# Building an Ontology for Identity Resolution in Healthcare and Public Health

**Jeffrey Duncan, MS[1*], Karen Eilbeck, PhD[1], Scott P. Narus, PhD[1,3], Stephen Clyde[2], Sidney Thornton, PhD[1,3], Catherine Staes, PhD[1]**

1. Department of Biomedical Informatics, University of Utah, Salt Lake City, UT USA

2. Department of Computer Science, Utah State University, Logan, UT USA

3. Intermountain Healthcare, Salt Lake City, UT USA

**Abstract**

Integration of disparate information from electronic health records, clinical data warehouses, birth certificate registries and other public health information systems offers great potential for clinical care, public health practice, and research. Such integration, however, depends on correctly matching patient-specific records using demographic identifiers. Without standards for these identifiers, record linkage is complicated by issues of structural and semantic heterogeneity.

**Objectives:** Our objectives were to develop and validate an ontology to: 1) identify components of identity and events subsequent to birth that result in creation, change, or sharing of identity information; 2) develop an ontology to facilitate data integration from multiple healthcare and public health sources; and 3) validate the ontology's ability to model identity-changing events over time.

**Methods:** We interviewed domain experts in area hospitals and public health programs and developed process models describing the creation and transmission of identity information among various organizations for activities subsequent to a birth event. We searched for existing relevant ontologies. We validated the content of our ontology with simulated identity information conforming to scenarios identified in our process models.

**Results:** We chose the Simple Event Model (SEM) to describe events in early childhood and integrated the Clinical Element Model (CEM) for demographic information. We demonstrated the ability of the combined SEM-CEM ontology to model identity events over time.

**Conclusion:** The use of an ontology can overcome issues of semantic and syntactic heterogeneity to facilitate record linkage.

**Keywords: ontology, medical record linkage, integrated child health information systems**

**Abbreviations:** Health Information Exchange (HIE), U.S. Office of the National Coordinator for Health Information Technology (ONC), Universal newborn hearing screening (UNHS), newborn metabolic screening (NBS), Business Process Modeling Notation (BPMN), Electronic Health Record (EHR), Utah Statewide Immunization Information System (USIIS), enterprise master person index (EMPI).

**Correspondence:** jeff.duncan@utah.edu

**DOI:** 10.5210/ojphi.v7i2.6010

## 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 apublic 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 non-unique 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 among healthcare and public health entities for post-birth activities. We created process models using Business Process Modeling Notation (BPMN) [29] with the goal of documenting specific post-birth 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 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 1 depicts a high-level process model derived from these interviews. All of the process models created are included as supplemental materials.

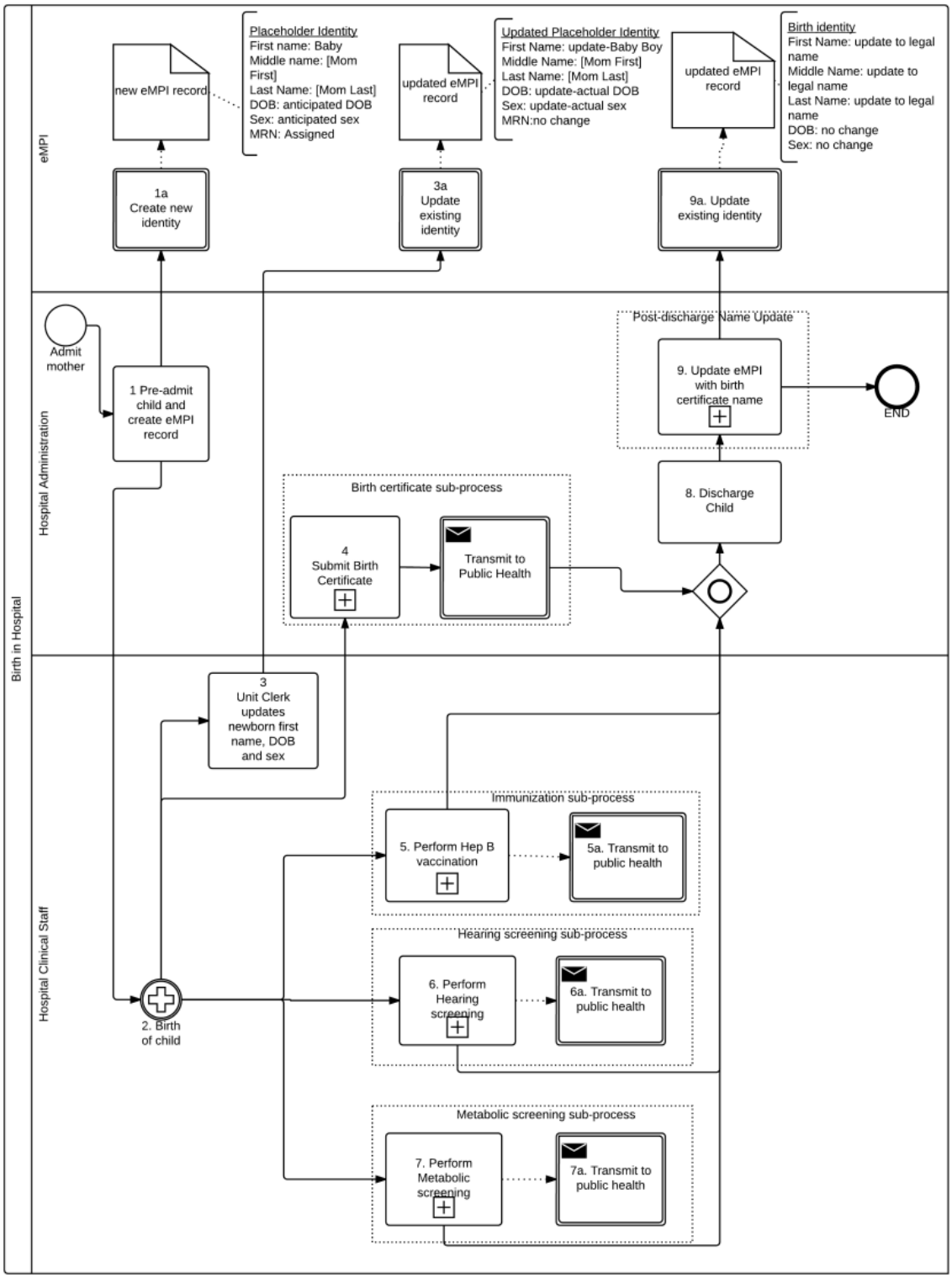


Figure 1. High-level process model for birth-related events in a hospital using BPMN

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 pre-registration, 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 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 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.

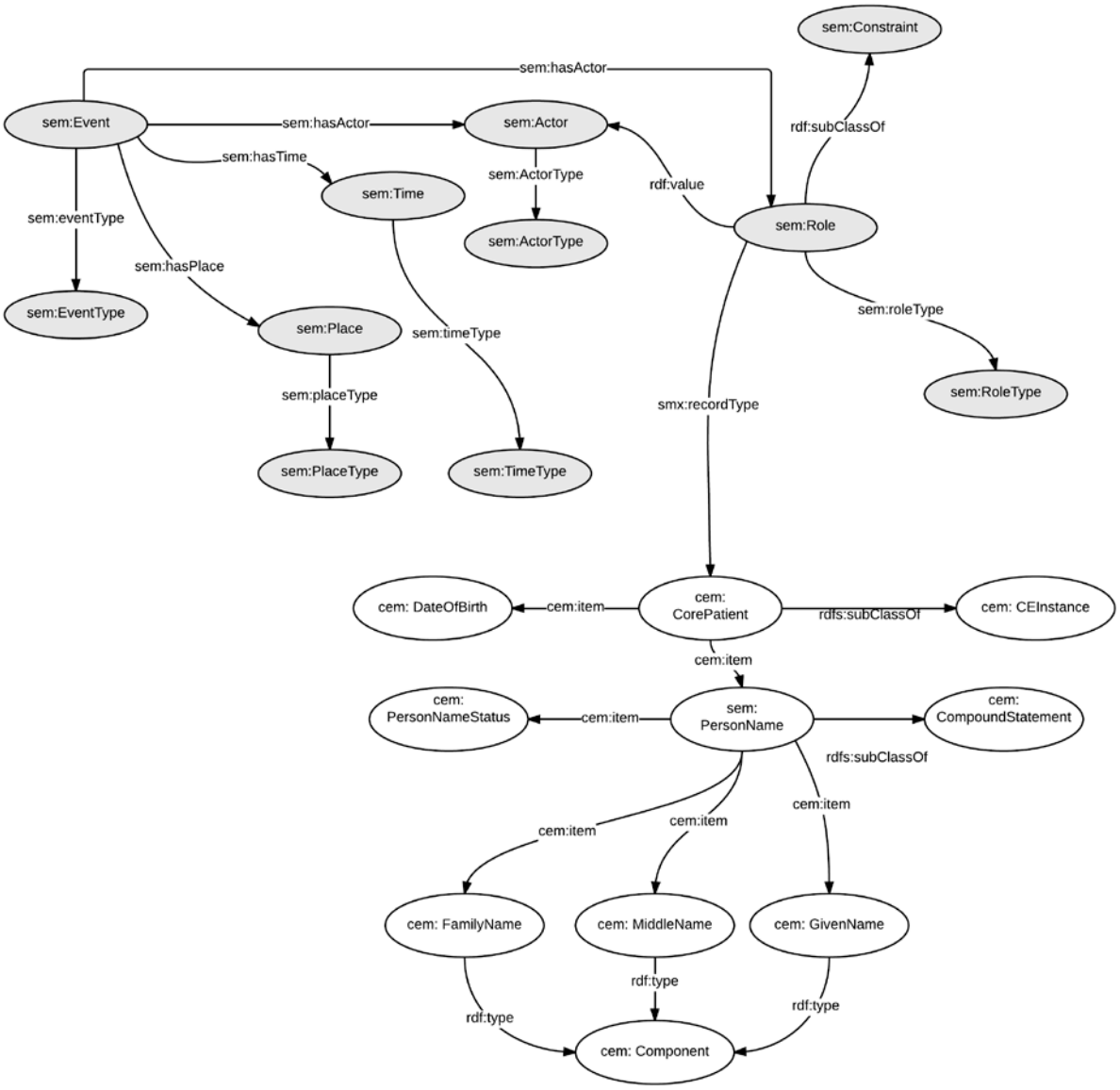


Figure 2. High-level overview of the combined SEM-CEM ontologies. (Classes are represented by ovals and relationships are represented by arrows.)

Figure 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.

After creating simulated instances in triple stores, we conducted both structural and functional validations of the combined SEM-CEM ontology [41]. Our structural evaluation was completed using the Pellet OWL2 Reasoner in Protégé to validate the classes and properties, and individuals [42].

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 1 describes the events, actors and places that were modeled.

Table 1. Information system events, actors, and places modeled in SEM-CEM Ontology

| Event Name | Place | Actors | Comments |
|---|---|---|---|
| BirthRegistrationEvent | Birth Registry | Child<br>Mother<br>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<br>EHR<br>IIS | Immunization recorded in EHR are reported to IIS in real-time messages or in batches |
| NewbornScreening ReportEvent[1] | | Child<br>Birth Facility<br>Laboratory | Birth facility submits blood spot and identifying information to laboratory for analysis. This is typically a manual process. |
| NewbornScreening Results ReportEvent[1] | | Child | Reporting results back to the source hospital may be done electronically |

| | | | |
|---|---|---|---|
| | | Laboratory EHR | or manually with a fax |
| HearingScreening RecordEvent | EHR | Child | |
| HearingScreening ReportEvent[1] | | Child<br>EHR<br>EHDI | EHDI = Early Hearing Detection and Intervention system |
| PaternityEvent | Birth registry | Child | |
| AdoptionEvent | Birth registry | Child<br>Child 2 | The original record is sealed<br>A new child record is created, using the original child's State File Number (unique identifier) |
| DeathRegistrationEvent | Death Registry | Decedent | |
| DeathReportEvent[1] | | Death Registry<br>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 |
| 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<br><br>RecordSplitEvent | | Person1<br>Person2<br>Person1<br>Person2 | A record repository such as an EMPI may merge multiple identities into one, or may split one into multiple |

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

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 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 4.

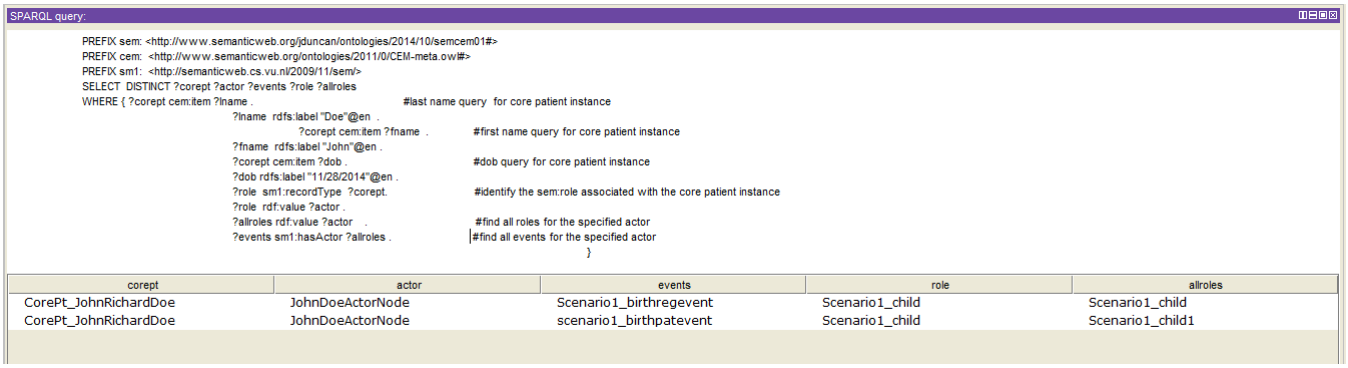


Figure 3. SPARQL query returns associated actors, roles, and events for an individual named John Doe, born 11/28/2014.

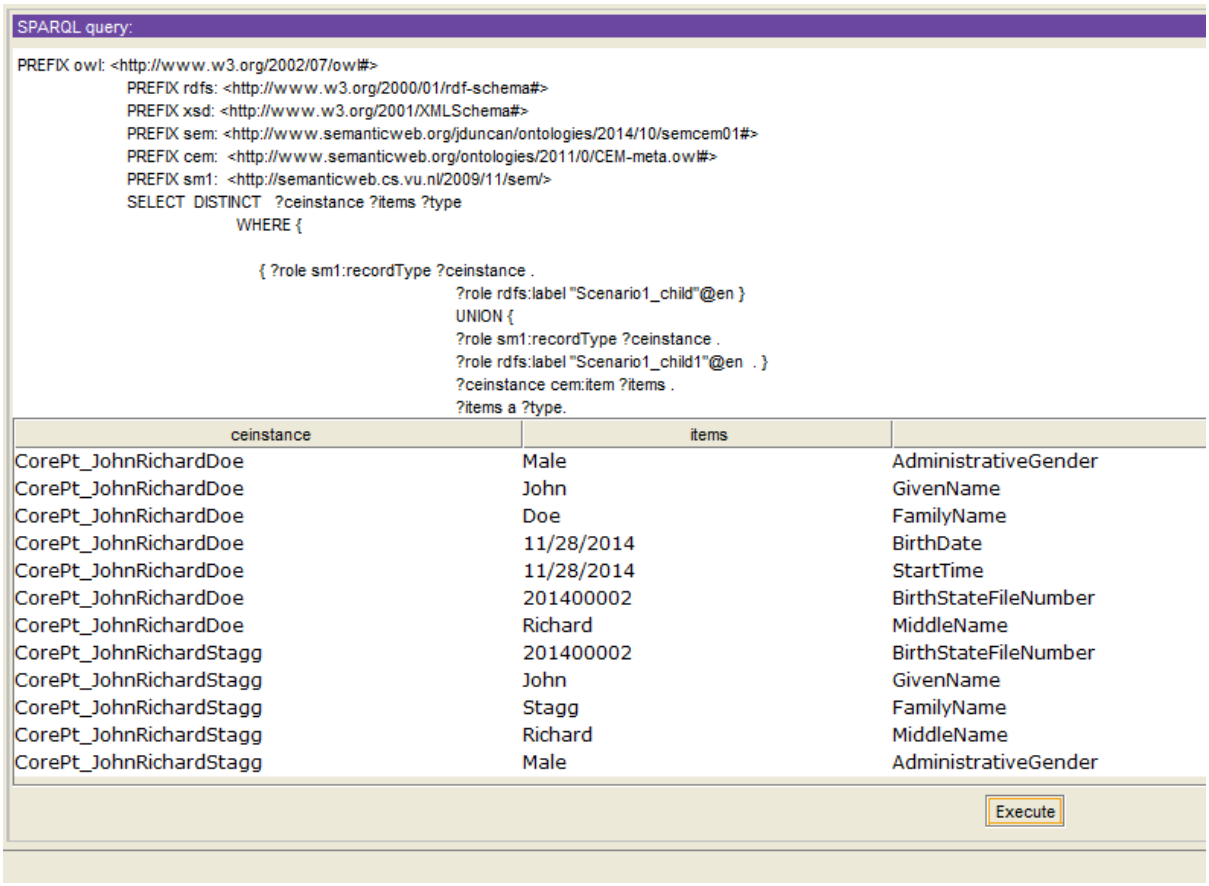


Figure 4. SPARQL query returns identity items and their corresponding types for the two CEM instances identified in Figure 3.

### *Additional strengths of model*

The approach used to model names in CEM, as depicted in Figure 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 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 [43].

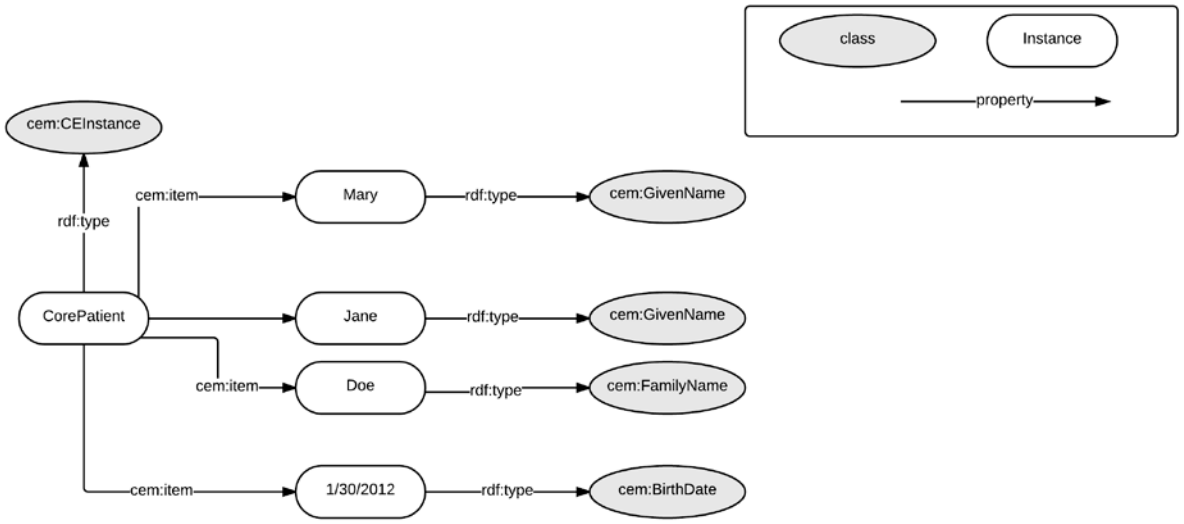


Figure 5. Example of the modeling of identity properties in SEM-CEM.

## 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 [44]. 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 un-merging 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 un-merged 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.

## References

1. 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.
2. Wache H, Voegele T, Visser U, Stuckenschmidt H, Schuster G, et al., eds. Ontology-based integration of information-a survey of existing approaches. IJCAI-01 workshop: ontologies and information sharing; 2001: Citeseer.
3. Fellegi IP, Sunter AB. 1969. A theory for record linkage. *J Am Stat Assoc*. 64(328), 1183-210. http://dx.doi.org/10.1080/01621459.1969.10501049
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. 2002. Analysis of identifier performance using a deterministic linkage algorithm. *Proc AMIA Symp*. 2002, 305-09. PubMed
6. Clark DE, Hahn DR, eds. 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; 1995: American Medical Informatics Association.
7. Campbell KM, Deck D, Krupski A. 2008. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and abasic'deterministic algorithm. *Health Informatics J*. 14(1), 5-15. PubMed http://dx.doi.org/10.1177/1460458208088855
8. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. 2011. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 64(5), 565-72. PubMed http://dx.doi.org/10.1016/j.jclinepi.2010.05.008
9. McClellan MA, ed. Duplicate medical records: a survey of twin cities healthcare organizations. AMIA Annual Symposium Proceedings; 2009: American Medical Informatics Association.
10. Duncan J, Narus SP, Clyde S, Eilbeck K, Thornton S, et al. 2015. Birth of identity: understanding changes to birth certificates and their value for identity resolution. *J Am Med Inform Assoc*. 22(e1), e120-29. PubMed
11. Martin DW, Lowery NE, Brand B, Gold R, Horlick G. 2015. Immunization information systems: a decade of progress in law and policy. Journal of public health management and practice. *J Public Health Manag Pract*. 21(3), 296-303. PubMed http://dx.doi.org/10.1097/PHH.0000000000000040
12. Yasuda K. 2006. Immunization information systems. *Pediatrics*. 118(3), 1293-95. PubMed http://dx.doi.org/10.1542/peds.2006-1723

13. Halpin KS, Smith KY, Widen JE, Chertoff ME. 2010. 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*. 21(3), 169-75. PubMed http://dx.doi.org/10.3766/jaaa.21.3.5

14. Wrightson AS. 2007. Universal newborn hearing screening. *Am Fam Physician*. 75(9), 1349-1352. PubMed

15. Yoshinaga-Itano C. 2004. Levels of evidence: universal newborn hearing screening (UNHS) and early hearing detection and intervention systems (EHDI). *J Commun Disord*. 37, 451-65. PubMed http://dx.doi.org/10.1016/j.jcomdis.2004.04.008

16. Kim S, Lloyd-Puryear MA, Tonniges TF. 2003. Examination of the communication practices between state newborn screening programs and the medical home. *Pediatrics*. 111(2), E120-26. PubMed http://dx.doi.org/10.1542/peds.111.2.e120

17. Hinman AR, Atkinson D, Diehn TN, Eichwald J, Heberer J, et al. 2004. Principles and core functions of integrated child health information systems. *J Public Health Manag Pract*. 10, S52-6. PubMed http://dx.doi.org/10.1097/00124784-200411001-00008

18. Hinman AR, Eichwald J, Linzer D, Saarlas KN. 2005. Integrating child health information systems. *Am J Public Health*. 95(11), 1923-27. PubMed http://dx.doi.org/10.2105/AJPH.2004.051466

19. Duncan J, Xu W, Narus SP, Nangle B, Thornton S, et al. 2013. A Focus Area Maturity Model for a Statewide Master Person Index. *Online J Public Health Inform*. 5(2), 210. PubMed http://dx.doi.org/10.5210/ojphi.v5i2.4669

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

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

22. Soldatova LN, Nadis D, King RD, Basu PS, Haddi E, et al. 2014. EXACT2: the semantics of biomedical protocols. *BMC Bioinformatics*. 15(Suppl 14), S5. PubMed http://dx.doi.org/10.1186/1471-2105-15-S14-S5

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

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. Jafarpour B, Abidi S, Abidi S. Exploiting Semantic Web Technologies to Develop OWL-Based Clinical Practice Guideline Execution Engines. Biomedical and Health Informatics, IEEE Journal of. 2014;PP(99):1-.

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. 1996. Ontologies: Principles, methods and applications. *Knowl Eng Rev*. 11(02), 93-136. http://dx.doi.org/10.1017/S0269888900007797

29. Mendling J, Weidlich M, Weske M, eds. 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 10-11-2103]. 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, eds. 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. 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. 2011. Design and use of the Simple Event Model (SEM). *Web Semant*. 9(2), 128-36. http://dx.doi.org/10.1016/j.websem.2011.03.003

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. 2003. Standards for detailed clinical models as the basis for medical data exchange and decision support. *Int J Med Inform*. 69(2-3), 157-74. PubMed http://dx.doi.org/10.1016/S1386-5056(02)00103-X

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. Journal of the American Medical Informatics Association. 2012:amiajnl-2012-001326.

41. Gangemi A, Catenacci C, Ciaramita M, Lehmann J, eds. A theoretical framework for ontology evaluation and validation. Semantic Web Application Platform (SWAP); 2005.

42. Parsia B, Sirin E. Pellet: An owl dl reasoner. Third International Semantic Web Conference-Poster2004.

43. Karakasidis A, Verykios VS, eds. Privacy preserving record linkage using phonetic codes. Informatics, 2009 BCI'09 Fourth Balkan Conference in; 2009: IEEE.

44. Grannis SJ, Overhage JM, Hui S, McDonald CJ, eds. Analysis of a probabilistic record linkage technique without human review. AMIA annual symposium proceedings; 2003: American Medical Informatics Association. Supplementary Material A. Identity Process Models for Births in Hospitals B. Identity Process Models in Public Health