Wikipedia and Wikidata in Consumer Health Informatics Applications

**Louise To**

Associate Fellow, 2019-2020

National Library of Medicine

User:To1518

**Project Sponsors**

Vojtech Huser (NIH/NLM/LHC)

User:EncycloABC

January 2020

Table of Contents

[Abstract 3](#_Toc30068242)

[Background 3](#_Toc30068243)

[Objective 4](#_Toc30068244)

[Procedures 4](#_Toc30068245)

[Researching PHRs 5](#_Toc30068246)

[Researching Infobutton 5](#_Toc30068247)

[Researching Wikidata 5](#_Toc30068248)

[Outcomes 5](#_Toc30068249)

[PHRs 5](#_Toc30068250)

[Infobutton 14](#_Toc30068251)

[Wikidata 19](#_Toc30068252)

[Discussion 21](#_Toc30068253)

[Wikidata Tools and Resources 22](#_Toc30068254)

[Next Steps 23](#_Toc30068255)

[References 23](#_Toc30068256)

# 

# Abstract

Background

Wikipedia is a widely referenced knowledge resource that patients are familiar with. Connecting patients to Wikipedia through infobuttons in their electronic health records and health applications is possible through Wikidata, a structured database tied to Wikipedia.

Objective

This project lays out how personal electronic health records, infobuttons, Wikipedia, and Wikidata intersect to become a knowledge resource for patients by examining the current ways patients access their health records, explaining the technology of infobuttons, and proposing how Wikipedia, through Wikidata, can serve as a viable infobutton knowledge resource.

Procedures

Research included literature searches, an interview with a project lead of OpenInfobutton, Guilherme del Fiol, and connecting with Wikidata leaders like Daniel Mietchen, Lane Raspberry, and Rob Fernandez. Wikidata and infobutton help pages supplemented the learning process.

Outcomes

From the two patient health systems that were analyzed (Kaiser and myPrivia), infobutton-like links, or knowledge links, seemed non-existent or limited. Infobutton technology is dependent on structured data and medical terminology annotations being on both the health information system side as well as the knowledge resource side. Querying Wikidata’s structured database can create an index of terminologies for infobutton technology to run against to locate the appropriate Wikipedia article’s URL for the patient to view.

Discussion

Wikidata needs to have more comprehensive coverage of medical terminology annotations before it can become an infobutton knowledge resource. Given sparse coverage, an alternative is building an application that can translate terminologies to find a match in Wikidata, but the former is preferred.

Next Steps

Since MeSH identifiers have been most assigned in Wikidata, evaluating those links and increasing the identifiers to near-completion would provide an environment to showcase Wikipedia as an infobutton knowledge resource. These efforts should be done in collaboration with WikiProject Medicine.

# 

# Background

Patients have access to their electronic health information as a result of the Health Information Technology for Economic and Clinical Health Act (HiTECH) of 2009. Meaningful Use, a clause of the HiTECH Act, initially spurred the use of infobuttons, or context sensitive links that connect the user to relevant knowledge resources. In 2012, infobuttons became a requirement for US EHR certification to address Meaningful Use.1 Infobuttons within electronic health record systems (EHRs) not only support clinicians’ decision-making, but they can also serve patients. Since then, few studies have evaluated clinicians’ use of infobuttons,2,3 and fewer have analyzed patient usage of infobuttons.4

While it is possible for patients to manually type each concept of interest from their health record separately into a search page of any given resource (e.g., UpToDate, MedlinePlus or simply search via Google, Bing or Yahoo), the seamless integration via a quick infobutton click creates a new mode of consumer information-seeking. For complex and lengthy health records, the seamless integration allows a patient to quickly explore many topics and adds a level of resource curation and refinement that cannot be done through individual searches for each item of interest.

For infobuttons to be useful, there must be suitable knowledge resources that cover a significantly large body of knowledge. The following knowledge resources have been paired with infobutton in the past: UpToDate, MedlinePlus, Micromedex, Krames, MayoClinic, and WebMD. This project investigates the viability of Wikipedia as an infobutton knowledge resource for patients. Patients would access or download their personal health records (PHRs), either from their medical insurance company or health care provider, and each diagnosis, medication, lab test, etc on their health record would have an infobutton that has the option to link to the corresponding Wikipedia article.

Infobuttons need structured data in order to link to an appropriate article within a knowledge resource and functionality relies on annotating each knowledge resource article with relevant codes. Medical terminologies, such as ICD-10-CM, MeSH, MedlinePlus, etc, act as the linking semantics between the electronic health record and Wikipedia. For example, a MedlinePlus article on “C-section” is annotated with relevant codes for C-Section in diagnostic terminologies such as ICD-10CM or SNOMED CT. Wikipedia has a parallel resource called Wikidata where each Wikipedia article is linked to the corresponding Wikidata item. The Wikidata layer of Wikipedia provides the structured data needed for infobutton technology to function.

Despite numerous studies demonstrating how the quality of health information in Wikipedia varies, Wikipedia continues to be widely referenced.5–7 One study on patient usage of infobuttons suggests that patients are looking for fast, relevant results and familiarity of a resource may contribute to higher usage.4 Since Wikipedia consistently ranks as one of the most visited websites, patients will probably consider Wikipedia a familiar resource, and would subsequently choose it over other unfamiliar resources regardless of concerns over quality. In a free market that will endlessly produce consumer health applications, Wikipedia is bound to be utilized as a knowledge resource because it does not require licensing fees or formal contractual coordination. Therefore, research into possible usages and recognition of its merits and limitations is significant when the patient’s adoption of Wikipedia is imminent.

## Objective

In this report, we examine the current ways patients access their health records, explain how infobuttons work, and propose how Wikipedia, through Wikidata, can serve as a viable infobutton knowledge resource.

# Procedures

Initial research involved a literature search on Wikipedia, Wikidata, and patient use of infobutton technology to learn the current state of infobuttons and whether Wikidata had been previously considered as a route for consumer health informatics. Following conversations with the project sponsor, Vojtech Huser, we divided the project into three areas of study:

## Researching PHRs

As an effort to learn what patients and consumers experience when trying to access their health records, we explored two patient portals; Kaiser Permanente, myPrivia by Athenahealth, and their corresponding mobile apps. Specifically, we are interested in how providers offer patients information related to the items listed in their medical record, or if the records contain any infobutton-like links to knowledge resources. We refer to those links as knowledge links. We obtained patient permission to use screenshots of the portals for the purposes of this report. All personally identifiable information has been redacted. The patients’ identities will be kept confidential.

## Researching Infobutton

For this project, we researched an open source product called OpenInfobutton to gain a general understanding of infobutton mechanics. On a conference call with one of OpenInfobutton’s project leads, Guilherme Del Fiol, an Assistant Professor from the Department of Biomedical Informatics at the University of Utah, we discussed their infobutton manager and a supporting configuration tool called LITE (Librarian Infobutton Tailoring Environment).

## Researching Wikidata

Researching Wikidata involved watching YouTube tutorials and reading Wikidata’s learning materials and help pages. Vojtech wrote the SPARQL queries and scripts that are noted in this report. Aimee Gogan recommended Stacy Brody, a former Associate Fellow who currently works at the National Agricultural Library. Stacy provided some names to reach out to including Lane Raspberry and Daniel Mietchen at the University of Virginia’s School of Data Science. They are active with the WikiProject Medicine group and have offered to collaborate. Nancy Fallgren, Metadata Librarian at NLM, also suggested Lane Raspberry as a contact. Through Jamie Flood, the Wikimedian-in-Residence at the National Agricultural Library, we connected with Rob Fernandez from Wikimedia DC. Rob offered to provide a workshop on Wikidata tools if we needed assistance.

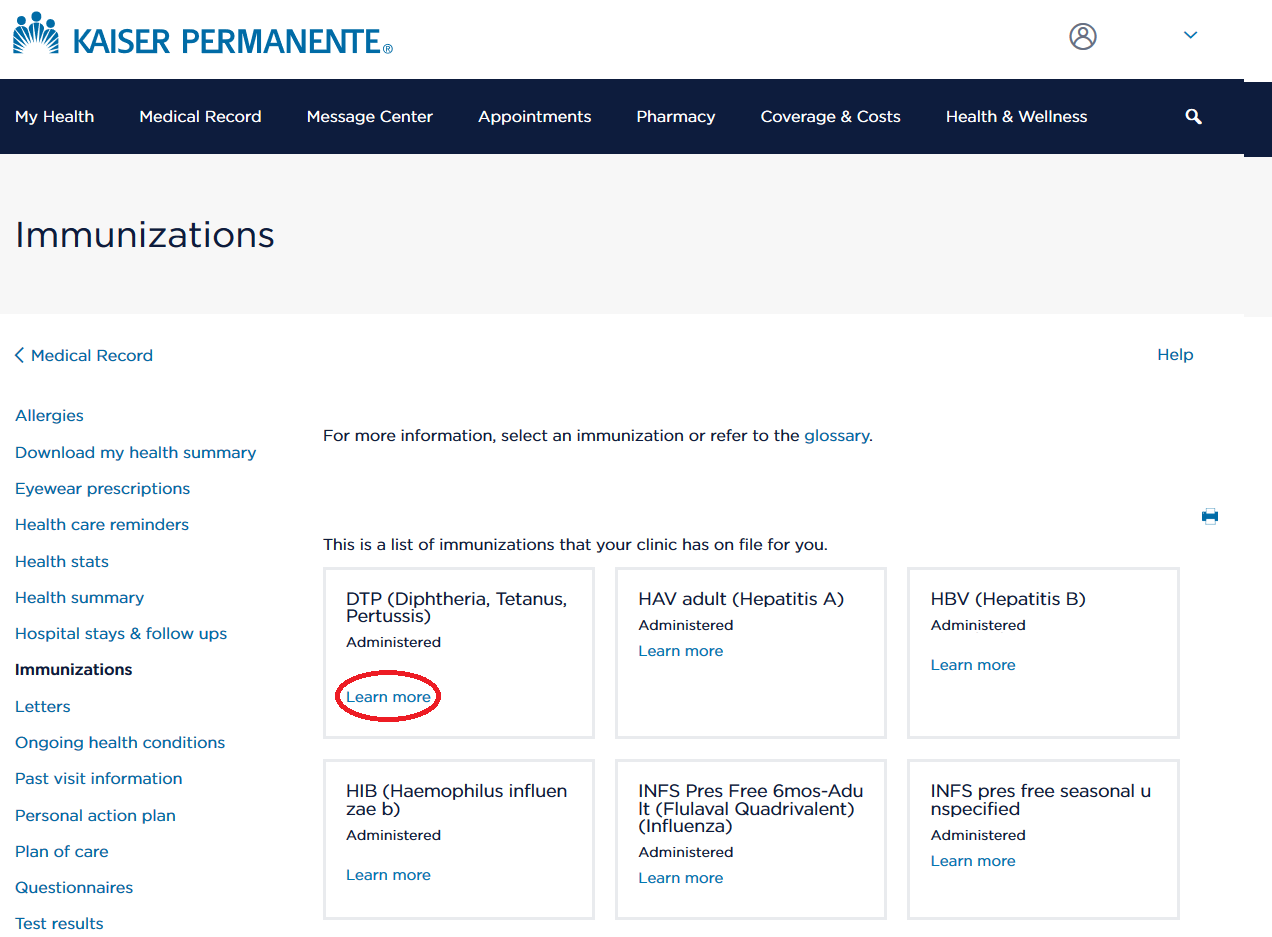
## Outcomes

### PHRs

#### Kaiser Permanente

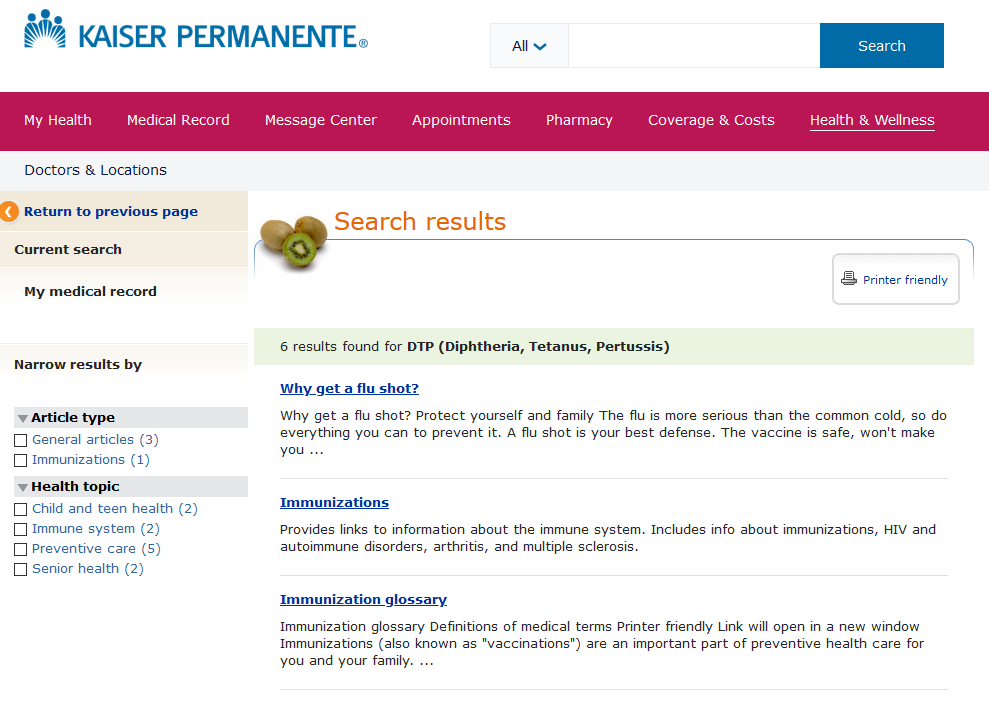
##### Online Portal

Under **Medical Record**, Kaiser offers knowledge links under some record items (Fig. 1).



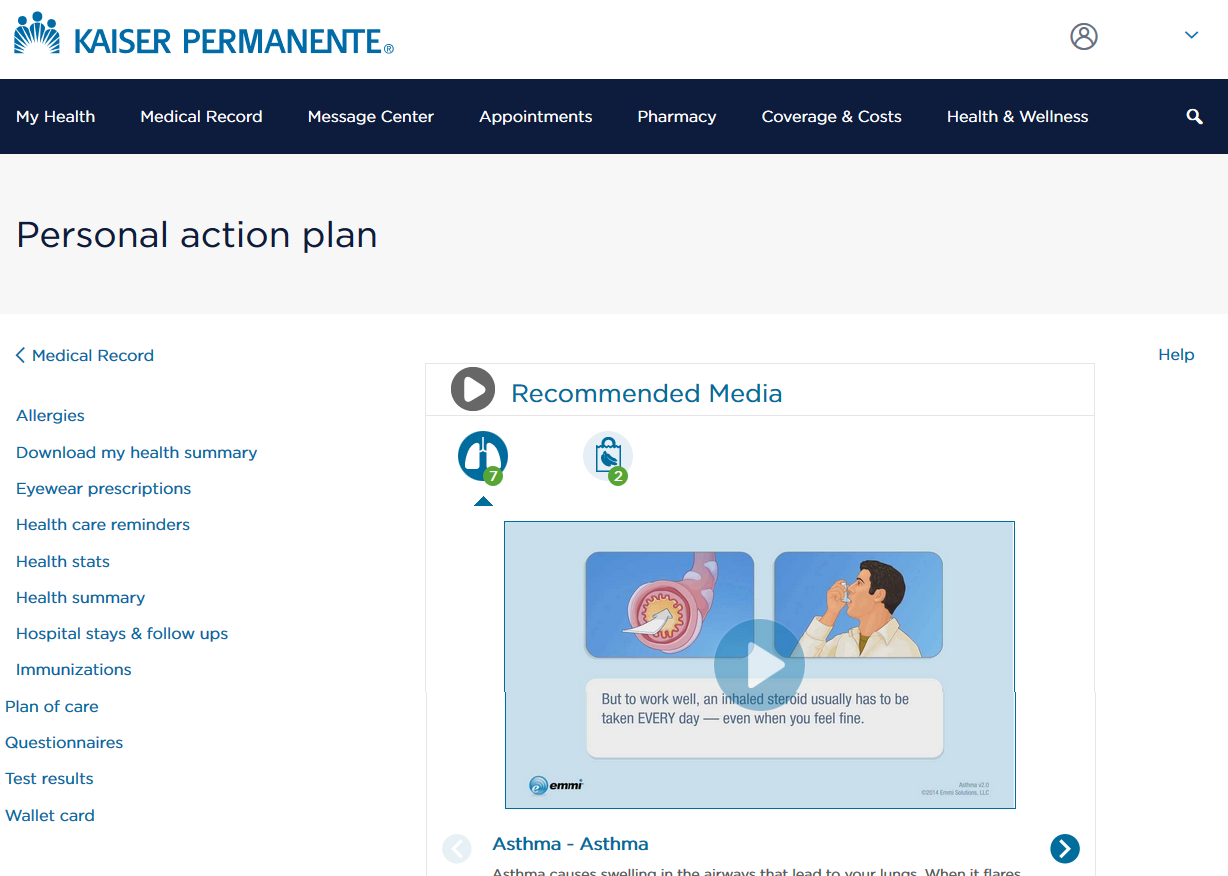
*Figure 1: Immunization history of a Kaiser Permanente Personal Health Record*

However, the patient is then taken to a list of search results that may or may not be specific to the item. For example, after clicking **Learn More** under **DTP (Diphtheria, Tetanus, and Pertussis)**, the patient sees a list of search results about the flu shot and general information on immunizations with nothing specific to DTP (Fig. 2).



*Figure 2. Search results page displayed after clicking on* ***Learn More***

The next closest information option is the **Personal Action Plan**, which is tailored to some of the patient’s conditions, health statistics, and relevant screenings (Fig. 3). However, it is unclear why certain conditions are covered while others are not.

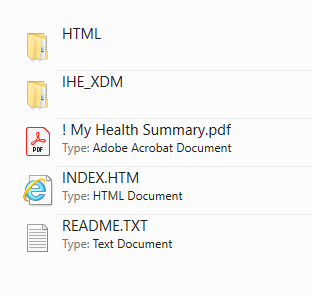


*Figure 3. Personal action plan in a Kaiser Permanente personal health record*

##### Blue Button logo that is a blue circle with a white outline. A downward pointing arrow into a horizontal bracket is in the middle of the blue circle.

##### Download (Blue Button)

Kaiser patients also have the option to download their health summary through the online portal. The ability to download health data is tied to the Blue Button initiative, which began as a response to the Meaningful Use clause of HiTECH. The Blue Button logo is a registered trademark owned by the Department of Health and Human Services that indicates the provider’s participation in making personal health records downloadable.

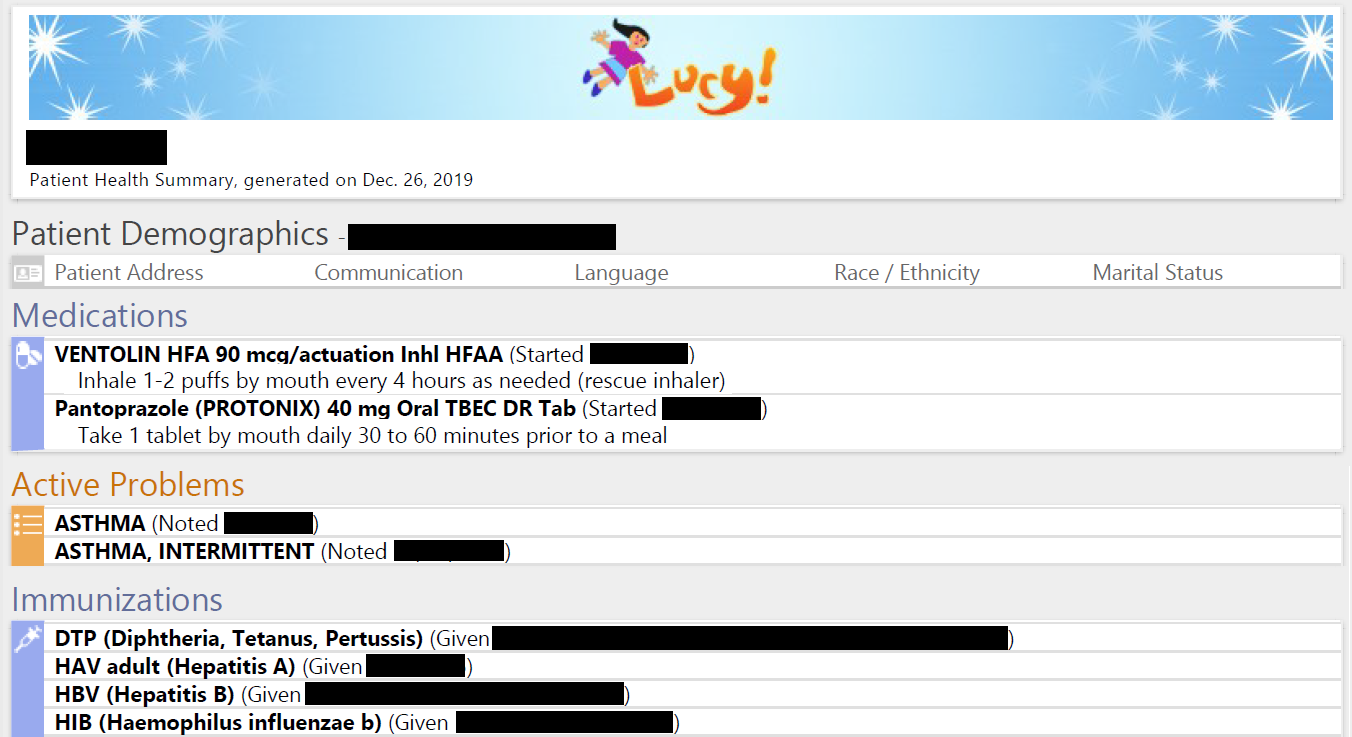


*Figure 4. Document package from downloading PHR from Kaiser*

Kaiser offers the option to download single visits, visits between a date range, all visits, a health summary, and password protection of the download. For this report, we downloaded the health summary (Fig. 4). The download is a ZIP file that opens up to five initial items:

1. a folder titled “HTML”
   1. which contains a folder for images and a style folder for CSS files,
2. a folder titled “IHE\_XDM”
   1. which contains a folder with the patient’s name. The folder with the patient’s name contains two XML files and one XSL file,
3. a PDF file titled “! My Health Summary.pdf”
4. an HTML document titled, “INDEX.HTM”
   1. which explains what a Lucy record is and how to share it with other healthcare providers
5. a text document titled “README.TXT”
   1. which explains each of the items in the zip file

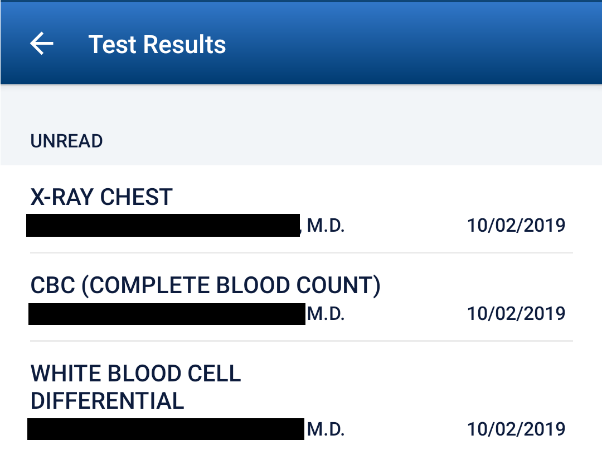
The average patient or consumer most likely will gravitate to the PDF titled “! My Health Summary.” The PDF has a banner at the top illustrating the words “Lucy!” (Fig. 5). Lucy refers to a portable personal health record that organizes a patient’s health information in one place. Lucy was created by EPIC Systems and allows patients the freedom to take their records with them to different providers. The health summary does not offer any knowledge links.



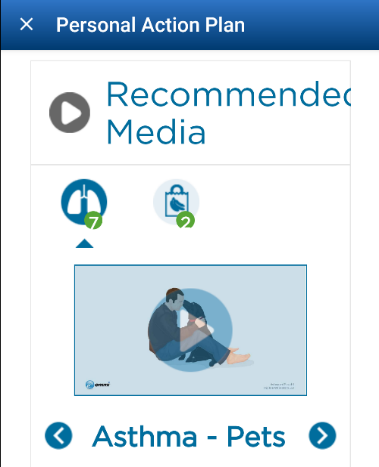
*Figure 5. Lucy!, the portable personal health record, provided by Kaiser*

##### Mobile Application

Similarly, Kaiser’s mobile application (for the purposes of this report we used the Android application; [Play Store link here](https://play.google.com/store/apps/details?id=org.kp.m&hl=en_US)) does not include knowledge links (Fig. 6). The closest information resource they offer is through their Personal Action Plan, which is the same as what is offered through the online portal (Fig. 7).



*Figure 6. Test results of PHR seen through the Kaiser Mobile App*

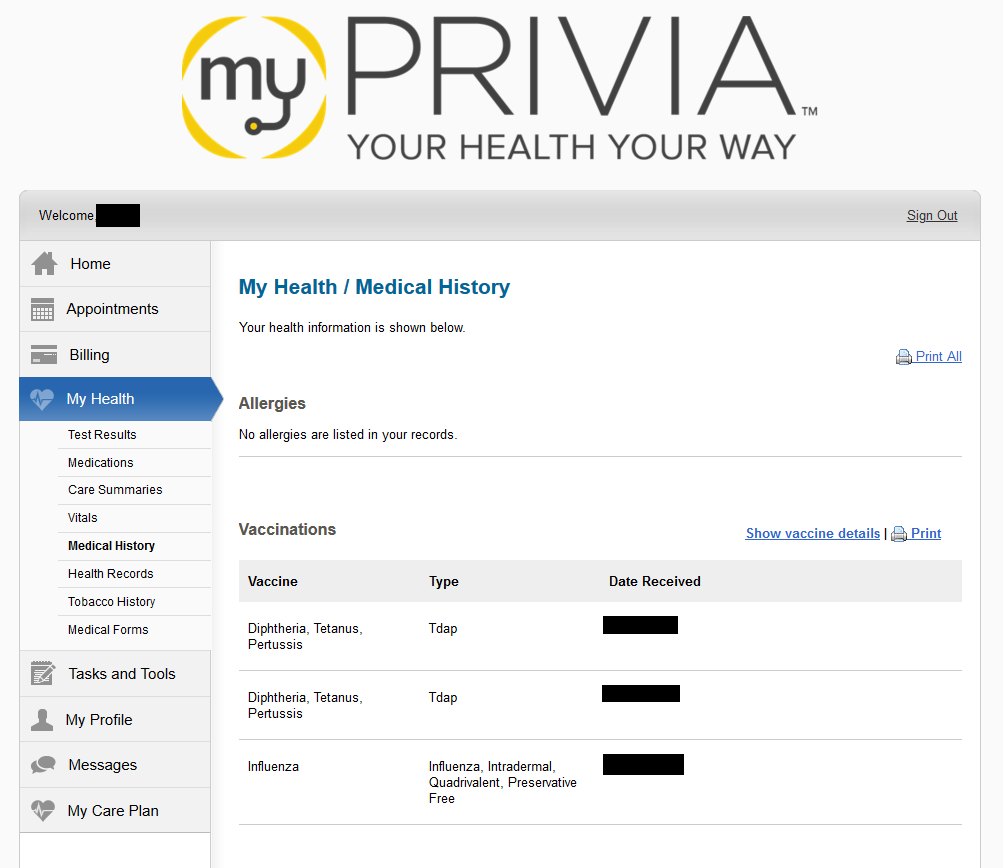


*Figure 7. Personal action plan seen through the Kaiser mobile app*

#### myPrivia

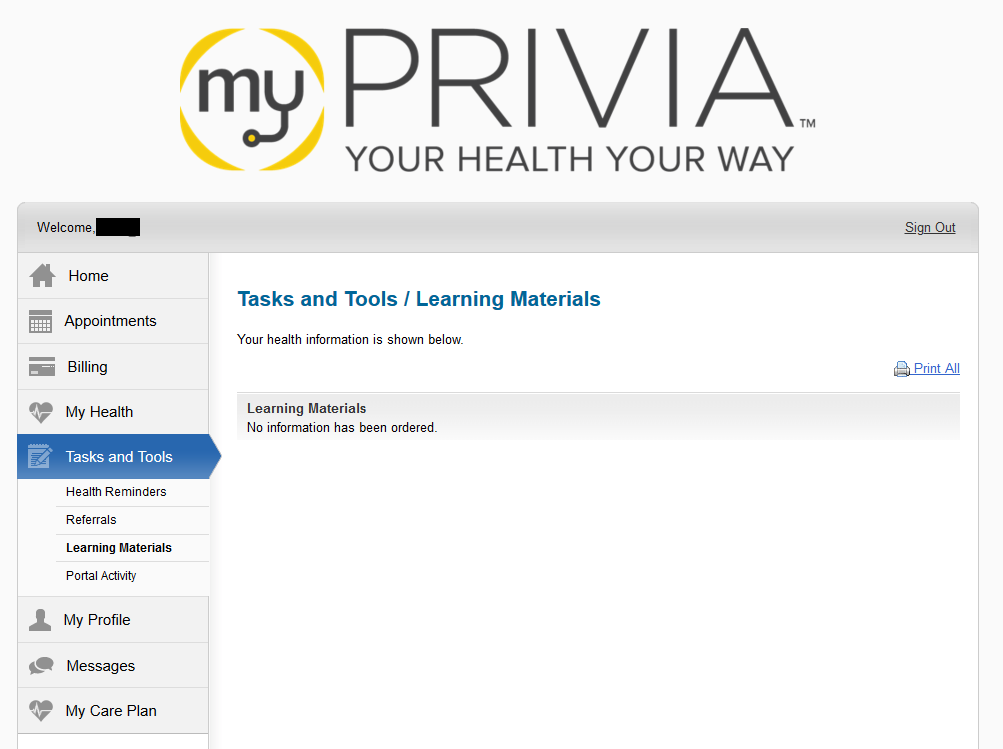
##### Online Portal

Upon logging into myPrivia, the landing page includes notifications and a summary of the patient’s health history. Under **My Health**, the patient can see their health history divided into test results, medications, etc (Fig. 8). However, the portal does not offer any knowledge links.



*Figure 8. Medical history of a myPrivia personal health record*

Under **Task and Tools**, there is a menu choice for **Learning Materials**, but nothing is displayed if learning materials have not been ordered for the patient, presumably by the health professionals that manage patients’ documentation (Fig. 9).



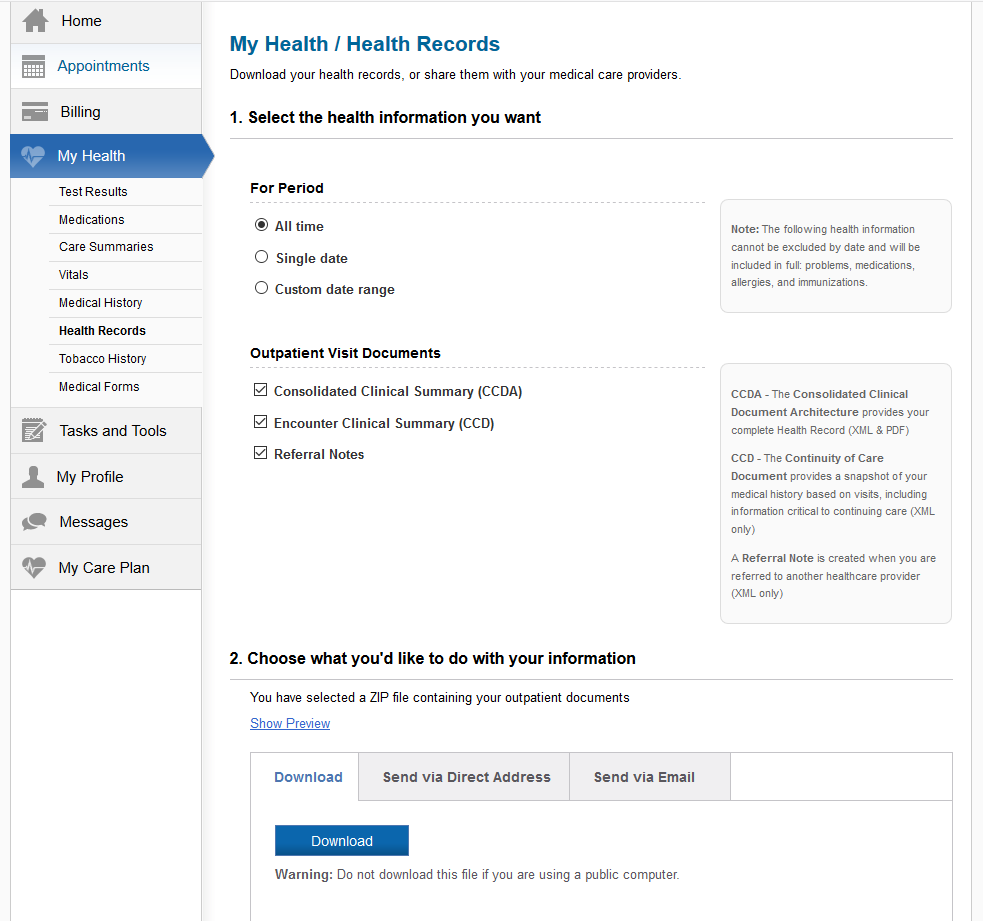
*Figure 9. Tasks and Tools page of myPrivia personal health record*

##### Download (Blue Button)

To download one’s health records from myPrivia, the patient navigates to **My Health** on the left hand menu, and clicks **Health Records**. myPrivia offers patients the ability to select a time period and three options:

1. Consolidated Clinical Summary (Consolidated Clinical Document Architecture)
   1. which provides the complete health record in XML and PDF
2. Encounter Clinical Summary (Continuity of Care Document)
   1. which provides a snapshot of the patient’s medical history in XML
3. Referral Notes
   1. which includes referrals to other healthcare providers in XML

The patient can download their record immediately as a ZIP file, or send by email or direct address (Fig. 10).

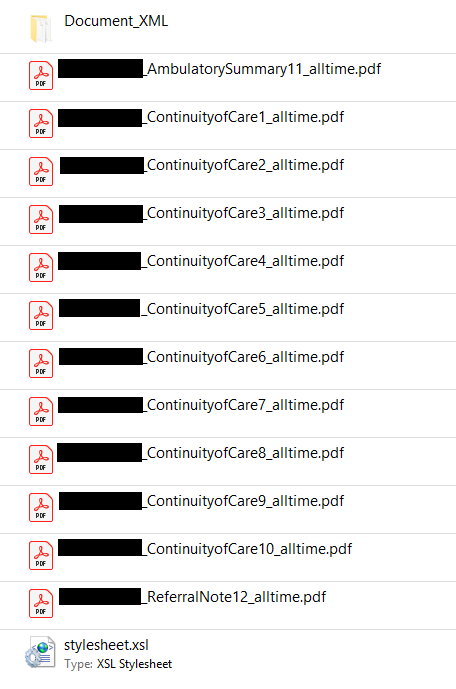


*Figure 10. Options to download health records in myPrivia*

With all three options selected, the download included (Fig. 11):

1. twelve PDFs
   1. that include an “Ambulatory Summary,” ten “Continuity of Care” documents, and a “Referral Note”
2. one folder titled, “Document\_XML”
   1. which includes the twelve XML versions of the PDFs
3. one XSL stylesheet

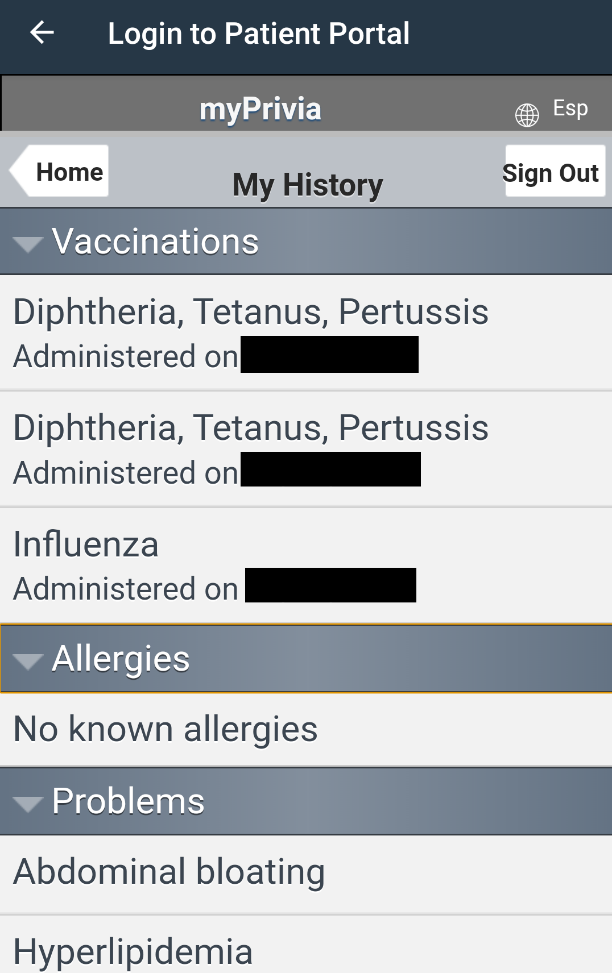
Again, patients will likely open the PDFs as they are the most human-readable. The PDFs do not offer knowledge links.



*Figure 11. Document package from downloading PHR from myPrivia*

##### Mobile Application

Patients navigate to their record by tapping on **My Health** and then **History**. myPrivia’s mobile application does not offer any links or learning materials (Fig. 12).



*Figure 12. Medical history of PHR seen through myPrivia mobile application*

### Infobutton

Infobutton technological infrastructure is generally composed of the following:

1) An **infobutton manager:** software that calls a knowledge resource, fetching the response from all or selected resources

2) A **list of knowledge resources**: licensed or free information resources that are queried for information

3) **Knowledge presentation:** an interface that displays the information to the patient in various platforms such as through a web browser window pop-up or on a mobile app screen

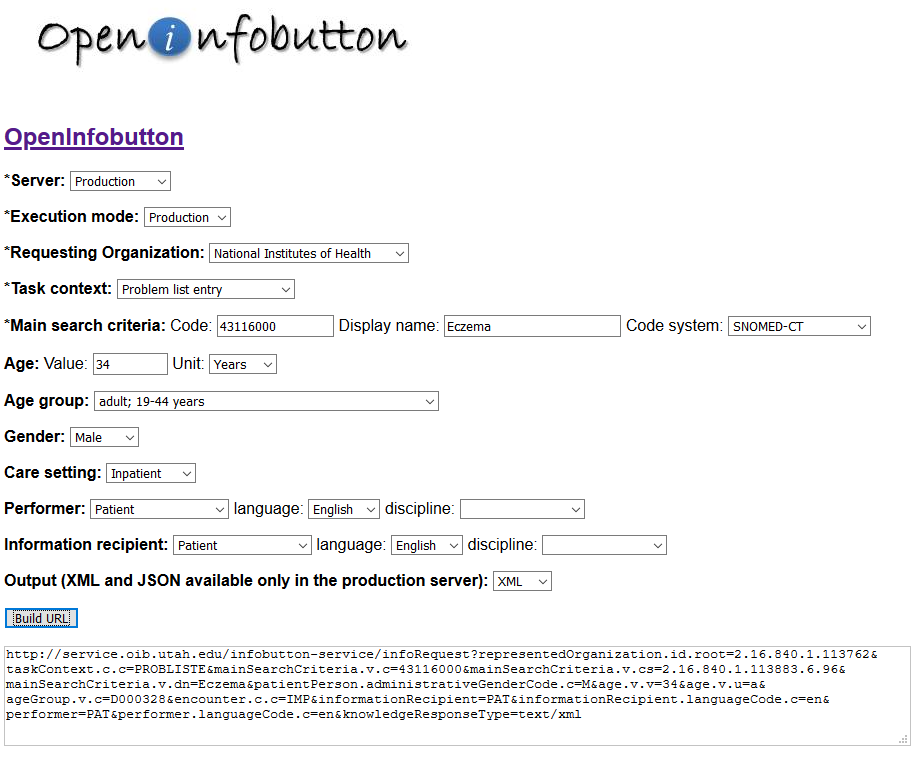
To implement infobuttons, health information systems require an API that understands terminologies, such as ICD-10-CM, SNOMED CT, MeSH, HCPCS, etc. The ability to link to the most appropriate knowledge article relies not only on code annotation of the knowledge resource, but codes must also be present on the health record. If a health information system uses a terminology that does not match the terminology used by the knowledge resource, the infobutton manager will translate the terminologies. OpenInfobutton uses UMLS Terminology Services’ (UTS) RESTful API to translate. UMLS, or United Medical Language System, is a meta-thesaurus provided by the National Library of Medicine that brings together many biomedical vocabularies to enable interoperability between systems.

An infobutton manager uses the patient’s age, gender, language, etc provided by the health information system as filters to narrow down to the most relevant articles from a knowledge resource like UptoDate, or as we propose, Wikipedia. OpenInfobutton offers a test tool that illustrates an example of how URL requests are structured with these parameters in order to retrieve the appropriate resource for the patient (Fig. 13).

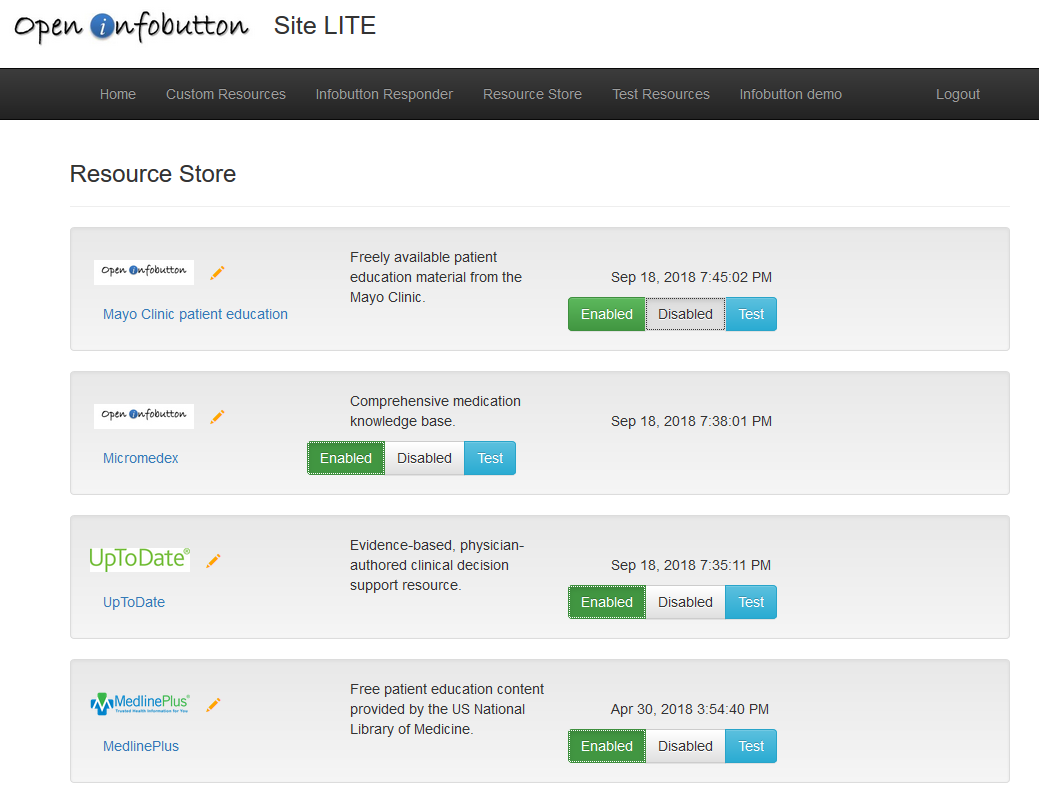
Most infobutton knowledge resources are HL7-compliant.8 In other words, they are standardly structured to enable interoperability. Non-HL7-compliant resources, like Wikipedia, can still be configured for infobutton as long as there is a search API that allows mapping to the infobutton standard parameters. The mapping process is essentially building an index for the API to run against.

LITE (Librarian Infobutton Tailoring Environment) is an OpenInfobutton configuration tool that allows individuals with no IT background to manage infobutton knowledge resources.9 Health information systems that implement infobutton can browse through the resource profiles available in LITE’s Resource Store, and select which knowledge bases they want to enable (Fig. 14). Some of the resources include MedlinePlus, an offering from the National Library of Medicine, and UpToDate. A future aim is to create a Wikipedia resource profile for infobutton implementers to enable for their own health information system.

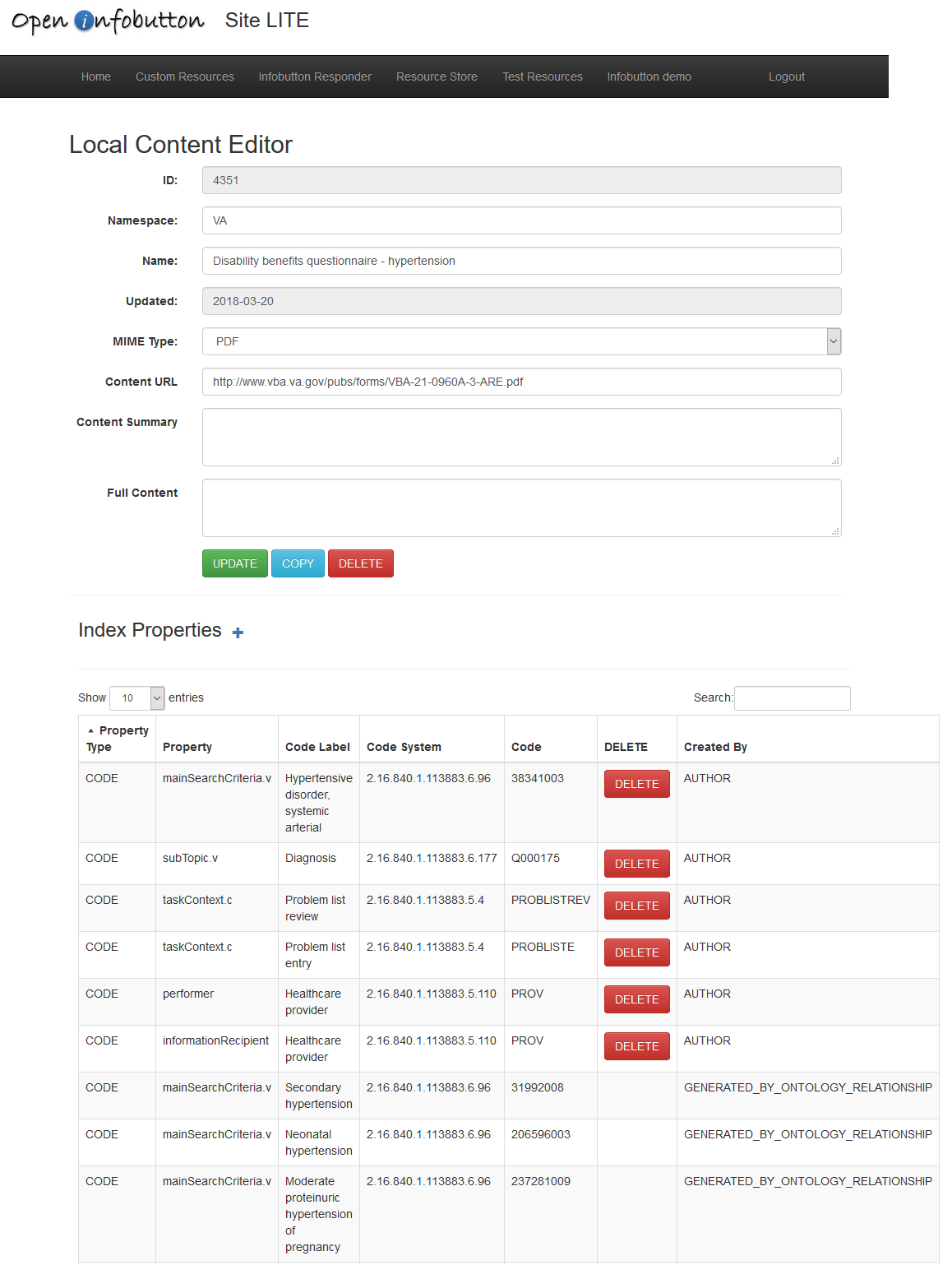
Within LITE, users can manually map to individual resources through the Infobutton Responder, which is useful for health information systems that have internal non-HL7 compliant resources that they want to make available to the patient (Fig. 15). However, manual mapping is not the most efficient route to making Wikipedia an infobutton knowledge resource. Wikidata provides a route to circumvent manual mapping.



*Figure 13. OpenInfobutton test tool that generates HL7 requests*



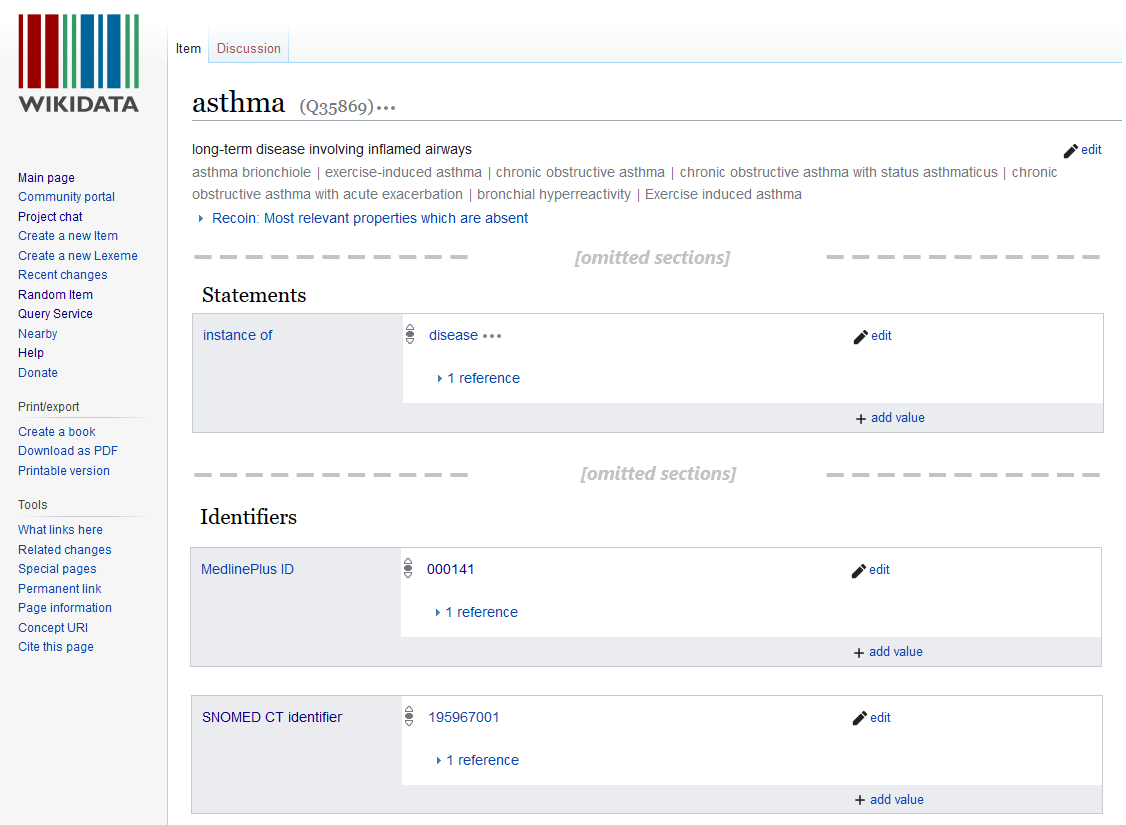
*Figure 14. LITE Resource Store*



*Figure 15. Infobutton Responder in LITE*

### Wikidata

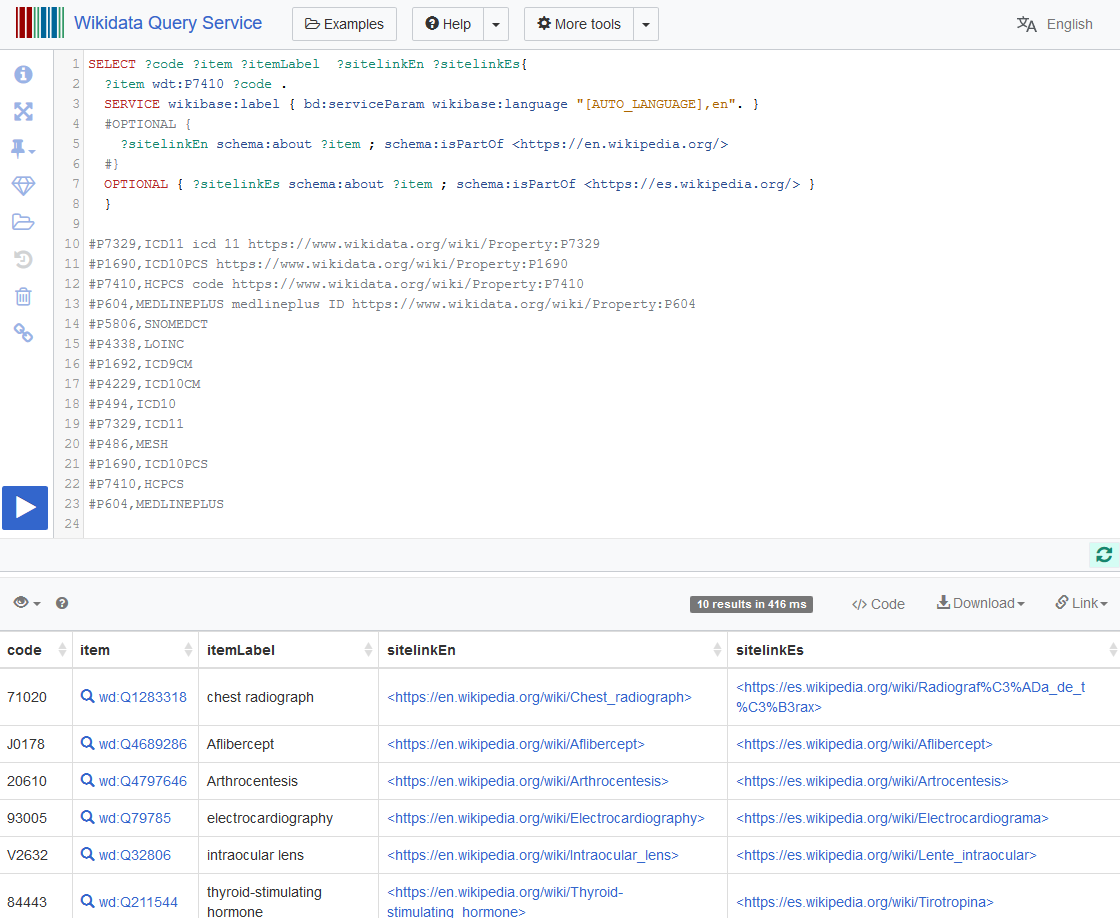
Wikidata is a semantic editable and computable database created in 2012 that links to Wikipedia articles.10 It provides the infrastructure and structured data needed to route patients from their PHRs to Wikipedia. Wikidata consists of items (Q-items) and properties. Each Q-item serves as a concept in Wikidata ontology and contains properties that describe the relationship between items. For example, Fig. 16 illustrates the Q-item “asthma” (Q35869), which contains a property called “instance of” (P31) with the value being “disease” (Q12136). Each entry of a property and a value is called a statement, and if applicable, each statement preferably has a reference to address reliability. Another class of properties are identifiers. Databases and systems may describe a given item using different terminologies; MedlinePlus’ identifier for asthma is 000141, while SNOMED-CT’s identifier for asthma is 195967001. Wikidata Q-items collect these various identifiers, and subsequently provides access points for information systems that use medical terminologies like the infobutton.



*Figure 16. Example of a Wikidata Q-item. Sections have been omitted to illustrate the general make-up of a Q-item.*

#### Wikidata Query Service and SPARQL

Wikidata can provide the index that infobutton needs through its query service, which utilizes the SPARQL language. Queries would produce indexes of health-related Q-items, related medical terminology identifiers, and the URL for the corresponding Wikipedia article in multiple languages (Fig. 17). If a Q-item contains all medical identifiers, all health information systems should be able to navigate to the Wikipedia article. However, Wikidata currently does not have the comprehensive coverage needed to be a robust infobutton knowledge resource.



*Figure 17.* [*Query on all Q-items with HCPCS identifiers and the link to the corresponding Wikipedia article*](https://w.wiki/FTH) *in English and Spanish.*

#### Current state of WD

|  |  |
| --- | --- |
| Terminology | n |
| HCPCS | 10 |
| ICD10 | 5283 |
| ICD10CM | 10515 |
| ICD10PCS | 13 |
| ICD11 | 80 |
| ICD9CM | 5192 |
| LOINC | 184 |
| MEDLINEPLUS | 1508 |
| MeSH | **33232** |
| SNOMEDCT | 32 |

*Table 1. Number of Q-items that contain the medical terminology in question.*

Table 1 lists the medical terminologies and the number of Q-items that exist for each (note: Wikidata’s editable nature means numbers can be in flux). The numbers vary drastically, with MeSH being the terminology that has been assigned the most (Fig. 18) and HCPCS being the least (possibly affected by copyright). The queries that generated these numbers can be found [here](https://w.wiki/FTK).

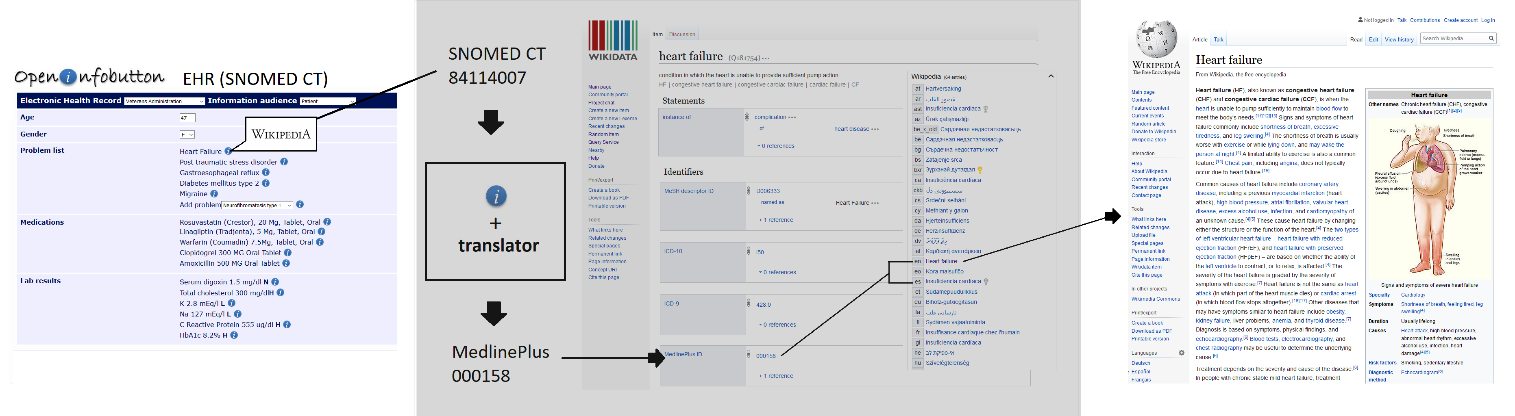
*Figure 18. Number of entries for each medical terminology*

Ideally, Wikidata would contain all identifiers from all terminologies to ensure that any PHR item could link to the correct Wikipedia article.

# Discussion

For infobutton to effectively route to Wikipedia, two avenues can be taken:

1. **Strive towards comprehensive coverage**: Each medical Wikidata Q-item includes all relevant medical identifiers. This option would require a mass upload of data to create missing Q-items and map terminologies, or any external data item, to its correct Wikidata Q-item. Subsequently, the data would need to be refreshed as terminologies are updated periodically.
2. **Build a translating application**: This avenue is a response to how Wikidata is incomplete. Since health-related Q-items likely contain only a few, if any, medical identifiers, we need a way to route to the Wikipedia article using the identifiers that are already listed on the Q-item. The application would work with an infobutton manager and the EHR to produce a list of synonymous and hierarchically relevant codes to run against what is available in the Wikidata index. Once a match is found, infobutton will retrieve the appropriate Wikipedia link to display on the patient’s interface (Fig. 19).



*Figure 19. An illustration of how Option B would work. OpenInfobutton’s demo tool on the left emulates what a patient would see through their EHR and the gray section outlines backend activity.*

Although these avenues can be taken simultaneously, Option A is the preference. Since MeSH has the highest coverage, increasing MeSH identifiers to near-completion would set up a testing environment to showcase Wikipedia as an infobutton knowledge resource.

### Wikidata Tools and Resources

To support the next iteration of this project, we summarize a few resources that will be useful or should be investigated further.

#### QuickStatements and Bots

User:Magnus Manske authored a tool called QuickStatements that can create and import statements and references for Q-items through text commands. QuickStatements can be used to increase the coverage of medical identifiers (Option A). Large batches are executed through Wikimedia’s server, and therefore, may run into server traffic. A large batch may also be labeled as a User:QuickStatementsBot, and therefore may need to abide by Wikidata’s bot protocol.

#### Mix’n’match

Another tool created by User:Magnus Manske is Mix’n’match which aims to serve Galleries, Libraries, Archives, and Museums (GLAM). The tool is used to import datasets from external knowledge bases, such as MeSH terms, and match the data to corresponding Wikidata items. Matching is necessary because external databases are structured differently, and we want to make sure the external item is referring to the same thing as the Wikidata item. We had corresponded with Daniel Mietchen from the University of Virginia to learn his experience with using Mix’n’Match to import the “MeSH Health Care” catalog. He explained that Mix’n’Match was the easiest option at the time of the import, and he suggested that OpenRefine has evolved into the better tool for imports.

#### OpenRefine

OpenRefine, a free data wrangling tool, can link tabular data from external knowledge bases with Wikidata items. The differences between Mix’n’match and OpenRefine were not thoroughly researched, but OpenRefine may have more capacity. At NLM, the use of OpenRefine needs to be cleared through certain security measures.

#### WikiProject Medicine

Wikipedia users can join WikiProjects, or teams that work to improve specific topic areas. WikiProject Medicine is a group that can be mobilized to support this project, and as a first step, we’ve created a subproject page titled, “[Wikipedia and Infobutton](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine/Wikipedia_and_Infobutton).” Correspondence should continue with Lane Raspberry and Daniel Mietchen, members of WikiProject Medicine. They mentioned another colleague named Aidong Zhang who is also looking into linking Wikidata more closely with MeSH.

# Next Steps

This project explored the current status of personal health records and feasibility of linking patients to a knowledge resource that has not been previously explored.

As possible follow-up work, we identified the following tasks:

1) Evaluate MeSH links for accuracy. Currently, over 30,000 articles are linked with a MeSH term. The accuracy of these links needs to be verified to ensure that an infobutton manager links to the correct item.

2) Use computational and tool-based methods (not manual editing) to increase the number of Wikipedia articles that contain an annotation in Wikidata suitable for infobutton. This process involves creating Q-items in Wikidata that link to Wikipedia articles, and adding identifiers to said Q-items.

3) Correspond with WikiProject Medicine, including Daniel Mietchen and Lane Raspberry. Daniel brought up how Wikidata can only ingest public domain content, meaning SNOMED CT and HCPCS’s copyright needs to be investigated.

4) Investigate the Medical Resources box created by User:RexxxS. This box seems to be supplied by the medical identifiers listed in Wikidata Q-items.

# 

# References

1. Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. *N Engl J Med*. 2010;363(6):501-504. doi:10.1056/NEJMp1006114

2. Cook DA, Teixeira MT, Heale BS, Cimino JJ, Del Fiol G. Context-sensitive decision support (infobuttons) in electronic health records: a systematic review. *J Am Med Inform Assoc*. August 2016:ocw104. doi:10.1093/jamia/ocw104

3. Teixeira M, Cook DA, Heale BSE, Del Fiol G. Optimization of infobutton design and Implementation: A systematic review. *J Biomed Inform*. 2017;74:10-19. doi:10.1016/j.jbi.2017.08.010

4. Long J, Hulse NC, Tao C. Infobutton usage in Patient Portal MyHealth. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci*. 2015;2015:112-116.

5. Candelario DM, Vazquez V, Jackson W, Reilly T. Completeness, accuracy, and readability of Wikipedia as a reference for patient medication information. *J Am Pharm Assoc*. 2017;57(2):197-200.e1. doi:10.1016/j.japh.2016.12.063

6. Thomas GR, Eng L, de Wolff JF, Grover SC. An Evaluation of Wikipedia as a Resource for Patient Education in Nephrology: WIKIPEDIA FOR RENAL PATIENT EDUCATION. *Semin Dial*. 2013;26(2):159-163. doi:10.1111/sdi.12059

7. Shafee T, Masukume G, Kipersztok L, Das D, Häggström M, Heilman J. Evolution of Wikipedia’s medical content: past, present and future. *J Epidemiol Community Health*. August 2017:jech-2016-208601. doi:10.1136/jech-2016-208601

8. Cimino JJ, Jing X, Del Fiol G. Meeting the electronic health record “meaningful use” criterion for the HL7 infobutton standard using OpenInfobutton and the Librarian Infobutton Tailoring Environment (LITE). *AMIA Annu Symp Proc AMIA Symp*. 2012;2012:112-120.

9. Jing X, Cimino JJ, Del Fiol G. Usability and Acceptance of the Librarian Infobutton Tailoring Environment: An Open Access Online Knowledge Capture, Management, and Configuration Tool for OpenInfobutton. *J Med Internet Res*. 2015;17(11):e272. doi:10.2196/jmir.4281

10. Turki H, Shafee T, Hadj Taieb MA, et al. Wikidata: A large-scale collaborative ontological medical database. *J Biomed Inform*. 2019;99:103292. doi:10.1016/j.jbi.2019.103292