

Computable Records: The next generation of the EMR conversation

Sanna Rimpilainen

Document reference number	DHI+DDMMYY+doctype+000X <i>DHI211215RR0001</i> <ul style="list-style-type: none"> ○ E = exploratory report ○ L = lab report ○ F = factory report ○ S = summary document ○ LR = literature review ○ RR = research report ○ MR = market research ○ MAP = mapping ○ V=video ○ O= other
----------------------------------	--

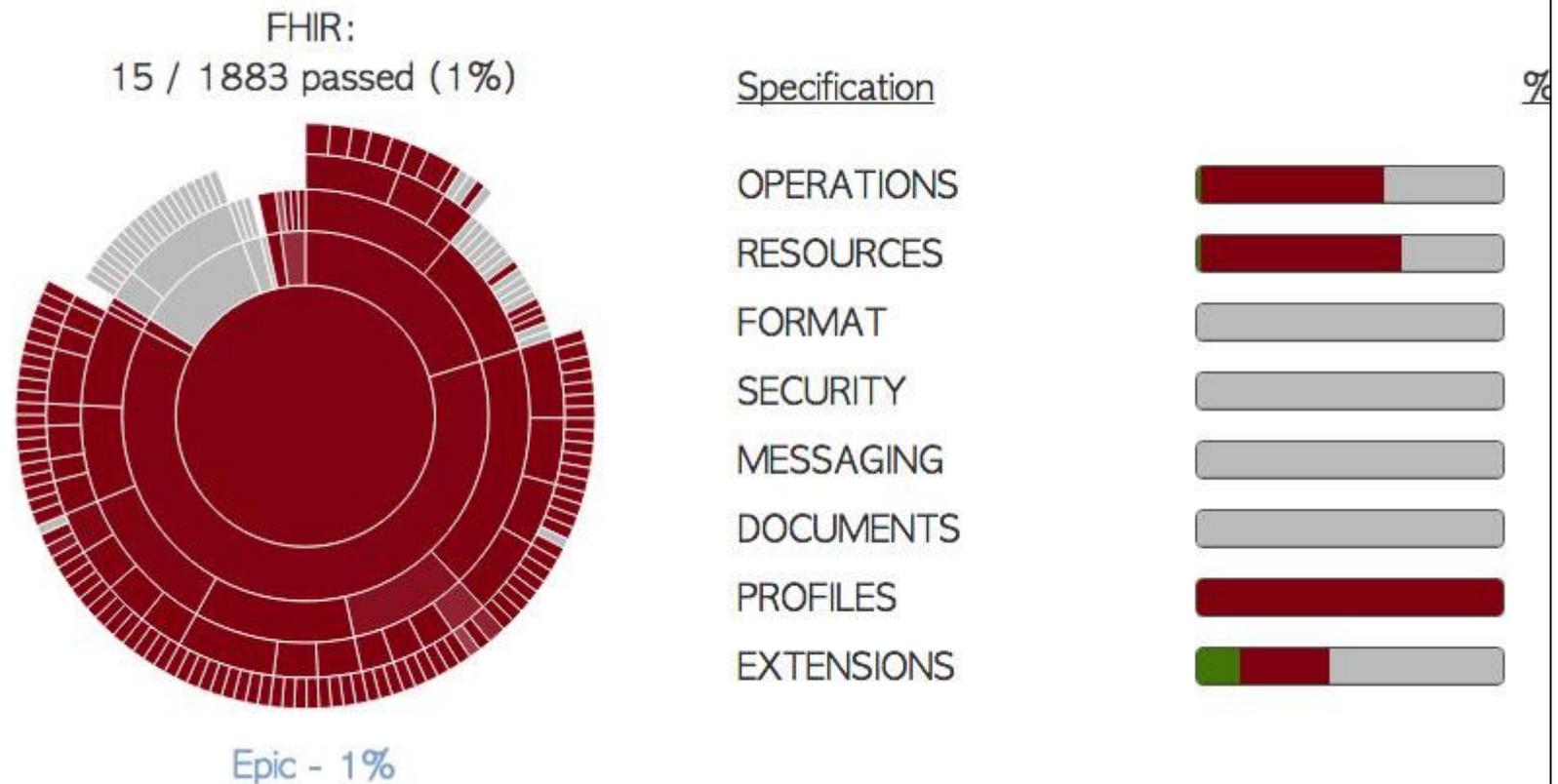
Publication date	21/12/2015
Revision date	
Revision number	

Purpose of document	Research piece into utilising computable records for EHRs
Other detail (delete row if appropriate)	

Related projects	Names and doc reference numbers
Keywords	Electronic health records; EHRs; computable phenotyping;

Follet and
Herlin
2015

6. Computable Records: The next generation of the EMR conversation.



In 2016 and onward, computable medical records will fuel the next generation of EHRs, as the quest for interoperable, portable, and comprehensive health data continues.

Computable medical records, readable by both human and machine, will house a patient's entire record from conception to death. Importantly, such records will declare their fidelity level — their degree of completeness and accuracy — so that users can not only identify what data is there, but also what's missing.

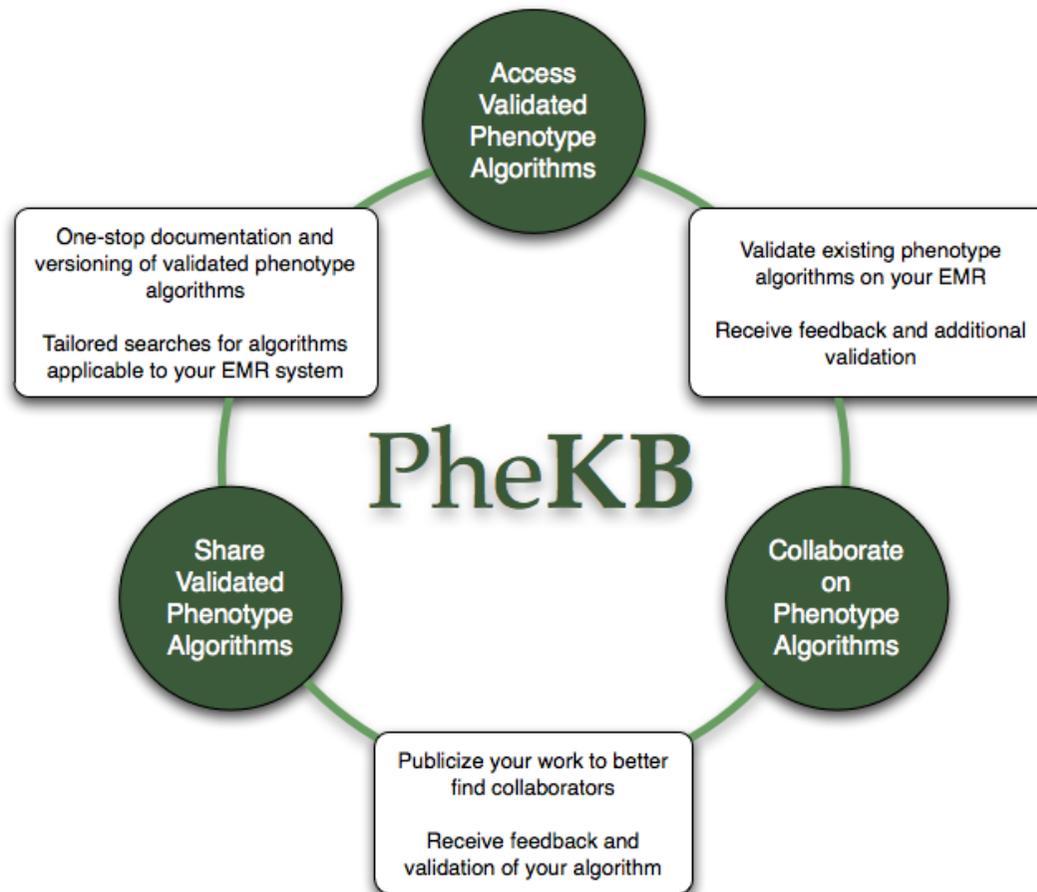
The computable medical record will be unique, enabling users to find the right record for the right person; will support a health status scoring system; and will ideally be open source to drive adoption across software vendors, hospital systems, and government.

FHIR (HL7's latest attempt at a health data exchange) is a step in the right direction. [The Argonauts](#) are working on a handful of profiles and core data services, [according to John Halamka, CIO of Beth Israel Deaconess Hospital](#). They're implementing the API [as recommended](#) by the [MITRE JASON report](#). While a far cry from a 'computable medical record', it's a hopeful signal flare for U.S. progress in digital healthcare.

<http://mobihealthnews.com/content/digital-healthcare-services-2016-and-beyond>

PheKB
2014

What is the Phenotype KnowledgeBase?



Health Data is becoming an increasingly important source for clinical and genomic research. Researchers create and iteratively refine algorithms using structured and unstructured data to better identify cohorts of subjects within the health data.

The Phenotype Knowledgebase website, PheKB, is a collaborative environment to building and validating electronic algorithms to identify characteristics of patients within health data. PheKB was functionally designed to enable such a workflow and has purposefully integrated tools and standards that guide the user in efficiently navigating each of these stages from early stage development to public sharing and reuse. PheKB has tools to enable cross-site collaboration for algorithm development, validation, and sharing for reuse with confidence.

On PheKB you can: [View existing algorithms](#); [Enter or create new algorithms](#); [Collaborate with others to create or review algorithms](#); [View implementation details for existing algorithms](#)

Phenotype algorithms can be viewed by data modalities or methods used:

- ICD and CPT codes
- Laboratories
- Medications
- Vital Signs
- Natural Language Processing

<p>Richesson and Smerek 2015</p> <p>EXCELLE NT HOW TO</p>	<p>What is a phenotype?</p> <p>A <i>phenotype</i> is the observable physical or biochemical expression of a specific trait in an organism, such as a disease, stature, or blood type, based on genetic information and environmental influences. The phenotype of an organism includes factors such as physical appearance, biochemical processes, and behavior. In short, the phenotype of an organism is the appearance it presents to observers.</p> <p>A more contemporary interpretation of the term <i>phenotype</i> is understood as measurable biological (physiological, biochemical, and anatomical features), behavioral (psychometric pattern), or cognitive markers that are found more often in individuals with a disease or condition than in the general population.</p> <p>What is a computable phenotype?</p> <p>A <i>computable phenotype</i> is a clinical condition, characteristic, or set of clinical features that can be determined solely from the data in EHRs and ancillary data sources and does not require chart review or interpretation by a clinician. These can also be referred to as <i>EHR condition definitions</i>, <i>EHR-based phenotype definitions</i>, or simply <i>phenotypes</i>.</p> <p>We use the term EHR broadly to reference data that are generated through healthcare delivery and reimbursement practices; in practice, these functions may be covered in multiple systems and can contain both practice management data and data that are strictly limited to the clinical domain. We use <i>ancillary data sources</i> to refer to sources such as disease registries, claims data, or supplemental data collection that are related to health care delivery but may not be directly integrated into the EHR system.</p>
---	--

What are computable phenotype definitions?

Computable phenotype definitions are specifications for identifying patients or populations with a given characteristic or condition of interest from EHRs using data that are routinely collected in EHRs or ancillary data sources. Computable phenotype definitions can support reproducible queries of EHR data from multiple organizations. These queries can then be replicated at multiple sites in a consistent fashion, enabling efficiencies and also ensuring that populations identified from different healthcare organizations have similar features, or at least were identified in the same way.

Phenotype definitions are composed of [data elements](#) and logic expressions (AND, OR, NOT) that can be interpreted and executed by a computer. In other words, the [syntax](#) defining a computable phenotype is designed to be interpreted and executed programmatically without human intervention. Computable phenotype definitions rely on [value sets](#) derived from standardized coding systems and may employ hierarchies and [weighting factors](#) for data elements. Data elements and the difference between data elements and phenotypes will be described further in this chapter.

Why are computable phenotype definitions important?

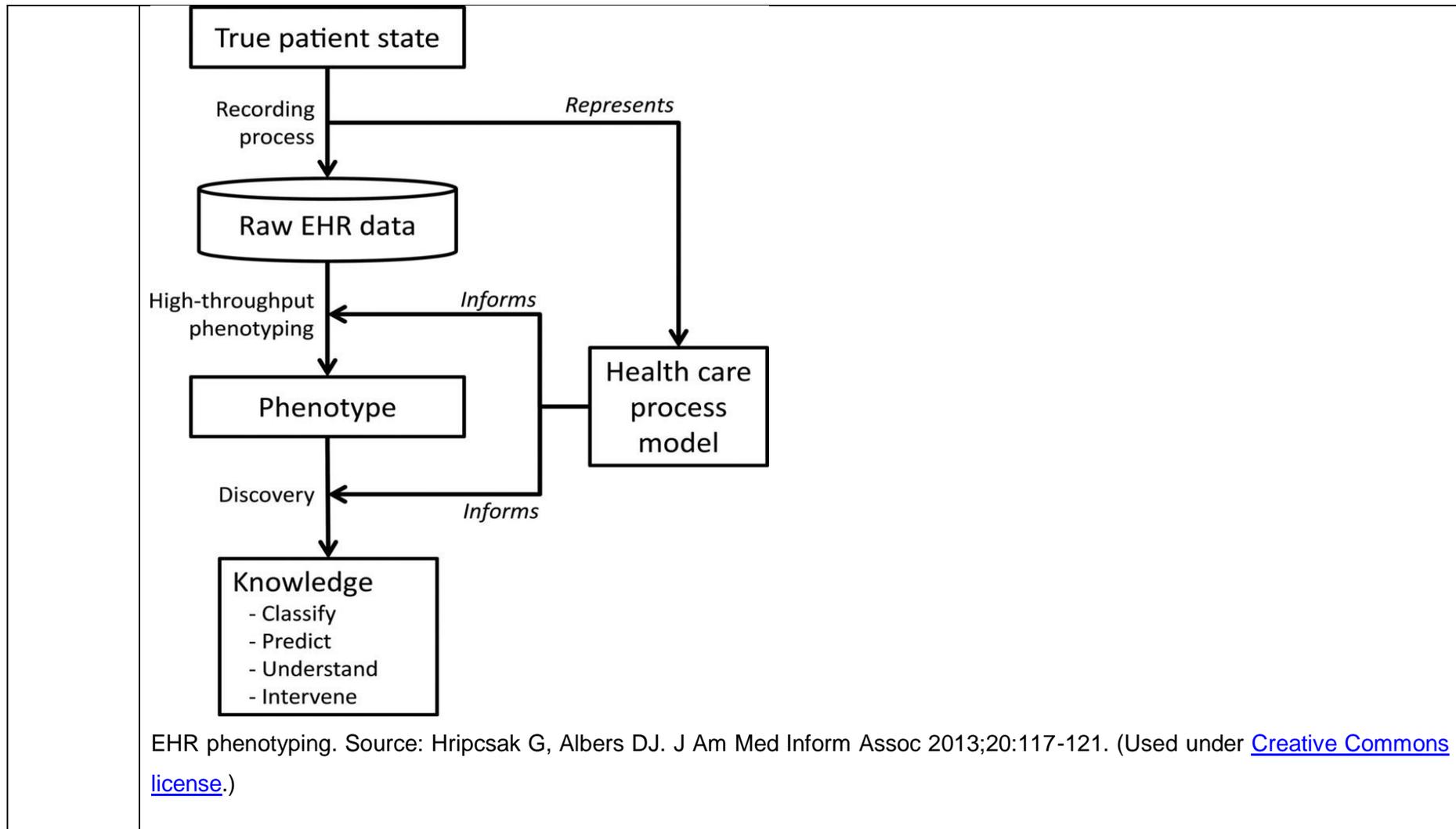
The ability to identify people with particular conditions across healthcare organizations by using common definitions has value for clinical quality measurement, health improvement, and research. Standard phenotype definitions can enable direct identification of cohorts based on population characteristics, risk factors, and complications, allowing decision-makers to identify and target patients for screening tests and interventions that have been demonstrated to be effective in similar populations. This identification process can be integrated with the EHR for real-time [clinical decision support](#).

Standard phenotype definitions can also streamline the development of [registries](#) and applications using healthcare data and can enable consistent inclusion criteria to support regional surveillance in the identification of infectious diseases and rare disease complications.

Finally, computable phenotype definitions are essential to the conduct of pragmatic clinical trials and comparative effectiveness research. These studies, which may involve multiple hospitals or health systems, rely on standard phenotype definitions for EHR-based inclusion/exclusion of participants and consistent data analysis and reporting across data sources. Computable phenotype definitions have applications in interventional, observational, prospective, and retrospective studies [\[1\]](#).

How do computable phenotypes relate to the true presence of a condition?

As shown in the figure below, phenotype definitions are composed of data constructs and coding systems available for providers to record patient data in EHR systems. These data from EHRs may reflect a patient's state or disease status, but the data are generated from the perception, interpretation, and recording by the clinical staff that are observing the patient. The data in EHRs, therefore, represent a limited view of a patient's condition, and are by definition incomplete and often biased.



EHR data are available only for those patients who are motivated (often by disease or illness) and able to see a provider. Other attributes related to the healthcare provider and providing organization influence the nature of the data in EHRs, including the experience of the provider, availability and use of diagnostic equipment and therapeutic procedures, interactions with clinical specialists, insurance coverage and limitations, and coding and reimbursement practices of the organization [\[2\]](#). The quantitative impact of each of these features on the performance of phenotype definitions is largely unknown. The measurement and estimation of these factors, and the development of strategies to mitigate their impact on data quality, are active methodological research areas in health services research and informatics.

What are the benefits of “standard” phenotypes or condition definitions?

The explicit documentation of computable phenotype definitions can support their use in many different organizations or settings for the consistent identification of patient populations for various purposes. It is important to identify appropriate phenotype definitions for health policy and research. Differences across phenotype definitions can potentially affect their application in healthcare organizations and subsequent interpretation of data.

It is not proposed that a single phenotype definition—of type 2 diabetes mellitus or heart failure, for example—will be sufficient for all intended uses. Rather, the Collaboratory intends to research existing phenotype definitions and document a set of common, well-defined phenotype definitions appropriate for a given characteristic or condition and intended use. This work will support future standardization efforts, including realizing the vision of a [standardized Table 1](#) for reporting baseline patient characteristics in research studies.

Standardization—the process of reconciling differences—can be applied in many different ways within the arena of phenotypes. We distinguish between *data capture standardization*, *phenotype definition standardization*, and *phenotype representation standardization*. These distinctions are important because researchers using secondary data for research purposes do not normally have the ability to enforce data capture standardization for the originating system.

The standardization of one or more phenotype definitions is a complex process that will necessarily engage many stakeholders, representing clinical, research (industry and academia), and patient perspectives. Future work of the Collaboratory will be to identify and promote standards in this area by supporting broader vetting and promotion of scientifically and clinically validated phenotype definitions.

What data sources are used?

Unfortunately, there are still only a limited number of data fields that are routinely collected across different EHR systems. Most phenotype definitions, therefore, use some combination of International Classification of Diseases codes ([ICD-9](#)), medication names, and/or laboratory tests. [ICD-9-CM](#) diagnosis codes can be found in technical billing, professional billing, and/or problem lists. In the future, EHRs will use [ICD-10](#) codes for diagnoses and potentially [Systematized Nomenclature of Medicine–Clinical Terms \(SNOMED CT\)](#) codes for problem lists and other aspects of EHRs. EHRs also contain narrative (unstructured) data. The use of [natural language processing](#) techniques within the biomedical domain is evolving and may offer opportunities for leveraging clinically rich, narrative data within EHRs [\[3\]](#). There are many opportunities to validate and improve these algorithms [\[4\]](#).

The U.S. Department of Health and Human Services' (HHS) [Office of the National Coordinator of Health Information Technology \(ONC\)](#) maintains standards and implementation specifications for EHR systems to ensure that certified systems support the

achievement of [Meaningful Use](#) criteria [\[5\]](#). Accordingly, data elements required by ONC are able to be collected within all [certified](#) EHR systems in the United States in a manner consistent with ONC specifications.

Because EHR data may be available from different types of encounters, including inpatient, outpatient, and emergency department visits, phenotype definitions should take into consideration which sources are relevant to answering the question at hand. In some cases, multiple sources will be needed for complete data capture. For example, medication data can be obtained from reconciliation of various contexts, such as inpatient administration, provider ordering, or outpatient dispensing.

What terms are related to phenotype definitions?

Informatics and data standards groups use the following terms related to phenotype definitions:

- **Data element:** the unit of data being queried, exchanged, or analyzed, which includes a descriptive name that represents the concept being described plus a specified value set and other descriptive metadata, such as a definition. As illustrated in the next section, phenotype definitions can be represented using one or more data elements.
- **Value set:** the set of possible values, categories, or responses (and their codes) that are associated with a particular data element, often derived from established vocabularies or data standards [\[6\]](#).
- **Metadata:** descriptive data about objects, including data objects. Metadata are data about data [\[7\]](#), such as version, author, concept, identifier, data type, definition, and preferred label for a particular data element in a data collection system or form.
- **Operationalization:** a process by which a researcher defines how a concept is measured, observed, or manipulated within a particular study and available data sources; this process translates a theoretical, conceptual variable of interest into a

set of specific operations or procedures that define the variable’s meaning in a specific study, allowing for examination of a hypothesis [8]. A phenotype definition can be considered an operationalization of a disease concept in electronic health data systems or clinical data repositories.

The standardization of data elements and their associated value sets will support consistent phenotype definitions across healthcare providers and organizations using different EHR systems. This is the goal of the ONC, using the Meaningful Use incentive program, and is supported in part by the [NIH Common Data Element](#) initiatives and the [Value Set Authority Center](#) of the National Library of Medicine.

How are data elements and phenotypes different?

Every data element has a value set, and value sets can vary in size and complexity. A value set might include a limited set of categorical values, or a more extensive list of codes from standardized coding systems such as ICD-9-CM or [RxNorm](#). For example, the data element for “sex” includes a single variable with that name, along with a set of discrete values, and perhaps with a definition and associated descriptive metadata. To query the sex of a person, a single data element is assessed.

Example Data Elements with Associated Value Sets (Categorical Value Types)

Data Element	Value Set (Categorical Values)
Sex	Male, Female, Unknown/Not reported

Race American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White, Unknown/Not reported

Demographic characteristics are generally data elements in and of themselves, not combinations of data elements. They might be considered phenotypes themselves, but more often are used as component data elements for phenotype definitions of particular medical conditions.

Many data elements include long lists of values, called nominal value sets. Data elements with nominal value sets can reference entire coding systems or enumerated lists from standardized coding systems or controlled vocabularies.

Example Data Elements with Associated Value Sets (Nominal Value Types, Using Coding Systems)

Data Element	Value Sets (Nominal)
Final diagnosis	ICD-9-CM codes (all)
Final diagnosis of diabetes	249.xx, 250.xx, 357.2, 362.01-06 , 366.41 (from ICD-9-CM)
Medications ordered	Local medication list; clinical drugs coded in RxNorm
Diabetes-related medications ordered	Acarbose, Precose, Acetohexamide, Dymelor, etc.

Phenotype definitions are represented as logical query criteria using one or more data elements with a defined value set. For example, to infer that a patient has a clinical characteristic such as diabetes, evidence can come from one or many data elements:

Possible Data Elements to Identify the Presence of Diabetes

Data Element	Value Sets (Nominal)
ICD-9-CM codes for diabetes	249.xx, 250.xx, 357.2, 362.01-06 , 366.41
Diabetes-related medications	Acarbose, Precose, Acetohexamide, Dymelor, etc.
Hemoglobin A1c values suggestive of uncontrolled diabetes	≥6.5%

Any one of the elements in the table above, or all of the elements collectively, could be used to define a phenotype definition for diabetes. Such a definition would specify that any or all of the data elements (and associated value criteria) must be present to classify a patient as having diabetes on the basis of the data recorded in the EHR.

Demographic characteristics such as sex are not really phenotype definitions, but rather data elements whose value sets are relatively short lists of category variables. However, they are included in this discussion because they are important person characteristics, frequently reported in research, and need to be standardized across pragmatic clinical trials. Data elements for such patient characteristics can also be part of a phenotype. For example, male sex could be component of a prostate cancer phenotype definition.

Evaluating Phenotype Definitions

What makes a “good” phenotype definition?

Computable phenotype definitions should be explicit, reproducible, reliable, and valid. Specific details of the components of a definition (e.g., data elements; value sets) should be provided and should be sufficient to reproduce the query in another system or by another data operator. For a phenotype definition to be reliable, it must be able to produce a similar result with the same data set every time it is applied. For a phenotype definition to be valid, it must identify the condition for which it was developed and claims to identify and meet the desired degree of sensitivity and specificity.

Various performance metrics are used to measure the performance of a phenotype definition in different data sources or populations, analogous to measuring the performance of a case definition or diagnostic technique. These metrics include [sensitivity](#), [specificity](#), [positive predictive value \(PPV\)](#), and [negative predictive value \(NPV\)](#).

In addition, to become consistently used, computable phenotype definitions must leverage data that are routinely collected in most, if not all, EHRs and/or ancillary systems.

How can the validity of a phenotype definition be determined?

The validity of a phenotype definition refers to its ability to correctly measure or detect people with and without the intended condition; i.e., its ability to correctly identify which individuals exhibit the true phenotype and which do not.

The estimation of validity requires a gold standard, defined as the best classification available for assessing the true or actual phenotype status. Assessment of a gold standard is a resource-intensive process requiring careful manual review of current and historic individual patient data. Due to logistical and efficiency considerations, multiple clinical reviewers are usually involved in the process. However, to ensure consistency between conclusions drawn from the patient records, an initial training of the reviewers is crucial. Most studies utilize expert clinicians to review identified cases, but do not specify the training of the individuals or the details of their assessment of true disease/case status.

Many phenotype developers have conducted validation studies [\[9–11\]](#), but none appear to have used a controlled approach. Some investigators attempt to characterize the validity of a phenotype definition using [agreement rates](#) between the definition and a known standard, whereas others report the sensitivity or specificity of the definition compared with a known or gold standard. In this context, sensitivity is the ability to correctly identify individuals who have the phenotype, and specificity is the ability to correctly identify those who do not have the phenotype. The true phenotype status must be known to assess validity. Positive predictive value (PPV) provides an estimate of the prevalence of the true condition among individuals who have the phenotype, and negative predictive value (NPV) provides an estimate of the prevalence among those who do not have the phenotype. PPV and NPV give an indication of the success rate of the phenotype definitions when they are to be used in practice. Similar to sensitivity and specificity, PPV and NPV require knowledge about the true phenotype. These can be estimated based on sensitivity, specificity, and prevalence of the condition in the population being examined.

Researchers at Duke University’s Center for Predictive Medicine are developing and testing methods to quantify the validity and reliability of certain computable phenotype definitions (see [presentation](#)).

Determination of a gold standard is a critical complicating factor related to questions about data quality in EHRs and ultimately the “source of truth.” For conditions in which laboratory values are diagnostic, a laboratory value can be the gold standard, although the clinical context is critical in many cases. For behavioral or mental health conditions, the gold standard or best source of data to approximate the “truth” is often from the patient or from an observation by an expert clinician. For many diseases with complex etiology, subjective diagnosis, or a broad range of clinical presentations, the best source of data (or “truth”) is not clear. Likely, a variety of data sources must be used to determine a patient’s true state of disease or identify the condition.

How can the reliability and reproducibility of a phenotype definition be determined?

Reliability refers to the extent to which an experiment, test, or measuring procedure (or phenotype definition) yields the same results on repeated trials [\[12\]](#). Reliability is an attribute of any computer-related component (software, hardware, or a network, for example) that consistently performs according to its specifications. One method for assessing reliability is to implement the phenotype definition algorithm multiple times and see if the results on the same patients are the same over repeated implementations.

In contrast, *reproducibility* refers to the consistency of results/implementation of the algorithm multiple times under similar conditions (perhaps with different person implementing). For reliability, one would repeatedly implement the algorithm on the same set of patients and check whether the phenotype results for the same patients match. For reproducibility, the algorithm can be implemented on either different or the same patient populations by different “coders.”

Ultimately what is required is an unequivocal algorithm that is implemented without any room for confusion. For most diseases (especially those with a subjective diagnosis or broad range of clinical presentations), a variety of data sources must be included

in a phenotype definition. Unfortunately, the more complex the phenotype definition, the more difficult it can be to reproduce and the more likely errors can influence the reliability of the algorithm [\[13\]](#).

Several well-known issues can affect reliability, including coding terminology changes over time and coding practice variations at the provider, healthcare system, and regional levels. An active and future area of research involves studying data quality and testing various phenotypes in different settings or time periods to represent variations in data quality.

How can the reproducibility of a phenotype definition be optimized?

Two features of phenotype definitions can enhance the likelihood that they will be applied consistently: clearly articulated specifications for the definition and guidance for implementers. However, the development of meaningful specifications and documentation is complicated by the variation in healthcare information systems and lack of data standards for EHR data.

Ideally, a phenotype definition should be reproducible across institutions, but many factors can affect reproducibility, including regional differences in patient populations, differences in EHR systems, variations in the work flows that generate data, and variations in coding practices.

What are potential limitations of EHR data and computable phenotypes?

The data contained in EHRs and ancillary systems are generated through the provision of clinical care. As such, the data are not optimized for secondary uses and are associated with multiple limitations when applied for research purposes [\[14\]](#).

Missing Data

Because EHR data are derived from patient encounters with a provider or healthcare system, data are only recorded during healthcare episodes. This can result in bias due to healthier individuals being missing from the dataset. “Missingness” is a frequent problem and is often nonrandom—a concept known as informative censoring [\[15,16\]](#). Patients are also lost to follow-up if they move out of the area or obtain care from a provider in a different healthcare system. In pragmatic clinical trials, it is therefore important to distinguish between “not present” in the dataset versus “did not assess.”

Inaccurate or Uninterpretable Data

Errors are common in data from EHRs or ancillary sources, because most data are entered by busy healthcare providers during a patient visit or afterwards from recall. Phenotype definitions based on coding that is influenced by billing are susceptible to systematic biases. In addition, data may be uninterpretable if, for example, units of measurement are missing or analyzable information cannot be gleaned from qualitative assessments.

Complex and Inconsistent Data

In healthcare, clinical definitions, coding rules, and data collection systems vary over time, creating challenges in the analysis of these data. Data collection practices can also vary by providers at different locations. Finally, much information is still captured as unstructured data and stored in narrative notes. Though many challenges exist in extracting unstructured data, these data are increasingly being used to support various types of clinical decision-making and research using an evolving set of tools [\[17\]](#).

<p>Mo et al., 2015 LUE!</p>	<p>Background Electronic health records (EHRs) are increasingly used for clinical and translational research through the creation of phenotype algorithms. Currently, phenotype algorithms are most commonly represented as noncomputable descriptive documents and knowledge artifacts that detail the protocols for querying diagnoses, symptoms, procedures, medications, and/or text-driven medical concepts, and are primarily meant for human comprehension. We present desiderata for developing a computable phenotype representation model (PheRM).</p> <p>Methods A team of clinicians and informaticians reviewed common features for multisite phenotype algorithms published in PheKB.org and existing phenotype representation platforms. We also evaluated well-known diagnostic criteria and clinical decision-making guidelines to encompass a broader category of algorithms.</p> <p>Results We propose 10 desired characteristics for a flexible, computable PheRM: (1) structure clinical data into queryable forms; (2) recommend use of a common data model, but also support customization for the variability and availability of EHR data among sites; (3) support both human-readable and computable representations of phenotype algorithms; (4) implement set operations and relational algebra for modeling phenotype algorithms; (5) represent phenotype criteria with structured rules; (6) support defining temporal relations between events; (7) use standardized terminologies and ontologies, and facilitate reuse of value sets; (8) define representations for text searching and natural language processing; (9) provide interfaces for external software algorithms; and (10) maintain backward compatibility.</p> <p>Conclusion A computable PheRM is needed for true phenotype portability and reliability across different EHR products and healthcare systems. These desiderata are a guide to inform the establishment and evolution of EHR phenotype algorithm authoring platforms and languages.</p>
--	--

	<p>http://computablemedicalrecords.org/</p> <p>A Computable Medical Record</p> <p>... houses a Patient's Data</p> <p>A patient's record from conception to death.</p> <p>... is Self Aware</p> <p>Does the record exist? Is this the right person?</p> <p>... is Machine & Human Readable</p> <p>... declares its Level of Fidelity</p> <p>What is the level of completeness and accuracy. What's missing?</p> <p>... supports a Health Status Score</p> <p>... and is Open Source (GPL3).</p>
<p>Sonin 2016</p>	<p>Computable Health Record:</p> <ul style="list-style-type: none"> • is a health data standard,

- a secure web health service,
- an interactive visual application

The service provides any citizen with an accurate record (complete or not) of their digital health data, health status, and care plans.

References:

- https://hiriresearch.files.wordpress.com/2011/10/the_computable_emr_3-0_explained1.pdf
- <https://hiriresearch.files.wordpress.com/2010/11/the-computable-medical-record.pdf>
- <https://hiriresearch.files.wordpress.com/2010/11/defining-the-clinical-content-of-the-medical-record.pdf>
- <https://hiriresearch.files.wordpress.com/2011/10/minutes-ad-hoc-emr-wg-meeting-november-3rd-2011.pdf>
- <http://ontology.buffalo.edu/medo/ACAAI-2012.pdf>
- <http://hiriresearch.org/category/uncategorized/>
- <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2274797/>
- <http://jamia.oxfordjournals.org/content/early/2015/02/09/jamia.ocu013>
- <http://www.napcrg.org/Conferences/AnnualMeeting/AbstractSearch?m=6&s=11983>

News References:

- <http://www.bostonglobe.com/business/2015/02/03/leung/PKOkXUsTSyG3tKGRwvZXnK/story.html>

<p>Leung 2015</p>	<p>For Massachusetts, digital health is about building a cloud to help hospitals house electronic medical records, writing software to keep patient data secure, connecting medical devices, analyzing big data for health care trends, and launching apps for consumers.</p> <p>It's also about helping people sign up for health insurance online — a seemingly simple task for any e-commerce outfit but complicated once health care is involved. Or it can be creating gee-whiz consumer ideas like PillPack, a local startup that takes the pharmacy into the digital age and to your doorstep.</p> <p>- -</p> <p>Jeff Leerink, the founder of Leerink Partners, a Boston health care investment bank, said digital health care is where biotechnology was in Massachusetts in the '80s and '90s.</p> <p>Health care's complexity "is why it's the last bastion of American industry that hasn't been undergoing as much of a technological revolution as other areas of the economy," said Leerink. "We do think now is a tipping point for the industry."</p> <p>But money alone can't buy love from tech startups. They are attracted to Silicon Valley, where the startup culture is as natural as the California sunshine. Mentors abound, seed capital is plentiful, and entrepreneurial activity on campuses is encouraged. No wonder so many tech companies that start out here move out West.</p>
-----------------------	--

	<p>“It is much more insular here,” said Tucci, the EMC chief. “When you go to Silicon Valley, there are thousands and thousands of people who made hundreds of millions of dollars. It’s the old adage we all learned: If you make money easy, you’ll spend it easy. It’s much easier to get an idea funded out there.”</p>
<p>Richesson et al, 2013</p>	<p>Abstract</p> <p>Widespread sharing of data from electronic health records and patient-reported outcomes can strengthen the national capacity for conducting cost-effective clinical trials and allow research to be embedded within routine care delivery. While pragmatic clinical trials (PCTs) have been performed for decades, they now can draw on rich sources of clinical and operational data that are continuously fed back to inform research and practice. The Health Care Systems Collaboratory program, initiated by the NIH Common Fund in 2012, engages healthcare systems as partners in discussing and promoting activities, tools, and strategies for supporting active participation in PCTs. The NIH Collaboratory consists of seven demonstration projects, and seven problem-specific working group 'Cores', aimed at leveraging the data captured in heterogeneous 'real-world' environments for research, thereby improving the efficiency, relevance, and generalizability of trials. Here, we introduce the Collaboratory, focusing on its Phenotype, Data Standards, and Data Quality Core, and present early observations from researchers implementing PCTs within large healthcare systems. We also identify gaps in knowledge and present an informatics research agenda that includes identifying methods for the definition and appropriate application of phenotypes in diverse healthcare settings, and methods for validating both the definition and execution of electronic health records based phenotypes.</p>
<p>Hughes 2011</p>	<p>The “computable” electronic Medical Record (eMR 2.0) defined</p>

It has been almost a half century since the benefits and principles of a computer based medical record were first described by Schultz and Weed. Why have we yet to achieved these putative benefits? Since the required computing horse power is now cheap enough the answer lies in understanding what computing is and how it can be applied to medical records.

An “Excel” spread sheet is computable (with regards to numbers) but a “Word” document is not. If you enter letters into the cells of a spread sheet they “do not compute” because the machine understands the semantics of numbers but not letters and words.

To make words and letters computable they must be represented as numbers. This is what lexicons such SNOMED CT and ICD9 do by assigning unique identifiers to clinical entities.

For the specified terms to participate in more sophisticated integration and semantic communication they require information to be attached. This meta data is the basis of information models such as the clinical document architecture (CDA) and detailed clinical models (DCM) from HL7 and the archetype from openEHR.

By using terminologies we can sort information in complex ways such as listing all patients aged 65 with diabetes type two. With detail rich terminologies such as SNOMED CT we can support clinicians at the point of care. By adding meta

-

data we can ask for a list of all patients with diabetes type two who have not had an A1C test in the last six months.

To achieve the type of assistance from the computer that will revolutionize clinical medicine and allow for real

-

time case specific research, such as published this week in the NEJM, the terminology with its attached meta data must be organized in a good ontology. A good ontology is a veridical relational hierarchy based on single parent child inheritance (i.e. each child has only one parent term); a limited number of upper level categories; and is congruent with the Open Bio

-

medical Ontology (OBO) which derives from the Basic Formal Ontology (BFO), one of the most widely used upper level ontologies in science. This is the process that the IHTSDO is trying to achieve for SNOMED CT.

We can now ask patient specific questions such as what is the best treatment strategy for this patient's situation according to the last one thousand patients of the same profile, taking into consideration his proteomic phenotype. Or we can ask what is the most common guideline non compliance in this group of medical residents.

The potential benefits of a computable medical record support the three classic domains of academic medicine; service, education and research. There is much talent in Canada and a lot of silo development is occurring. A national approach to specifying the clinical content of the eMR such that is truly "computable" will avoid waiting another half century.

John Hughes 2011/11/06

Follett, J., & Herlin, B. (2016, January 5). Digital healthcare services in 2016 (and beyond). Retrieved 8 January 2016, from <http://mobihealthnews.com/content/digital-healthcare-services-2016-and-beyond>

GOINVO. (n.d.). Computable Medical Records. Retrieved 8 January 2016, from <http://computablemedicalrecords.org/>

Hughes, J. (2011, November 6). The 'computable' electronic Medical Record (eMR 2.0) defined. Retrieved from https://hiriresearch.files.wordpress.com/2011/10/the_computable_emr_3-0_explained1.pdf

Leung, S. (2015, February 4). Mass. business leaders bet on health care tech - The Boston Globe. *BostonGlobe.com*. Boston. Retrieved from <https://www.bostonglobe.com/business/2015/02/03/leung/PKOkXUsTSyG3tKGRwvZXnK/story.html>

Mo, H., Thompson, W. K., Rasmussen, L. V., Pacheco, J. A., Jiang, G., Kiefer, R., ... Harris, P. A. (2015). Desiderata for computable representations of electronic health records-driven phenotype algorithms. *Journal of the American Medical Informatics Association*, ocv112. <http://doi.org/10.1093/jamia/ocv112>

PheKB. (2014). What is the Phenotype KnowledgeBase? | PheKB. Retrieved 8 January 2016, from <https://phekb.org/>

Richessaon, R., & Smerek. (2015). Electronic Health Records-Based Phenotyping | Rethinking Clinical Trials®. Retrieved from <http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/>

Richesson, R. L., Hammond, W. E., Nahm, M., Wixted, D., Simon, G. E., Robinson, J. G., ... Califf, R. M. (2013). Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *Journal of the American Medical Informatics Association: JAMIA*, 20(e2), e226–231. <http://doi.org/10.1136/amiajnl-2013-001926>



Sonin, J. (2015, February 14). [goinvo/ComputableMedicalRecords](https://github.com/goinvo/ComputableMedicalRecords) [Wiki]. Retrieved 8 January 2016, from <https://github.com/goinvo/ComputableMedicalRecords>