system R_H has a record $\rho_1 = r_{H(a_i)}$ that uniquely identifies a patient A. A different record ρ_2 is determined to be a duplicate if it can be shown that $\rho_2 = r_{H(a_j)}$, meaning that both ρ_1 and ρ_2 identify person A using different sets of identifiers, a_i and a_j. Record deduplication is the automated process of identifying duplicate records in a database. That is to say, a deduplication process may determine that $\rho_1 = r_H(a_i)$ and $\rho_2 = r_H(a_j)$, meaning that ρ_1 and ρ_2 both refer to individual A. Alternatively, it may happen that $\rho_1 = r_H(a_i)$ and $\rho_2 = r_H(b_j)$, meaning that each record identifies a different individual, A and B, respectively. In the first case, $\rho_1 = r_H(a_i)$ and $\rho_2 = r_H(a_j)$, we will write that $\rho_1 \sim \rho_2$, and for the case $\rho_1 = r_H(a_i)$ and $\rho_2 = r_H(b_j)$ we will write $\rho_1 \neq \rho_2$.

Probabilistic linkage software will compute a match likelihood score that quantifies the likelihood that $\rho_1 = \rho_2$ despite individual differences in attributes between the two records. In practice, two thresholds are established where scores above the upper threshold are considered duplicates and scores below are considered to not match. Scores between the thresholds are considered possible matches and are subject to human review.[16]

One approach to resolving possible matches is the use of a bridging file—an authoritative source such as a government registry that contains substantially error-free information that is free from duplicates—to assist with record linkage and deduplication.[17] In this project, we envision the state's birth certificate registry as such an authoritative source, available for queries of pairs of possible match records from EMPIs in healthcare or health information exchange (HIE).

A single record query from R_H to R_B for some record ρ_1 results in a search of R_B for a matching record. There are five possible outcomes of this query: (1) the record is not found, (2) the record is found but is flagged as pending adoption, (3) the record is found and is not flagged for adoption, (4) the record is sealed, (5) the record is the postadoption record for a completed adoption where the original record has been sealed. Under the current static policy of not releasing information about records that are pending adoption, R_B may respond that the record does not exist (false negative) when a record is flagged for adoption or sealed.

A query of a pair of records ρ_1 and ρ_2 from R_H to R_B to ascertain whether they represent the same child's pre- and post-adoption identities results in more possibilities, especially if R_B is enhanced to include all chronological records for individuals, including those where information has been amended or even sealed. That is to say, R_B' is an enhanced birth registry containing chronological identity records, so that when a record is changed for reasons such as amendment or paternity or adoption, a new record is created and added to the database and the previous record is retained for search purposes. This means that all records $r_B(a_1)$, $r_{B,F}(a_1)$, $r_{B,S}(a_1)$, and $r_{B,N}(a_1)$ are searchable in R_B' . An ontological approach to accomplish this is discussed in Chapter 5.[18]

Using the above formalization, we created and executed DEC axiomatizations for identity transactions in healthcare and public health systems following the birth of a child. We distinguished between Tier I events that happen to persons; Tier II (information) events that occur in information systems, typically in response to Tier I events; and Tier II (record) events, a special kind of Tier II event that occurs at a database level. Table 6.2 lists all of the events that were modeled by type of event.

We used DECReasoner,[19] an open-source, Python-based program that converts

DEC axioms to satisfiability problems, to execute our various common identity outcomes for different combinations of the events following birth in a hospital. Following is a brief description of how events and their actions were modeled in DEC.

In DEC, an event can be triggered with a *triggering axiom* as follows:

$Happens(BirthRegistration(person), t_1) \rightarrow$

$Happens(Record(person, identity, BirthRegistry), t_1)$ (Axiom 1)

Axiom 1 states that a birth registration event at time t_1 triggers the Record(person, identity, system) event. Similar triggering axioms were created for each of the trigger events listed in Table 6.2.

DEC allows for events with nondeterministic effects, such as when a healthcare visit may result in a new record being created or the verification of an existing record. In DEC this can be represented using a determining fluent, as follows:

$HoldsAt(AddRecordDF(system), t_1) \rightarrow$

$Initiates(ClinicVisit(person, system), AddNew(system), t_1)$ (Axiom 2)

$\neg HoldsAt(AddRecordDF(system), t_1) \rightarrow$

$Initiates(ClinicVisit(person, system), Update(system), t_1)$ (Axiom 3)

The truth value of the fluent AddRecordDF(system), true in (Axiom 2) and false in (Axiom 3) above, is used to determine whether a clinic visit results in the addition of a new record or the update of existing records.

Contextual policies

To examine the potential benefits and policy issues involved with releasing birth certificate information, we created a list of all possible outcomes, or contexts, for a paired-record query to an enhanced birth registry R'_B. We eliminated outcomes that were redundant or obviously impossible, such as a record that doesn't exist matching with one that does. For each remaining outcome, we created a DEC model that would result in the particular outcome. We then developed DEC descriptions to simulate a sequence of events resulting in that outcome.

We developed possible query responses based on the principle of minimum necessary information. While birth certificates are excluded from the Health Insurance Portability and Accountability Act (HIPAA), releases of birth information typically adhere to the HIPAA Privacy Rule's provision to release the minimum necessary information to accomplish a specific purpose. Table 6.3 shows the possible responses to a paired-record query to a birth certificate registry. A query may generate more than one response from the table.

The outcomes described above were modeled using deductive reasoning. That is, given an axiom of the form:

$Happens$(event e, time t) →HoldsAt(fluent f,time t+1), (Axiom 4)

if we know event e happened at time t, we can deductively reason that fluent f is true at time t+1. We also assessed the ability of a healthcare system to abductively reason about identity-changing events using the minimum necessary information identified above. Thus, given Axiom 4 above and the knowledge that fluent f holds at time t+1, we can abductively reason that event e happened at time t. For example, given the axiom that an adoption usually leads to a name change, and given a child with a name change, we can abductively reason that the child may have been adopted. For abductive reasoning, we considered three different fluents *f* as follows:

$DOB_1 \neq DOB_2 \rightarrow HoldsAt(DeltaDOB)$ occurs when the date of birth differs between the two records.

$Sex_1 \neq Sex_2 \rightarrow HoldsAt(DeltaSex)$ occurs when the sex differs between the two records.

$Name_1 \neq Name_2 \rightarrow HoldsAt(DeltaName)$ occurs when any name component (first, middle, or last), differs between the two records.

Results

Event models

Figure 6.2 illustrates the results of a basic sequence where there are no data entry errors, relinquishments, or adoptions, and records are updated as planned. Figure 6.3 illustrates the results of a scenario where a child is relinquished at birth, preventing use of the birth certificate to update other public health record systems.

Contextual policies

A query for the match status of two records $\rho_1$and ρ_2, where each record query has 5 possible outcomes and the records either match or not, can result in 30 possible outcome combinations. We can eliminate all of those that include a match with a record that doesn't exist, leaving us with the 25 possible outcomes shown in Table 6.3. Under the current static policy, the Birth Registry may not release information regarding records that are flagged as pending adoption or sealed. We assume that because $R_B^{'}$ maintains a chronological history for each person, if both submitted records match any two versions of a single person's chronological history, the match status of the two submitted records is true, otherwise it is false. Table 6.4 lists the possible birth certificate scenarios, or contexts, whether or not a response is allowed under the current static policy, and a potential response based on each scenario's context. We do not determine whether each contextual situation is allowable or not, only that it may be allowed. Figure 6.4 shows an overview of the paired-record query and response that we modeled.

Of the 25 outcomes shown in Table 6.4, ten cannot generate responses under current static policies that restrict the release of flagged or sealed information. For these ten outcomes, contextual rules may enable a birth registry to respond with information about match status. Rules may also be developed to allow the release of adoption status for certain scenarios where records are flagged or even sealed.

From the healthcare EMPI perspective, abductive reasoning may be used to determine the need to query a birth registry in the first place, as in the case of a name or date of birth discrepancy. Abductive reasoning may also be used to evaluate the responses received from the birth registry for a paired-record query. We considered the latter for three simple scenarios: name mismatch, date of birth mismatch, or sex mismatch. We did not evaluate more complex situations, for example, where both name and sex do not match. Given a fluent, for example HoldsAt(DeltaName), DEC can abductively reason to a hypothesis for a response fluent from a query to R'_B . Table 6.5 shows the abductively-determined hypothesis and possible conclusions for each of the response fluents modeled.

Discussion

Birth certificates are a valuable source of identity information, particularly for children, and birth certificate registries may be seen as an authoritative source for information regarding a child's legal name and date of birth. Use of birth certificates for identity resolution is complicated by the presence of adoption records. Because of state laws making child adoption information private, Vital Statistics agencies typically default to a static policy of not releasing identifying information for matching purposes when a child has been relinquished or otherwise flagged as pending adoption. This static policy ostensibly protects a child's birth identity from being linked to a postadoption identity, thus breaching the privacy of the adoption. The current reality, illustrated by our DEC models, is that because of the various flows of information between healthcare and public health, systems often contain an adopted child's pre- and post-adoption identities. The result of a static no-release policy is duplicate records for adopted children in healthcare and immunization information systems. With nearly 2.3% of all births in the U.S. resulting in adoption,[5] and given that adopted children are more likely to have special health care needs,[20] the ability to link pre- and post-adoption medical records is important. Record linkage is also expensive, costing as much as $60 per record according to a report issued by the Office of the National Coordinator (ONC) for Health Information Technology.[21]

We propose contextual rules for releasing birth certificate information, including information from records that may be flagged as pending adoption or sealed postadoption. Of the outcomes in Table 6.4, ten represent situations in which a birth registry would currently return no result. In many cases, hospitals may already know an adoption is pending, given that relinquishment at birth is a field on the 2003 standard US birth certificate,[22] and hospitals report birth certificates to the state.

DEC models can be used to demonstrate the variety and relative timing of events that result in identity conflicts across information systems, including events with indeterminate effects. The abductive reasoning scenarios illustrate the potential for a healthcare EMPI system initiating a query to reason about the response received. In particular, if the records identify the same individual, the system will want to determine which record is the correct record. Vector timestamps, a simple logical mechanism for ordering events in distributed systems, may be a useful way to accomplish this task.[23] In the case where one record of a record pair is not found by the birth registry, the abductive reasoner will hypothesize a possible typographical error in either of the records. Interestingly, the hospital system could possibly confirm a typographical error by slightly modifying data in one of the records and resubmitting. If the record matching the birth registry was the one modified, the query response would be that no records matched.

Commonsense reasoning of the sort enabled by DEC can be used to reason about specific changes in records, particularly using a minimal response from a birth certificate registry. The scenario where a hospital uses an automated query to validate a possible match can reduce the burden of manual review and preserve privacy for adoptees. Such a scenario may also be used to integrate authoritative source queries in identity resolution exchanges such as Cross-Community Patient Discovery (XCPD).[24]

Our analysis is limited by the fact that it has not been attempted using real identities and events. Our abductive models were simple in nature and did not consider complex events or combinations of potential changes in records. And while abductive reasoning can be used to identify the occurrence of a change event such as an adoption or paternity, the ability to infer which type of event based on name change patterns is limited. Other approaches, such as Bayesian networks[25] using name change data may also prove to be effective at hypothesizing about specific name change events.

We did not address the legal considerations in releasing limited birth certificate information for identity resolution purposes. Just as the digital age has impacted the ability to seal private adoption records, it increasingly creates challenges for patient privacy in the healthcare domain, particularly when information regarding persons is exchanged.[26] The *contextual integrity* methods, developed by Helen Nissenbaum for analyzing information sharing in emerging technologies, may provide insight into legal and policy issues.[27]

Conclusions

A birth certificate registry is an authoritative source for identity information. Contextual rules and artificial reasoning methods such as DEC may be used to overcome barriers to sharing information regarding children who are adopted or pending adoption, reducing the likelihood of duplicate health and immunization records for these children. Contextual rules also reduce the need for manual inspection of records which can lead to disclosure of pre-adoption identities that are supposed to be sealed.

Table 6.1. Basic elements of Discrete Event Calculus

Formula	**Interpretation**
HoldsAt(f,t)	Fluent f is true at time t
Happens(e, t)	Event e happens at time t
Initiates(e,f,t)	Event e causes Fluent f to hold at time t
Terminates(e,f,t)	Event e causes fluent f to not hold at time t
t1 < t2	Time t1 is before time t2

Table 6.2. List of Tier I, Tier II, and record events modeled in Discrete Event Calculus

Events	Description	Event Results
	Tier I (Person) Events	
P1. Birth	Child is born	Triggers P2, P3, P4, I1, I2
P2. Immunization	Child is immunized	Triggers I3
P3. Hearing Screening	Child receives Early Hearing Detection and Intervention (EHDI) screening	Triggers I4
P4. Metabolic screening	Child received metabolic screening	Triggers I5
P5. Relinquishment	Child is relinquished at birth	Triggers I8
P6. Adoption	Child is placed for adoption	Triggers I8, May trigger changes to real identity and subsequent record changes
P7. Healthcare visit	Child visits hospital or clinic	May cause R1, R2, R3
P8. Paternity	A father is added to child's birth certificate, resulting in a change to surname	Triggers R2
	Tier II (Information) Events	
I1. Hospital record creation	Hospital staff create administrative records for child	Triggers R1
I2. Birth certificate registration	Birth certificate is filed with state Vital Statistics	Triggers R1
I3. Immunization report	An immunization is reported to an IIS	Triggers R1 or R2 or R3
I4. Hearing screening report	Newborn hearing screening results reported to public health department	Triggers R1
I5. Metabolic screening report	Newborn metabolic screening results reported to public health department	Triggers R1 or R2
I6. Birth certificate auto-update	Birth certificate information is matched to source program data using software methods	Triggers R6 or R1
I7. Adoption registration	Adoption is registered at Vital Records. Existing birth certificate is sealed, new birth certificate is created	Triggers R1, R8
I8. Birth record flagged	A birth certificate is flagged as pending adoption	Triggers R2
I9. Data entry error	Information is entered into a record incorrectly	May trigger R2
I10. Manual hospital identity update	Hospital staff manually update placeholder record with birth certificate identity	Triggers R2

Table 6.2 Continued

Events	Description	Event Results
I11. Birth certificate updates IIS, EHDI, NBS	Birth certificate identity information is linked to other registry information to update records	May trigger R2, R6
I12. Birth Registry query	An external system queries the birth registry for match status of two records	May trigger R6
	Tier II (Record) Events	
R1. Create record	A new record is created in an information system	May trigger I9
R2. Update record	An existing record is updated with new information	May trigger I9
R3. Verify record	An existing record is verified, no information changed	
R4. Delete record	An existing record is deleted	
R6. Merge records	Two existing records are combined into one	R2, R4
R7. Split records	A single record is split into two new records	R1, R4
R8. Seal record	A birth certificate electronic record is “sealed”	

Table 6.3. Possible responses to a paired-record query for possible match records ρ_1 and ρ_2

Responses
ρ_1 and ρ_2 are the same person
ρ_1 and ρ_2 are not the same person
Neither ρ_1 or ρ_2 was found
ρ_1(or ρ_2) was not found
ρ_1(or ρ_2) is flagged pending adoption
ρ_1(or ρ_2) is a pre-adoption identity
ρ_1(or ρ_2) is a post-adoption identity

Table 6.4. List of all possible outcomes for adoption match query to birth registry for two identity records ρ_1 and ρ_2, whether a query is currently allowed under static policy and a potential contextual response.

	Query results from birth registry: ρ_1	Query results from birth registry: ρ_2	Response allowed under Utah's static policy?	Contextual response
ρ_1 and ρ_2 identify the same person	$r_B(a_i)$	$r_B(a_j)$	Yes	ρ_1 and ρ_2 are the same
	$r_{B,F}(a_i)$	$r_{B,F}(a_j)$	No	ρ_1 and ρ_2 are the same
	$r_{B,N}(a_i)$	$r_{B,N}(a_j)$	Yes	ρ_1 and ρ_2 are the same
	$r_{B,S}(a_i)$	$r_{B,S}(a_j)$	No	ρ_1 and ρ_2 are the same
	$r_{B,N}(a_i)$	$r_{B,S}(a_j)$	No	ρ_1 and ρ_2 are the same
ρ_1 and ρ_2 do not identify the same person	$r_B(a_i)$	$r_B(b_i)$	Yes	ρ_1 and ρ_2 are not the same
	$r_{B,F}(a_i)$	$r_B(b_i)$	No	ρ_1 and ρ_2 are not the same
	$r_{B,N}(a_i)$	$r_B(b_i)$	Yes	ρ_1 and ρ_2 are not the same
	$r_{B,F}(a_i)$	$r_{B,F}(b_i)$	No	ρ_1 and ρ_2 are not the same
	$r_{B,N}(a_i)$	$r_{B,F}(b_i)$	No	ρ_1 and ρ_2 are not the same
	$b_N(a_i)$	$r_{B,N}(b_i)$	Yes	ρ_1 and ρ_2 are not the same
	$r_{B,S}(a_i)$	$r_{B,S}(b_i)$	No	ρ_1 and ρ_2 are not the same
	$r_{B,S}(a_i)$	$r_B(b_i)$	No	ρ_1 and ρ_2 are not the same
	$r_{B,S}(a_i)$	$r_{B,F}(b_i)$	No	ρ_1 and ρ_2 are not the same
	$r_{B,N}(a_i)$	$r_{B,S}(b_i)$	No	ρ_1 and ρ_2 are not the same
one or both records are not found	Not found	Not found	Yes	Neither record was found
	$r_B(a_i)$	Not found	Yes	ρ_2 was not found
	$r_{B,F}(a_i)$	Not found	No	ρ_2 was not found
	$r_{B,N}(a_i)$	Not found	Yes	ρ_2 was not found
	$r_{B,S}(a_i)$	Not found	No	ρ_2 was not found
not possible due to rules	$r_{B,N}(a_i)$	$r_B(a_i)$	N/A-rule 3*	
	$r_{B,F}(a_i)$	$r_B(a_i)$	N/A-rule 1**	
	$r_{B,S}(a_i)$	$r_B(a_i)$	N/A-rule 2***	
	$r_{B,N}(a_i)$	$r_{B,F}(a_i)$	N/A-rule 3	
	$r_{B,S}(a_i)$	$r_{B,F}(a_i)$	N/A-rule 2	

*Rule 3: When a new birth certificate $r_{B,N}(a_i)$ is created, all birth certificates are sealed for previous identities a_j where $0 \leq j < i$.

**Rule 1: If a birth certificate is flagged at i, it will be flagged for all prior identities a_j where $0 \leq j < i$.

***Rule 2: If a birth certificate is sealed at i, it will be sealed for all previous identities a_j where $0 \leq j < i$.

Table 6.5. Abductive modeling scenarios, hypotheses, and possible conclusions from a healthcare EMPI perspective

Response fluent	Hypothesis	Conclusion
Match	Change Event	Need to determine which record is current
No match	Two different persons	If DOB mismatch, may be a parent/child relationship, else no relationship If name mismatch, may be multiple birth
One record not found	Typographic error	Need to determine which record, if either, is correct
Both records not found	N/A	

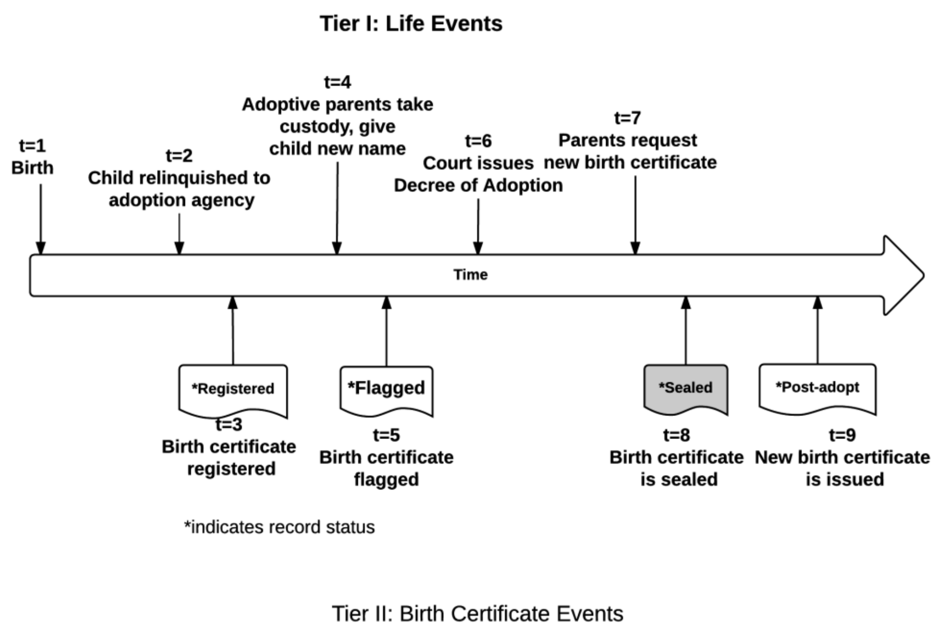

Figure 6.1. Example timeline of an adoption.

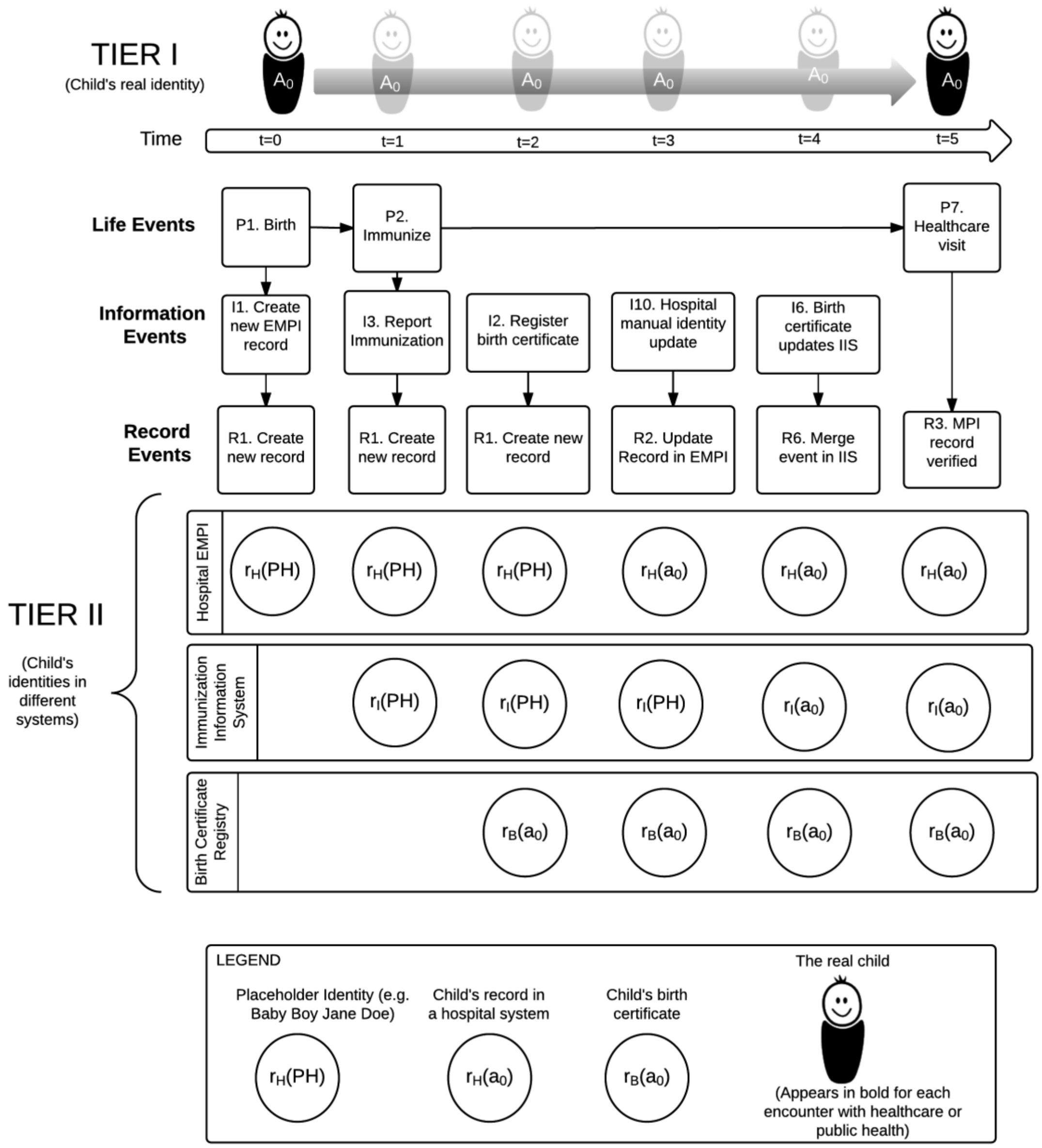

Figure 6.2. Basic sequence of events and outcomes when there are no adoptions or amendments

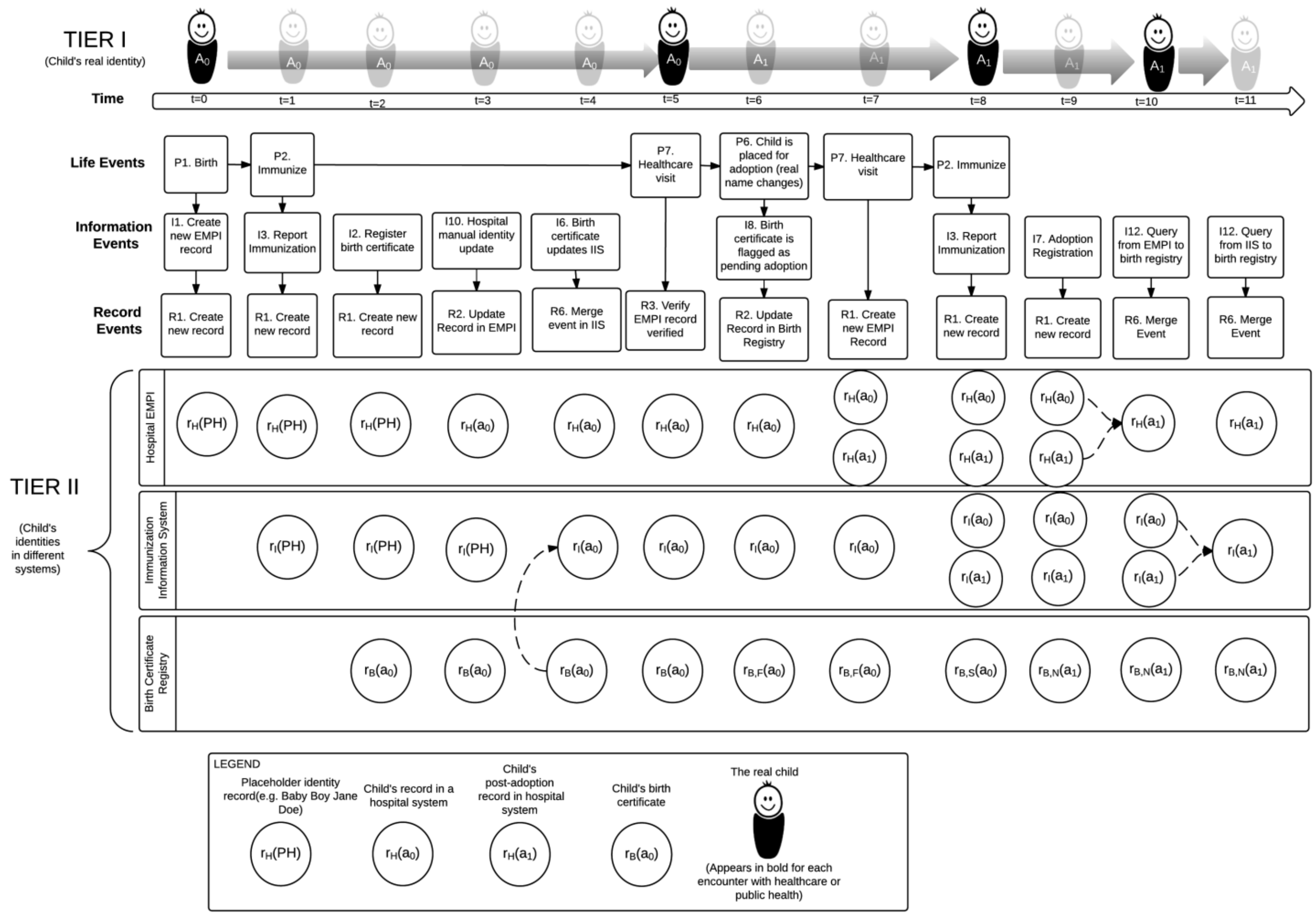

Figure 6.3. Sequence of events and outcomes showing the occurrence of an adoption and creation of duplicate records

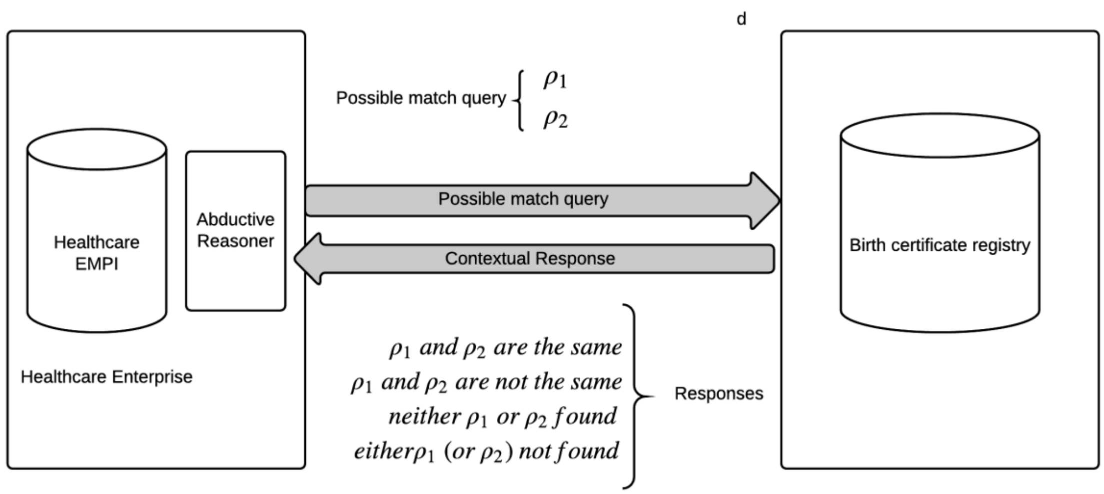

Figure 6.4. Paired-record query from healthcare EMPI to state birth certificate registry and possible responses

References

1. Centers for Disease Control and Prevention. Immunization Information Systems Functional Standards 2015 [cited 2015 04/01/2015]. Available from: http://www.cdc.gov/vaccines/programs/iis/func-stds.html.

2. Hinman AR, Eichwald J, Linzer D, Saarlas KN. Integrating child health information systems. *Am J Public Health*. 2005;95(11):1923.

3. Brumberg HL, Dozor D, Golombek SG. History of the birth certificate: from inception to the future of electronic data. *J Perinatol*. 2012;32(6):407-11.

4. Duncan J, Narus SP, Clyde S, Eilbeck K, Thornton S, Staes C. Birth of identity: understanding changes to birth certificates and their value for identity resolution. *J Am Med Inform Assoc*. 2015;22(e1):e120-9.

5. Kreider RM, Lofquist DA. Adopted Children and Stepchildren: 2010. *Adoption Quarterly*. 2010;13:268-91.

6. Gateway CWI. *Access to Adoption Records*: Washington, DC 2012 [cited 03/13/2015]. Available from: https://www.childwelfare.gov/pubPDFs/infoaccessap.pdf.

7. Salemi JL, Tanner JP, Bailey M, Mbah AK, Salihu HM. Creation and evaluation of a multi-layered maternal and child health database for comparative effectiveness research. *J Registry Manag*. 2013;40(1):14-28.

8. Johnson KE, Beaton SJ, Andrade SE, Cheetham TC, Scott PE, Hammad TA, et al. Methods of linking mothers and infants using health plan data for studies of pregnancy outcomes. *Pharmacoepidemiol Drug Saf*. 2013;22(7):776-82.

9. Jones VF. Comprehensive health evaluation of the newly adopted child. *Pediatrics*. 2012;129(1):e214-23.

10. Shanahan M. The event calculus explained. *Artificial Intelligence Today:* Springer; 1999. p. 409-30.

11. Kowalski R, Sergot M. A logic-based calculus of events. *Foundations of knowledge base management*: Springer; 1989. p. 23-55.

12. Bandara AK, Lupu EC, Russo A. Using event calculus to formalise policy specification and analysis. *Policies for Distributed Systems and Networks*, 2003 Proceedings POLICY 2003 IEEE 4th International Workshop on: IEEE; 2003. p. 26-39.

13. Farrell AD, Sergot MJ, Sallé M, Bartolini C. Using the event calculus for tracking the normative state of contracts. *International Journal of Cooperative Information Systems*. 2005;14(02n03):99-129.

14. Cicekli NK, Yildirim Y. Formalizing workflows using the event calculus. *Database and expert systems applications*: Springer; 2000. p. 222-31.

15. Mueller ET. Automating commonsense reasoning using the event calculus. *Communications of the ACM*. 2009;52(1):113-7.

16. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. *Proc AMIA Symp*. 2003: American Medical Informatics Association; 2003. p. 259.

17. Winkler WE. *Overview of record linkage and current research directions*. Research Report Series: Statistical Research Division, U.S. Census Bureau; 2006.

18. Duncan J, Eilbeck, K. , Narus, S., Clyde, S., Thornton, S., Staes, C. Building an ontology for identity resolution in healthcare and public health. *Online Journal of Public Health Informatics (accepted 06/14/2015)*. 2015.

19. Mueller ET. Discrete event calculus reasoner documentation. *Software documentation, I*BM Thomas J Watson Research Center. 2008;704.

20. Bramlett MD, Radel LF, Blumberg SJ. The health and well-being of adopted children. *Pediatrics*. 2007;119(Supplement 1):S54-S60.

21. Audacious Inquiry L. Patient Identification and Matching Final Report Baltimore, MD2014 [cited 2015 May 22, 2015]. Available from: http://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf.

22. National Center for Health Statistics. National Vital Statistics System 2013 [cited 2013 10-11-2103]. Available from: http://www.cdc.gov/nchs/nvss.htm.

23. Bechini A, Tai K-C, editors. Timestamps for programs using messages and shared variables. *Distributed Computing Systems, 1998 Proceedings 18th International Conference on*; 1998: IEEE.

24. Integrating the Healthcare Enterprise (IHE). Cross-Community Patient Discovery (XCPD) 2011 [cited 2015 03/11/2015]. Available from: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_XCPD_Rev2-3_TI_2011-08_19.pdf.

25. Scotch M, Duggal M, Brandt C, Lin Z, Shiffman R. Use of statistical analysis in the biomedical informatics literature. *J Am Med Inform Assoc*. 2010;17(1):3-5.

26. Mizani MA, Baykal N. A software platform to analyse the ethical issues of electronic patient privacy policy: the S3P example. *J Med Ethics*. 2007;33(12):695-8.

27. Nissenbaum H. Privacy as contextual integrity. *Washington Law Review*. 2004;79(1).

CHAPTER 7

CONCLUSIONS

In today's digital age, connecting identity within and across systems is critical. Birth Certificates are a valuable resource for identity resolution, but their use has been limited because of problems related to their quality, relevance, and privacy concerns, particularly in cases involving adoptions. Birth certificates are important identity documents in the United States, used to establish date of birth, parentage, and citizenship. They are a source for most other government-issued forms of identification, such as driver licenses, social security numbers, and passports. Because of their frequent use to establish identity, it can be expected, and our research has shown, that birth certificate information is validated and amended to ensure its correctness. Overcoming issues of semantic and syntactic heterogeneity, combined with using artificial intelligence techniques such as event calculus, will enable birth certificate registries to act as "authoritative sources" for identity resolution, provided concerns about privacy, quality, and policy are addressed.

Given that birth certificates in the United States are registered in the state where the birth occurred, it is important to ask, "what proportion of the adult population lives in their state of birth?" According to the U.S. Bureau of the Census, in 2010, 59% of all Americans and 89% of children under five years old lived in the state in which they were born.[1] This means that state birth certificate registries are relevant sources of information for the majority of Americans, particularly for younger children.

Significance to field

Much of the research on identity resolution has focused on linkage methodology, obtaining high sensitivity with almost perfect specificity. While these gains are impressive, it is unrealistic to think that such methods will ever reach 100% sensitivity and specificity. This research contributes in several important ways.

First, we establish a framework for understanding changes to identity in terms of tiers, distinguishing between events that happen to people (Tier 1) and events that are recorded in databases (Tier 2). Using this framework, we document the time-dependency of identity attributes recorded on the birth certificate. Typically viewed as a static document capturing the facts of parentage and identity at the time of birth, the birth certificate is actually a dynamic Tier 2 document that changes in response to Tier 1 life events, such as adoptions and paternities. Additionally, the use of birth certificates to obtain other identification documents makes them somewhat self-correcting in that errors in name spelling, date of birth, and sex are corrected when the documents are issued for other purposes. Because of their mutable, self-correcting nature, birth certificates are a valuable source of identity information.

Second, while the dynamic nature of birth certificate information promises value for identity resolution purposes in diverse systems for both healthcare and public health, capturing that dynamic nature in a usable form is problematic. This dissertation proposes an ontological approach to capture both identity attributes and the events that cause those attributes to change, a reflection of the tiers of identity framework we established. The SEM-CEM ontology that is introduced provides both the capacity to model identity attributes and the events that cause them to change over time, as well as to overcome issues of syntactic and semantic heterogeneity that complicate the sharing of identity information among diverse, heterogeneous systems. In simple terms, the ontology provides the means to document changes in Tier 2 identity attributes in response to Tier 1 events in the life of a person.

Third, the very events and changes to birth certificates that make them valuable for identity resolution also create significant policy concerns for state officials charged with protecting the privacy of individuals who are identified in the records. Laws in most states provide for the creation of a new birth certificate and "sealing" of the original birth certificate following an adoption. These laws cause state vital statistics agencies to adopt static policies limiting the release of information for children who have been relinquished at birth or otherwise flagged as pending adoption. This dissertation demonstrates that this static policy is not effective at protecting privacy, and it limits the ability to connect the identity dots for children who are likely more at risk for adverse health outcomes.[2] A more flexible, contextual policy is proposed to enable birth certificate registries to share limited information for identity resolution as an authoritative source.

Finally, this dissertation proposes the use of logical frameworks such as discrete event calculus in information systems to automate reasoning about identity discrepancies between records that match on most, but not all, attributes. This "commonsense" reasoning, similar to the reasoning humans use when reviewing possible matches, would enable systems to reason about and resolve identity discrepancies in an automated fashion.

To summarize, Tier 2 records in various systems are snapshots of Tier 1 identities over time. Those Tier 1 identities change over time in response to life events, causing those snapshots to differ. Understanding the events that cause identities to change over time, using an ontology to capture and represent those changes in time, using logical formalisms to reason and infer about changes, and using contextual rules to promote information flow between disparate, heterogeneous systems will enable healthcare providers to link information in the most difficult cases.

Birth certificate registries may be useful as an authoritative source but they are by no means the only systems that may perform this role, nor are they the only systems with significant privacy considerations. For example, healthcare EMPIs or immunization information systems may also serve as authoritative sources in an interconnected system of identity verification among healthcare and public health systems.

Future directions

Identity resolution will continue to be a critical problem in the increasingly connected healthcare environment in the United States. The research presented in this dissertation should be furthered in three important areas.

First, our study of changes to birth certificate identities showed that societal factors play a role in affecting patterns of changes to personal identities due to events such as adoptions and paternity acknowledgments. Further research is needed to understand how current trends such as advances in assisted reproduction, gestational surrogacy, and same sex marriage may affect identity, changes to identity, and resolution of identity across disparate information systems. Events such as these are driving an evolution in the concepts of biological versus legal parentage that has implications for identities. As healthcare evolves toward a more personal, individualized process relying on genetics and personal histories, informatics research is needed to resolve both biological and legal relationships.

Second, we described the use of an ontology to semantically characterize and store versions of identity information in a searchable triple store. Given that ontology development is necessarily an iterative process, practical experience implementing the SEM-CEM ontology presented in this dissertation is needed.

Third, from a practical standpoint, ontology-based triples are typically stored in relational database management (RDBMS) systems and have shown problems with scalability.[3, 4] Further work is needed to develop and evaluate a scalable and reliable solution that will accommodate large numbers of births and identity changes. Addressing these further research questions can lead to identity resolution methods that are able to automatically link the most difficult cases, avoid costly human review, and avoid the threats to privacy that come from such review.

References

1. US Bureau of the Census. *Migration/Geographic Mobility: 2010 to 2011*. 2011 [cited 3/11/2015. Available from: https://www.census.gov/hhes/migration/data/cps/cps2011.html.

2. Bramlett MD, Radel LF, Blumberg SJ. The health and well-being of adopted children. *Pediatrics*. 2007;119(Supplement 1):S54-S60.

3. Abadi DJ, Marcus A, Madden SR, Hollenbach K, editors. Scalable semantic web data management using vertical partitioning. *Proceedings of the 33rd international conference on Very large data bases*; 2007: VLDB Endowment.

4. Benjamins R, Contreras J, Corcho O, Gomez-Perez A. The six challenges of the Semantic Web. *Knowledge Representation and Intelligent Systems for the Web*, Toulouse, France, 2002.

APPENDIX A

BUSINESS PROCESS MODELS FOR SELECTED BIRTH-RELATED PROCESSES AT INTERMOUNTAIN HEALTHCARE

Introduction

The birth of a child in a modern hospital equipped with an electronic health record (EHR) system results in the creation of a new EHR for the child as well as several reports from the hospital to public health authorities reporting hearing screening, metabolic screening, immunizations, and establishment of the child's birth certificate. Each of these public health reports, in turn, result in a record in a program-specific database at the public health department. Without a common unique identifier, efforts to link records across these databases are complicated by both data entry errors and events such as adoptions that result in changes to identity.

The purpose of this document is to identify specific events, workflows, data exchanges, and other processes surrounding the creation of and subsequent recording of a newborn child's identity following birth in a hospital and the subsequent reporting of those events to public health.

Methods

The goal of each process model is to identify the following:

Events: For the purpose of these models, an event is an action or occurrence in a hospital or public health setting that results or may result in the recording, changing, merging, or transmission of identity information from.

Actions and actors: specific tasks required to record, verify, modify or transmit identity information

Workflows: what is sequence of events after birth that the record is created

Data artifacts: What identifying information is stored, modified or transmitted for each event?

Exchange methods and standards: How are data exchanged between entities? What standards are used?

List of models

Table A.1 provides a listing of all process models and accompanying narrative by page location.

Topic	BPMN Model	Narrative
Overall Process Model for Hospital Births	Page 110	Page 111
Birth certificate process model for hospital birth	Page 113	Page 113
Newborn metabolic screening process model	Page 115	Page 115
Immunization Process Model	Page 117	Page 117
Newborn hearing screening process model	Page 118	Page 118
Post Discharge Birth Name Update	Page 119	Page 119

Process models are created using Business Process Modeling Notation (BPMN) 2.0. In the following process models, each BPMN model is followed by a step-by-step narrative. Numbered steps in each narrative correspond to numbered symbols in the preceding model.

Below is a summary of the symbols used:

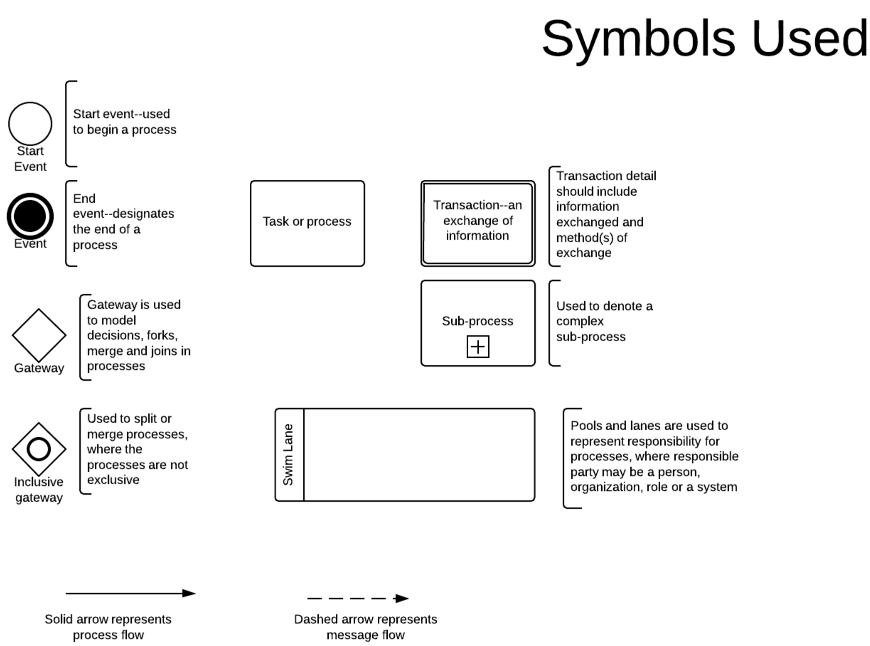

Figure A.1 BPMN Legend

Overall process model for hospital birth at Intermountain Healthcare

The following process model illustrates the overall process of a birth in an Intermountain Hospital. Detailed process models follow for reporting of birth certificates, immunizations, hearing screenings and metabolic screenings.

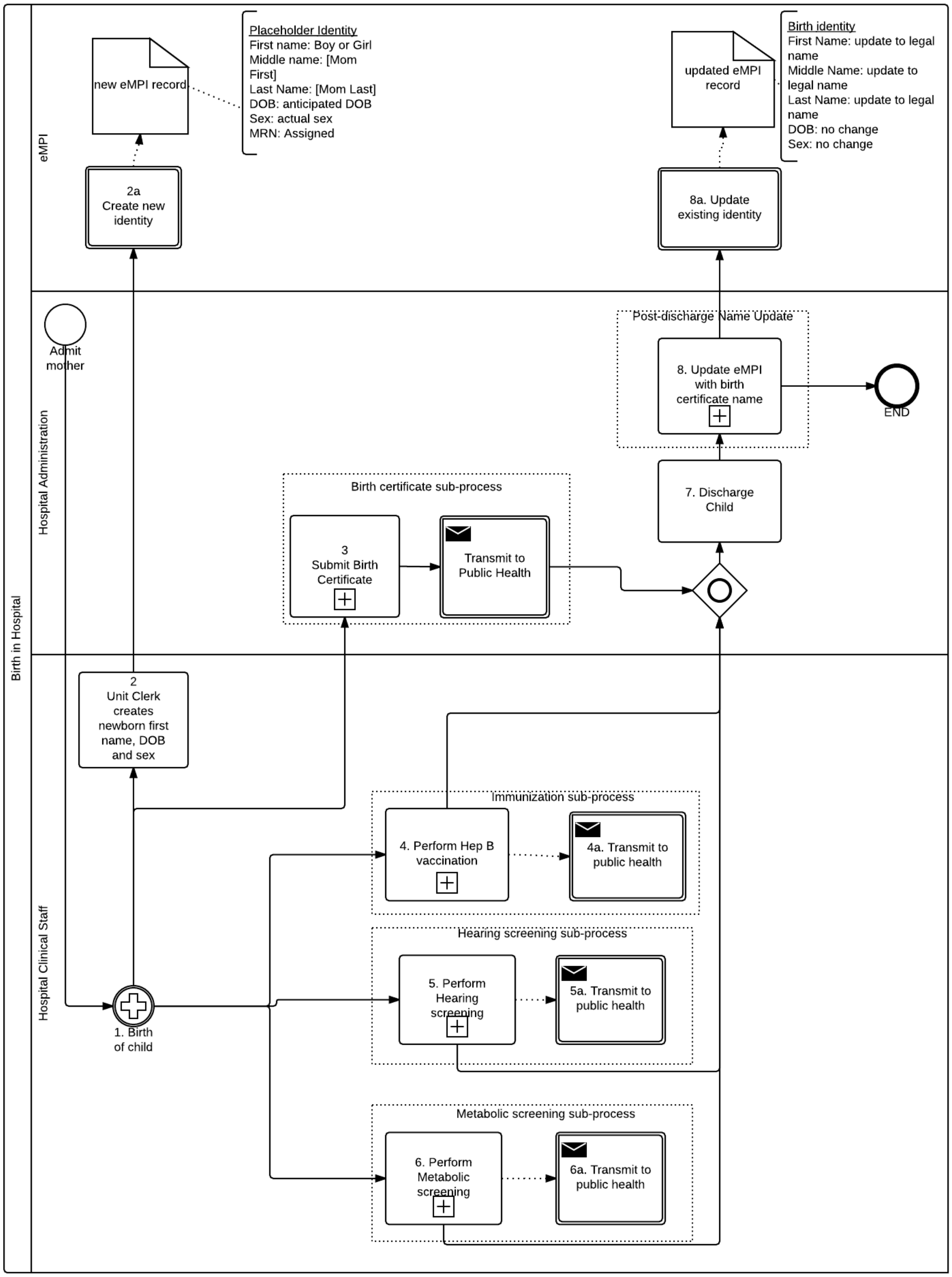

Figure A.2 Overall hospital birth process model

Step-by-step narrative

1.	Event: Birth of child	Results:	Newborn child or children are delivered in the hospital clinical environment.

The birth of a child triggers multiple actions in the clinical and administrative environments that are carried out in no particular order.

2.	Actor: Hospital Unit Clerk (HUC)	Action:	HUC creates new record in the hospital's EMPI system.

A placeholder name is used with "Boy" or "Girl" as the child's first name, depending on the child's sex.

2a.	Actor: EMPI	Action:	Create new record.

A new record is created for the newborn child.

3.	Actor: Birth Certificate Clerk	Action:	Submit birth certificate.

A birth certificate clerk in the administrative environment begins the sub-task of collecting birth certificate information and submitting it to public health. A more detailed model of the birth certificate process is included in Section 3 below.

3a.	Actor: Birth Certificate Clerk	Action:	Transmit birth certificate to public health.

A completed birth certificate is transmitted to public health using a web-based electronic birth registration system provided by the state.

4.	Actor: Clinical staff	Action:	Administer vaccination

Clinical staff administer a Hepatitis B vaccine to the newborn child. The vaccination is recorded in the child's EHR.

4a.	Actor: Clinical staff	Action:	Report vaccination to public health.

Clinical staff report immunizations using the WebKids web application that reports directly to the Utah Statewide Immunization Information System. A more detailed model of the immunization reporting process is included in Section 5 below.

5.	Actor: Audiologist	Action:	Perform hearing screening.

An audiologist performs a hearing screening for each newborn in the clinical setting. Results are recorded in the child's EHR.

5a.	Actor: EHR system	Action:	Transmit hearing screening results to public health EHDI program

Screening results are transmitted to public health in batch mode.

6.	Actor: Clinical staff	Action:	Perform metabolic screening

A clinical staff person completes the newborn metabolic screening card, takes heelstick blood samples from newborn and affixes them to the card. The NBS number on the card is entered into the

EHR and is also affixed to the birth certificate.

6a.	Actor: Clinical staff	Action:	Send metabolic screening to state

Completed screening card with blood spot sample and child information is mailed to the state lab for processing. . A more detailed model of the immunization reporting process is included in Section 4 below.

7.	Actor: Admin staff	Action:	Discharge child

The discharge is an administrative action that takes place after all clinical activities, including screenings and immunizations, take place.

8.	Actor: Admin staff	Action:	Post-discharge birth name update

After discharge the child's legal name, as reported on the birth certificate, is updated in the hospital's EMPI. At Intermountain hospitals, staff download an Excel Spreadsheet from OBTraceVue and use it to manually update each child's name in Epic. A more detailed model of the post-discharge name update process is included in section 7.

8a.	Actor: EMPI	Action:	Update existing identity

The existing EMPI record is updated to show the child's name as reported on the birth certificate. At this point, the placeholder name is replaced with the child's real name.

Birth certificate process model for hospital birth

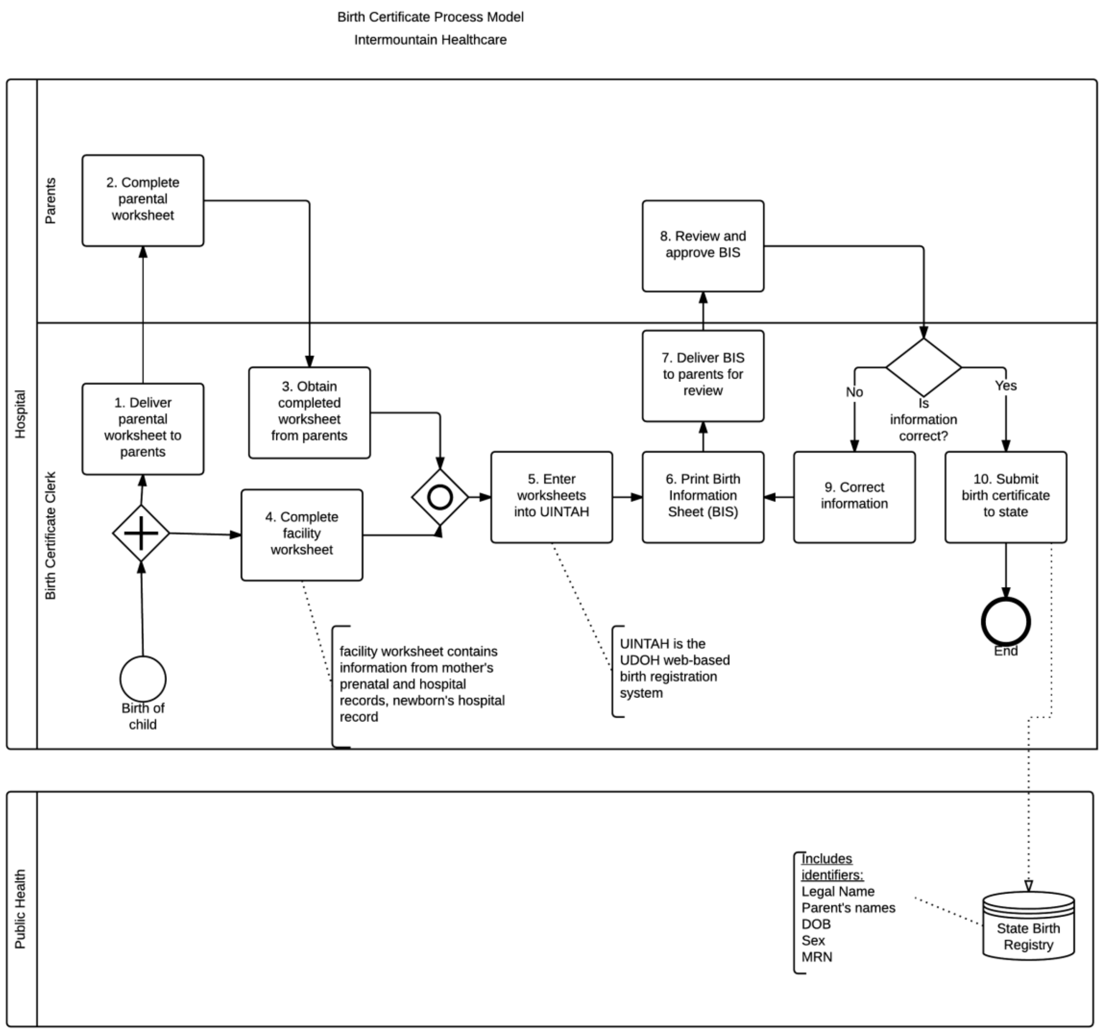

Figure A.3 Birth registration process model

Step-by-step narrative

1.	Actor: Birth certificate clerk	Action:	Deliver parental worksheet to parents

Following the birth of a child, a birth certificate clerk will deliver a parental worksheet to the mother (and father, if present).

2.	Actor: Mother (and father)	Action:	Complete parental worksheet

Parents manually complete a standard worksheet that contians demographic information about the parents and the newborn child, including the desired name of the newborn child. If a mother is married, the husband is presumed to be the father. If unmarried, a voluntary declaration of paternity must be signed by both mother and father in order to include father information on the birth certificate. The parental worksheet also allows parents to provide permission for the state to furnish information to the Social Security Administration (SSA) so that a social security number (SSN) may be issued.

3. Actor: Birth certificate clerk Action: Obtain completed parental worksheet from parents.

The birth certificate clerk will obtain the completed parental worksheet from the parents.

4. Actor: Birth certificate clerk Action: Complete facility worksheet

The birth certificate clerk will complete the facility worksheet by abstracting information from various medical records including the mother's prenatal record, mother's hospital record, and newborn's hospital record.

5. Actor: Birth certificate clerk Action: Enter worksheets into UINTAH

When both the facility and parental worksheets are complete, the birth certificate clerk will create a new birth record in UINTAH, Utah's web-based electronic birth reporting system.

6. Actor: Birth certificate clerk Action: Create birth information sheet (BIS)

After keying information into a new birth record, the clerk will print a BIS for parental review. The BIS contains demographic information provided on the parental worksheet and will be used to verify that information.

7. Actor: Birth certificate clerk Action: Deliver BIS to parents for review

The BC clerk will deliver the BIS to the mother (and father, if present), for review.

8. Actor: Parents Action: Review and approve BIS

Parents review the BIS, particularly information regarding the child's legal name and parental demographics. If the name is spelled incorrectly or any demographic information is not correct, parents may correct the information on the BIS.

9. Actor: Birth certificate clerk Action: Correct information

In the event there are errors on the BIS identified by parents, the BC clerk will correct the information then repeat the verifcation process (B6-B8).

10. Actor: Birth certificate clerk Action: Submit birth certificate to state

Once demographic information is verified by parents, the birth certificate clerk will submit the record to the state using UINTAH. This results in a birth certificate being registered for the child. If parents indicated permission on the parental worksheet, an SSN will be issued by Social Security Administration based on information received from the state.

Newborn metabolic screening process model

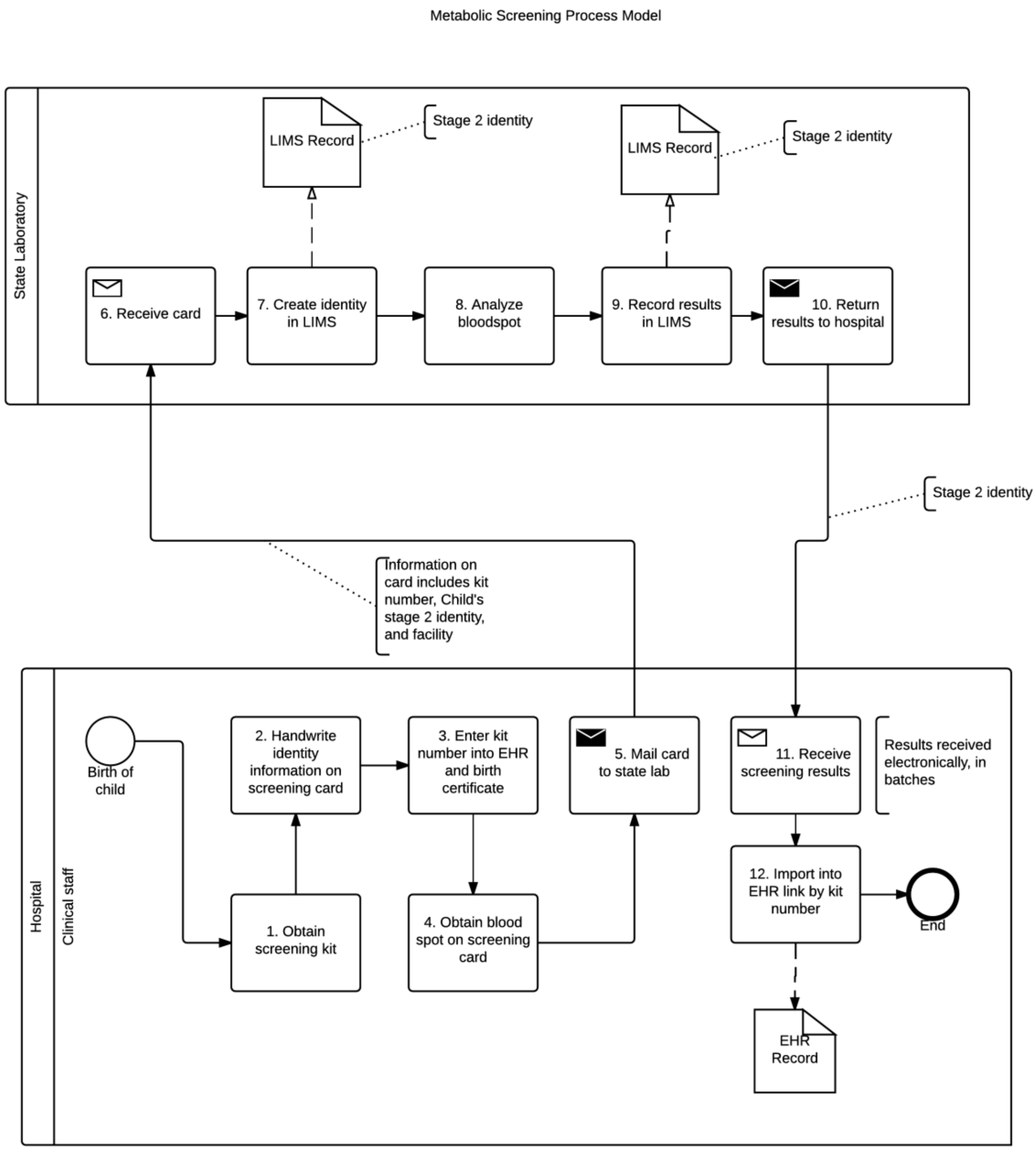

Figure A.4 Newborn metabolic screening process model

<u>Step-by-step narrative</u>

1.	Actor: Hospital clinical staff	Action:	Obtain screening kit

Following the birth of a child, a member of the hospital clinical staff will obtain an unused metabolic screening kit that is identified with a unique, bar-coded identifier.

2.	Actor: Hospital clinical staff	Action:	Handwrite information on screening card

Hospital clinical staff will handwrite identity information of the newborn on the screening card..

3.	Actor: Hospital clinical staff	Action:	Enter kit number into child's EHR.

Staff performing the heel stick sample will enter the kit number, into the child's EHR. This number is later abstracted by the birth certificate clerk and entered on the facility worksheet.

4.	Actor: Hospital clinical staff	Action:	Obtain blood spot on screening card

Clinical staff perform a heel stick and extract blood spot specimens directly onto the kit card for analysis. This is done within 48 hours and 5 days after birth, and before the infant is discharged from the hospital.

5.	Actor: Hospital clinical staff	Action:	Mail completed kit to state lab

Completed kit with blood spot specimens is mailed to the state lab for analysis.

6.	Actor: State laboratory staff	Action:	Receive completed kit in mail

Completed newborn screening kits are received in the mail at the state health department's laboratory for processing.

7.	Actor: State laboratory staff	Action:	Create identity in Laboratory Information Management System (LIMS)

For each initial screening kit received, state laboratory staff manually create a new child-specific record in LIMS by keying information as it is printed on the kit card. This identity typically, but not always, reflects the placeholder identity assigned in the hospital EHR, such as "Baby Boy Jane Doe" for a child born to mother Jane Doe.

8.	Actor: State laboratory staff	Action:	Analyze bloodspot specimens

In the laboratory, bloodspot specimens are analyzed for 37 different endocrine, metabolic, and hematologic conditions.

9.	Actor: State laboratory staff	Action:	Record initial screening results in LIMS

Results for each screening test are recorded in LIMS.

10.	Actor: State laboratory staff	Action:	Return results to hospital of birth

State laboratory staff report screening results by transfer of electronic batch files of results.

11.	Hospital admin staff	Action:	Import results into EHR and link by kit number

Upon receipt of results files from the state laboratory, hopsital staff upload the file and update results in the hospital's EHR. Results are linked to children by the unique kit number.

Immunization process model

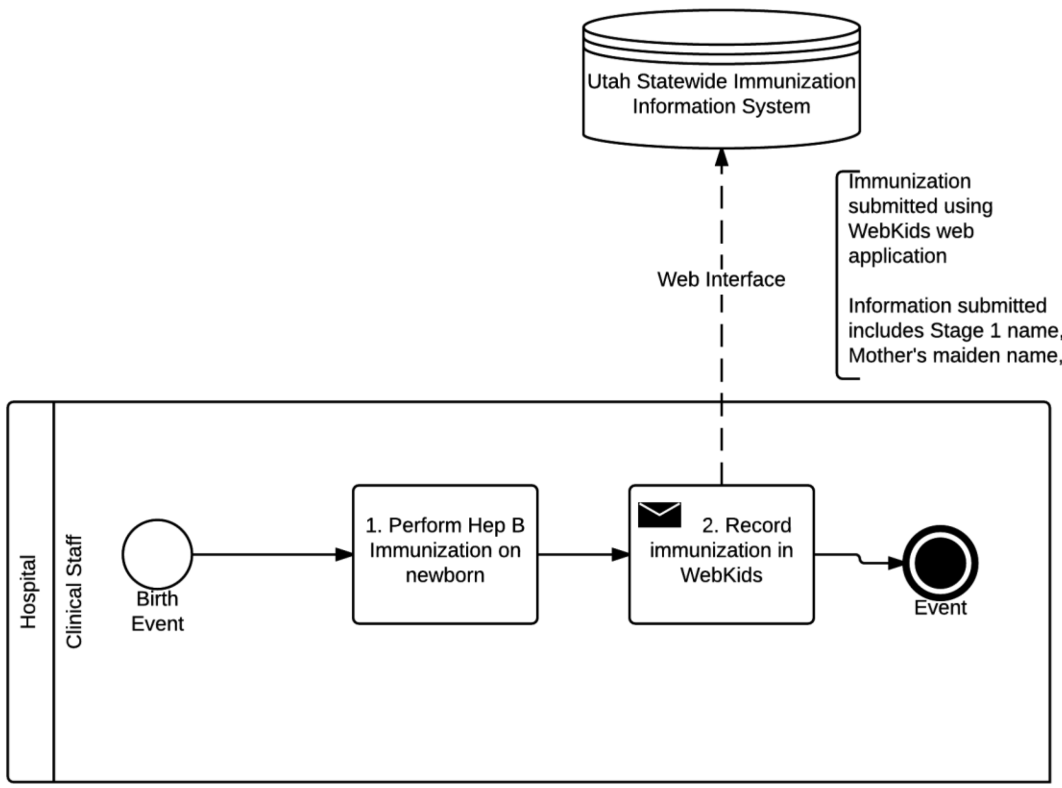

Figure A.5 Newborn immunization process model

<u>Step-by-step narrative</u>

1.	Actor: Hospital clinical staff	Action:	Perform hepatitis B immunization on newborn

Following the birth of a child, a member of the hospital clinical staff will administer a Hep B vaccination.

2.	Actor: Hospital clinical staff	Action:	Record immunization in WebKids

Hospital clinical staff will record the vaccination in WebKids, the web portal for Utah's Statewide Immunization Information System (IIS).

Newborn hearing screening process model

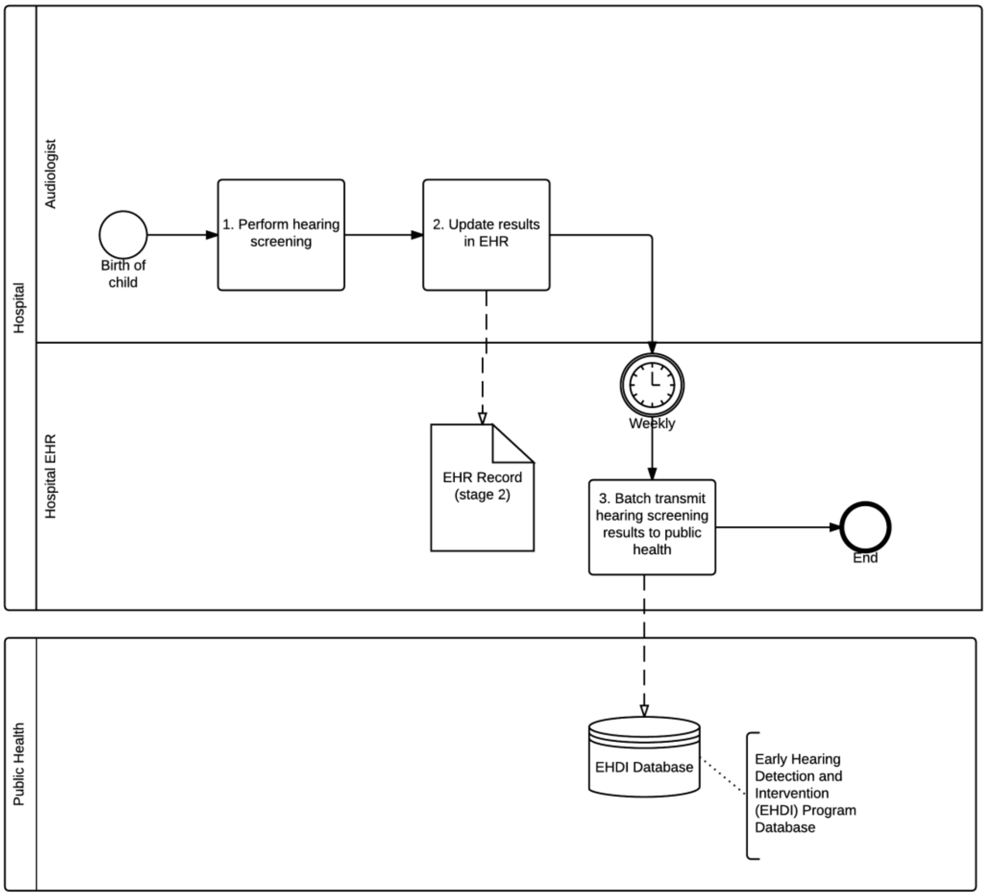

Figure A.6 Newborn hearing screening process model

Step-by-step narrative

1. Actor: Audiologist — Action: Perform newborn hearing screening

Following birth and prior to discharge, an audiologist will perform an initial hearing screening on each newborn.

2. Actor: Audiologist — Action: Update screening results in EHR

The audiologist will record hearing screening results in each child's EHR.

3. Actor: EHR system — Action: Report hearing screenings to state.

The EHR system produces a batch file that is transmitted to the state's Early Hearing Detection and Intervention (EHDI) Program database where the information is used to coordinate followup screening for children who fail the initial screen.

Post discharge birth name update

Once a child is discharged, the placeholder name in the EHR must be updated with the child's legal name. At Intermountain this process involves updating the hospital's EMPI from name information submitted to the state.

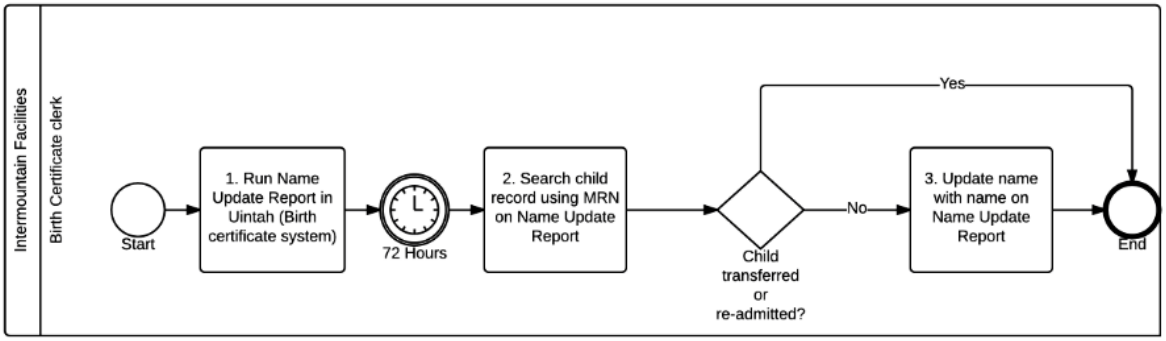

Figure A.7 Post-discharge birth name update process model

<u>Step-by-step narrative</u>

1. Actor: Birth certificate clerk — Action: Run Name Update Report in Uintah

The birth certificate clerk runs and prints a report from the state's Uintah birth certificate program that lists given names for each baby born during a selected time period.

2. Actor: Birth certificate clerk — Action: Search child using MRN on the Name Update Report

Birth certificate clerk enters searches for children on the name update report by the MRN that was keyed into Uintah on the birth certificate. If the child has been transferred to another Intermountain facility or re-admitted, the clerk will not update the name.

3. Actor: Birth Certificate clerk — Action: Update name

The birth certificate clerk will update the name in Intermountain's eMPI with the child's name as reported on the birth certificate.

APPENDIX B

PROCESS MODELS FOR CHILD IDENTIFICATION IN PUBLIC HEALTH INFORMATION SYSTEMS

Introduction

The purpose of this document is to identify specific events, workflows, data exchanges, and other processes surrounding the creation of and subsequent recording of a newborn child's identity following birth in a hospital and the subsequent reporting of those events to public health.

The goal of each process model is to identify the following:

- Events: For the purpose of these models, an event is an action or occurrence in a hospital or public health setting that results or may result in the recording, changing, merging, or transmission of identity information from.
- Actions and actors: specific tasks required to record, verify, modify or transmit identity information
- Workflows: what is sequence of events after birth that the record is created
- Data artifacts: What identifying information is stored, modified or transmitted for each event?
- Exchange methods and standards: How are data exchanged between entities? What standards are used?

List of models

Table B.1. Listing of all process models and accompanying narrative by page location.

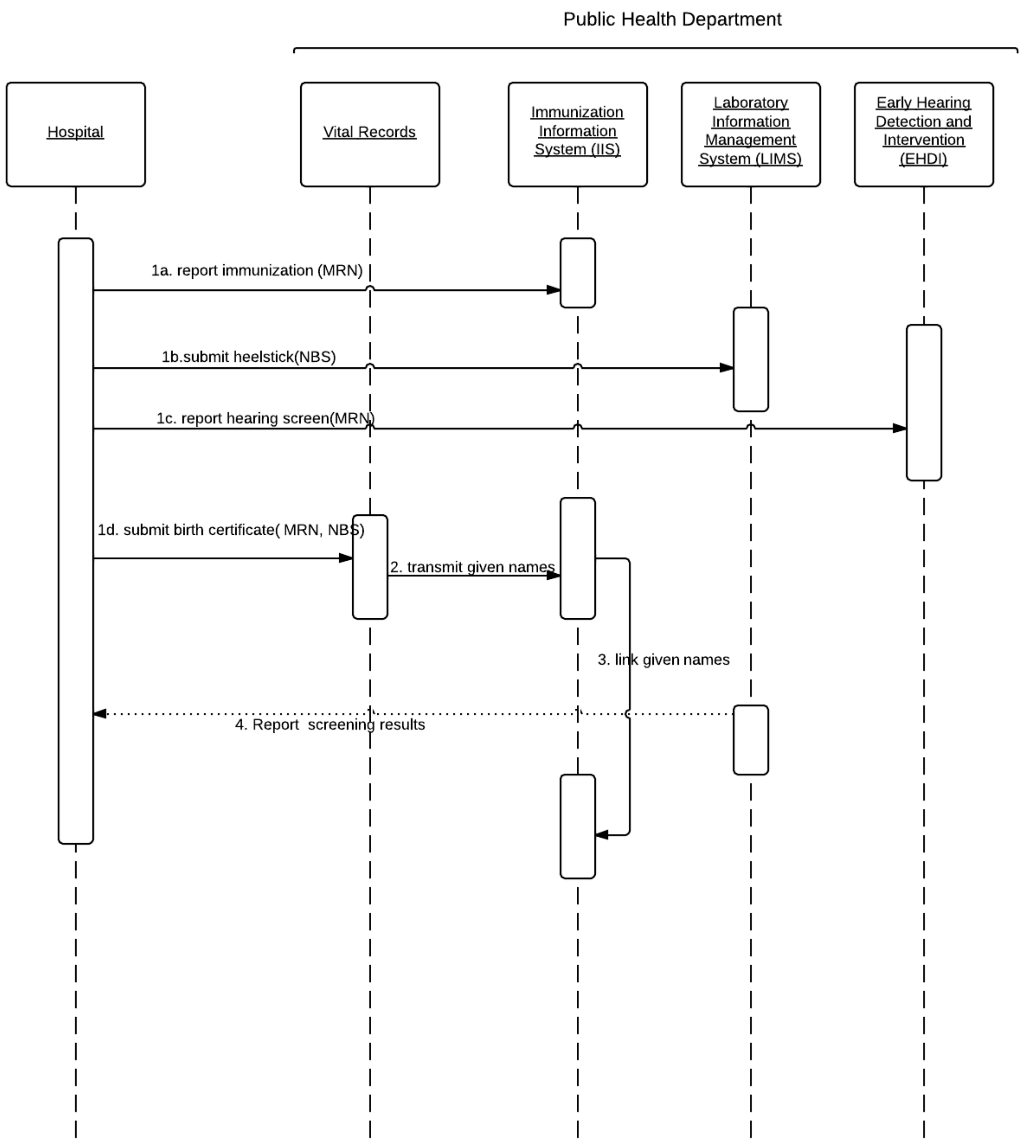

Figure B.1 Interaction diagram showing flow of information between hospital and public health following birth of a child

Birth registration process model (public health perspective)

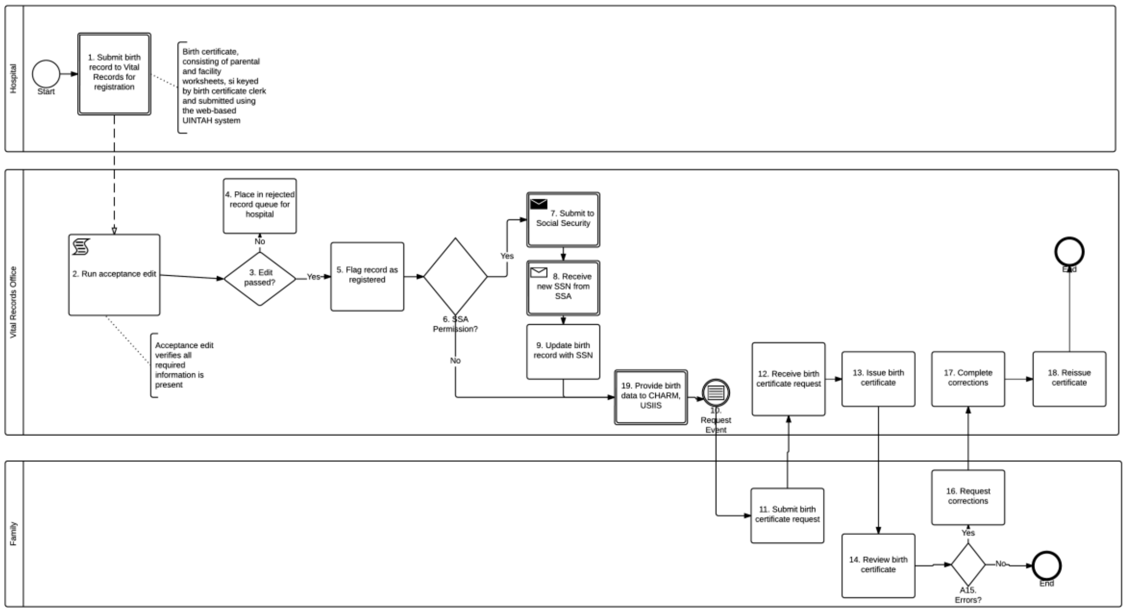

Figure B.2 Birth registration process model (public health perspective)

<u>Step-by-step narrative</u>

1.	Actor: Hospital admin staff	Action:	Submit birth record to Vital Records for registration

Hospital staff key demographic and health information from two worksheets—the parental worksheet and facility worksheet. (This workflow is defined in the hospital birth process model.) Information is keyed into UINTAH, Utah's web-based electronic birth certificate program. Once the record is complete it is submitted for registration by the state. The State Filenumber, the unique identifier for a birth record, is assigned when the record is created.

2.	Actor: Vital Records Office	Action:	Run acceptance edit

Staff in Vital Records daily run an "acceptance edit" on birth records awaiting registration. The edit is a process in UINTAH that determines if each record contains the minimum required information for registration, including date and time of birth, sex of child, name, date of birth of mother.

3.	Event: Acceptance edit	Results:	Record either passes or fails acceptance edit

Records that do not contain the minimum required information for registration are rejected.

4.	Actor: Vital Records office	Action:	Place record in rejected queue

The Vital Records office will place records that fail the acceptance edit into a hospital-specific queue for correction by hospital staff.

A5.	Actor: Vital Records office	Action:	Flag record as registered

Once a record passes the acceptance edit the system will flag it as registered and record the date and time of registration. The result of the registration process is a legal birth certificate that may be issued to authorized requestors. Any changes to the birth certificate after registration require an amendment to be recorded.

6.	Decision: SSA Permission	Results:	Decide whether to submit record to SSA

On the parental worksheet completed in the hospital, parents of a newborn may opt for Enumeration at Birth, a program sponsored by the Social Security Administration to assign social security numbers to newborn children. If parents opt in, the state vital records agency electronically submits birth certificate information to SSA and an SSN is issued for the newborn child. An SSN card is mailed to the parents of the child and the child's SSN is provided back to the state vital records agency.

7.	Actor: Vital Records Staff	Action:	Transmit information to SSA.

Several times a week, Vital Records employees submit batch files to SSA containing enumeration at birth information for children whose parents opt in to the program.

8.	Actor: Vital Records Staff	Action:	Receive SSN updates from Social Security Administration

At regular intervals, Vital Records staff wil download batch files from SSA containing newly issued social security numbers assigned to children born in the state.

9.	Actor: Vital Records Staff	Action:	Update SSNs.

After receiving batch files from SSA, Vital Records staff run an update process to add the information to each birth record in the birth certificate database.

10.	Event: Request Event	Results:	A request to issue a birth certificate is received

A request event may occur at some unknown time interval after a birth is registered. Typically parents will request birth certificates for newborns but it is not unusual for a child's certificate to remain unissued for years.

11.	Actor: Parent or authorized person		Action: Submit birth certificate request

At some point a parent or other person authorized by law may request a certified paper copy of the birth certificate. This request may be made in person, on the Internet, or by mail.

12.	Actor: Vital Records Staff		Action: Receive birth certificate request

Vital records staff receive requests for birth certificates in person, by mail or through an Internet applications. Requests are verified to be from persons authorized by law to receive the certificate and a search fee is collected.

13.	Actor: Vital Records staff		Action: Issue birth certificate

After verifying eligibility and payment of a fee, the paper birth certificate is printed on certified paper and issued to the parent or authorized person.

14.	Actor: Parent or authorized		Action: Review birth certificate

	person	

Receipt of the paper copy is the first opportunity parents or others may review what is in the birth record and identify errors. It is not uncommon to identify spelling errors in the name, missing name, error in the date of birth, or sex.

15.	Decision: Errors on birth certificate?	Results: Person may or may not identify factual errors on birth certificate

Correction of the error depends on the nature of the error, the age of the child, and the length of time since the certificate has been issued. Typically errors must be corrected with an affidavit.

16.	Actor: Parent or authorized person	Action: Request correction of errors

The actual means of correction depends on several factors, however the typical process requires completion of an affidavit. Only minor name changes and spelling errors may be corrected with affidavits, and documentation of actual name spelling, including school, medical or other records, is typically required.

17.	Actor: Vital Records Staff	Action:Complete corrections

When all specific requirements are met, Vital Records staff will make corrections to the birth record in the birth certificate database.

18.	Actor: Vital Records Staff	Action: Reissue certificate

Once corrections are made the certificate is reissued.

19.	Actor: Vital Records	Action: Share birth certificate information with other public health programs

Immunization registry process model

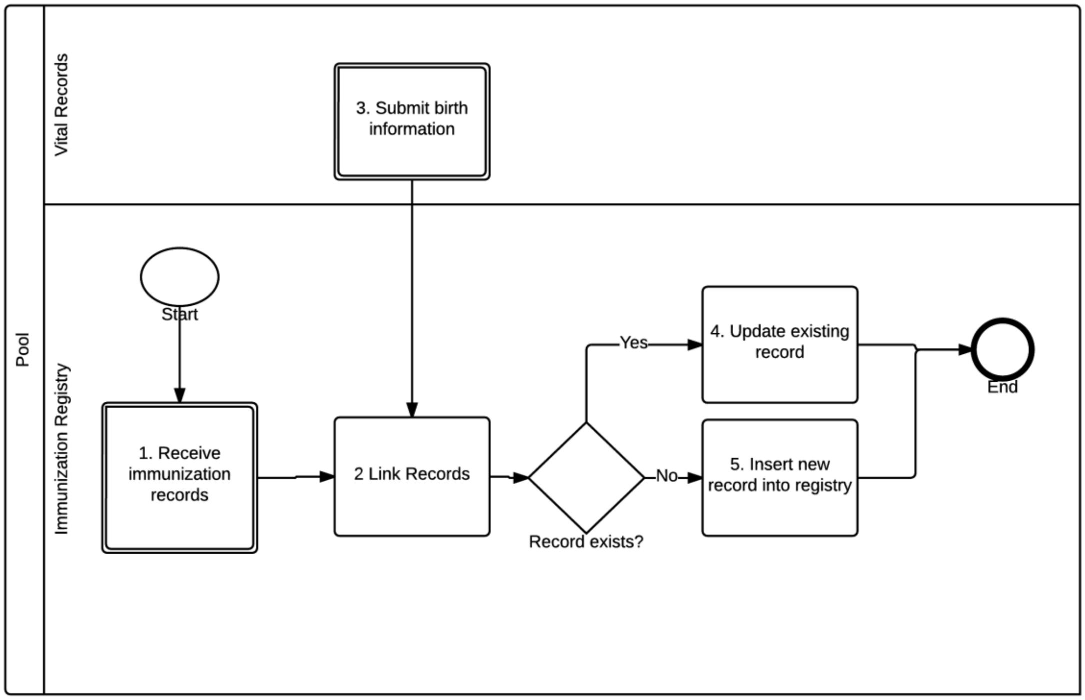

Figure B.3 Immunization registry process model

<u>Step-by-step narrative</u>

1.	Actor: Immunization Registry	Action:	Receive immunization records

The immunization registry routinely receives new immunization records for immunizations delivered by healthcare providers. Immunization records may be delivered individualy via a web interface (WebKids), delivered in real time from EHR to USIIS via HL7 messages, or submitted from EHRs to USIIS in batch mode.

2.	Actor: Immunization Registry	Action:	Link Records

The immunization registry attempts to link all incoming records with existing records. If a record exists, it is updated with the new record. If it does not exist, a new record is created.

3.	Actor: Vital Records	Action:	Submit birth records

On a periodic basis, at least monthly, vital records submits batches of birth records to the Immunization registry. Information from birth records is linked with existing data in the registry and used to improve names and other identifying information in the registry.

4.	Actor: Immunization	Action:	Update existing record

	Registry		

If a record is found in the immunization registry, information is updated with information from the new record.

5.	Actor: Immunization Registry	Action:	Insert new record into registry

If a record is not found in the immunization registry, a new record is created using the incoming information.

Newborn metabolic screening process model

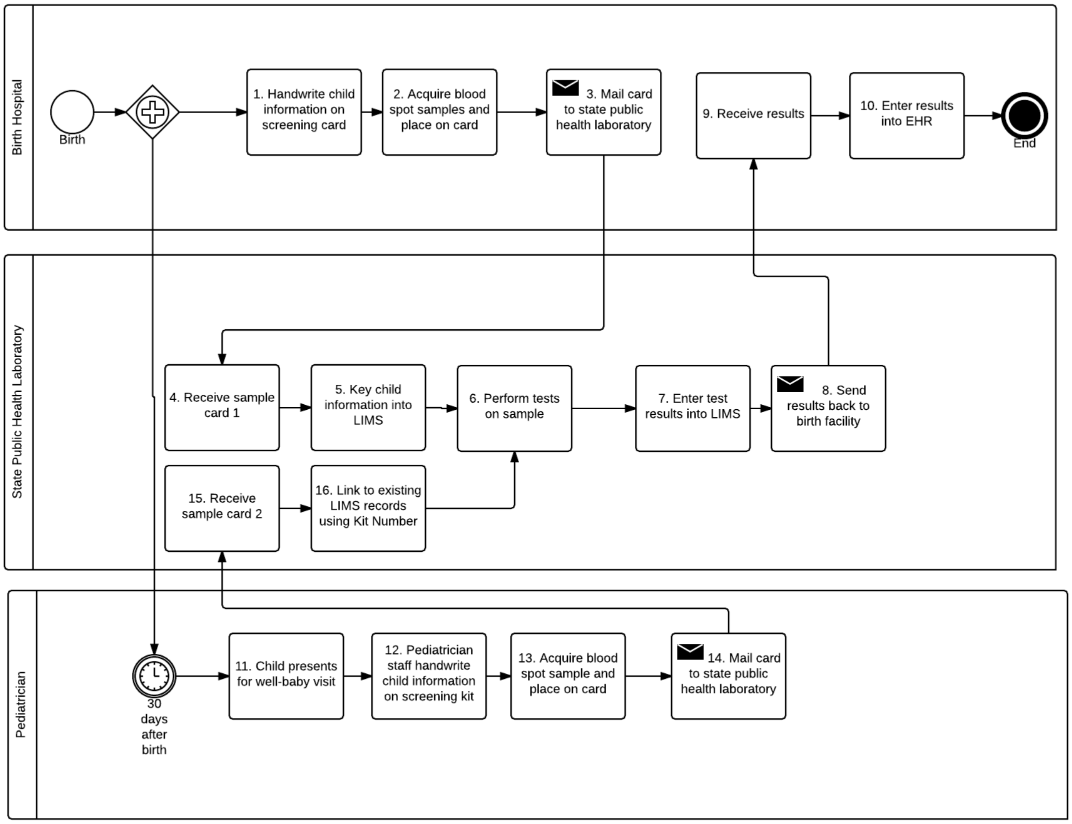

Figure B.4 Newborn metabolic screening process model

<u>Step-by-step narrative</u>

1.	Actor: Birth hospital clinical staff	Action:	Handwrite child demographic information on new screening kit card
	Following birth of a newborn and prior to discharge, clinical staff will take a new, uniquely numbered kit consisting of two sample cards. On one card staff will handwrite the child's demographic information, including name, sex and date of birth. This information is typically the placeholder name assigned by the hospital.		
2	Actor: Birth hospital clinical staff	Action:	Acquire blood spot and place on card
	Hospital clinical staff will perform a heelstick to acquire small blood samples from a newborn and will affix those blood spots to the card completed in the previous step.		
3	Actor: Birth hospital clinical staff	Action:	Mail completed kit to state health department newborn screening laboratory

After affixing blood spots to kit cards, clinical staff will send kits for evaluation to a designated laboratory that will perform analysis of the blood spots to screen for metabolic conditions.

4	Actor: public health laboratory staff	Action:	Receive completed kit card in mail

Sample 1 cards are received via regular mail by staff in designated public health laboratories

5.	Actor: public health laboratory staff	Action:	Key sample information into LIMS (laboratory information management system)

Laboratory staff key information from sample cards into LIMS system. Information include kit number, child name, sending facility, and other information that was written by hospital staff on the card.

6	Actor: public health laboratory staff	Action:	Perform analyses on samples

Laboratory staff conduct metabolic screenings on bloodspot samples , screening for a variety of metabolic and inherited conditions as prescribed by law.

7	Actor: public health laboratory staff	Action:	Enter results into LIMS

Laboratory staff enter screening results into LIMS. Depending on the LIMS system and laboratory equipment used, the level of manual effort varies.

8	Actor: public health laboratory staff	Action:	Return screening results to sending facility

Depending on the LIMS system and exchange procedures in place, results may be returned electronically using standard HL7 messages, or they may be returned by other methods such as fax.

9	Actor: Facility administrative staff	Action:	Receive screening results

In facilities using electronic messaging, results are received and updated in the hospital system automatically. If results are received via fax, administrative staff must manually locate the child's record and update screening results.

10	Actor: Facility administrative staff	Action:	Update screening results in EHR

When results are received via fax, administrative staff must manually locate a child's EHR and update results. Records may be located using the kit number, child's name, or a combination. Success of this update depends on the quality of identity information that was originally written on the card, and keyed into the LIMS system.