

Making clinical trials available at the point of care - connecting Clinical trials to Electronic Health Records using SNOMED CT and HL7 InfoButton standards

Jay Kola¹*^[0000-0002-7584-5003], Wai Keong Wong² and Bhavana Buddala¹

¹ Termlex Limited, Spaces, The Porter Building, Slough, SL1 1FQ, United Kingdom

² University College London Hospitals, 250 Euston Road, NW1 2PJ, United Kingdom
jay@termlex.com

Abstract. Making clinical trials discoverable at the point of care (patient encounter) is one of the holy grails of connecting clinical research with clinical practice [1, 2]. Semantic interoperability standards designed for hospital systems do not interface well with clinical trials, which are predominantly unstructured/free text. In this paper, we describe our experiences of using SNOMED CT and HL7 InfoButton standards to make clinical trials from a trial registry accessible to clinicians within an Electronic Health Record (EHR) system in University College Hospitals, London. In particular we discuss the use of HL7 InfoButton standard [15] as a standardised interface for a clinical trials repository, which we believe is a first of its kind in the UK. We discuss some of the barriers to making clinical trials more accessible in EHR systems, including considerations for using standards and associated challenges & opportunities.

Keywords: Clinical trials, SNOMED CT, HL7 InfoButton, Trial eligibility, Keytrials.

1 Introduction

1.1 Connecting clinical research to clinical practice

There is extensive literature that highlights how despite clinical research and trials being vital to advances in clinical medicine [1, 2], multiple challenges exist in patient recruitment [3], physician participation [3, 4] and identification of patient eligibility [4]. One of the key challenges in both patient recruitment and physician participation is the ability to expose existing local clinical study information (e.g. *eligibility, recruitment status*) to providers and patients [5]. While external trial registries like ClinicalTrials.gov [6] and UK Clinical Trials Gateway [7] exist, site-specific information in these registries are often not kept updated with on-going studies. At times coverage of on-going trials in external registries can be less than 50% [8]. In other cases, information in external registries might not be kept up to date with changes to the study.

In this paper (written as application notes), we describe our experience of creating *Keytrials*, a clinical trials discovery platform, designed to make local clinical

trials accessible to physicians and patients. While Keytrials makes existing clinical trials (either local or imported from external registries) available to users via a web (REST¹) API², our objective was to integrate trial-matching (on-demand) into the Electronic Health Record (EHR) system. There have been past attempts at creating electronic solutions and novel specifications for making local registries accessible to external consumers [5, 9, 10] including EHR systems. However, we based our integration between the EHR system and Keytrials on existing healthcare standards like SNOMED CT [11] and HL7, that are already in use in Electronic Health Record (EHR) systems within our setting and also internationally.

2 Keytrials Platform

Keytrials is an open source clinical trials discovery platform, designed to make it easier for clinicians and patients to find trials that are open, with a goal to increase trial recruitment and improve visibility of clinical trial activity at University College London Hospital (UCLH), UK. Keytrials is built using modern Web 2.0 and Java enterprise technologies. There is a clean separation of its backend layer from the user interface and backend layers using REST APIs as shown in Figure 1. This makes it easy for other 3rd party applications and other apps to plugin into the RESTful service layer.

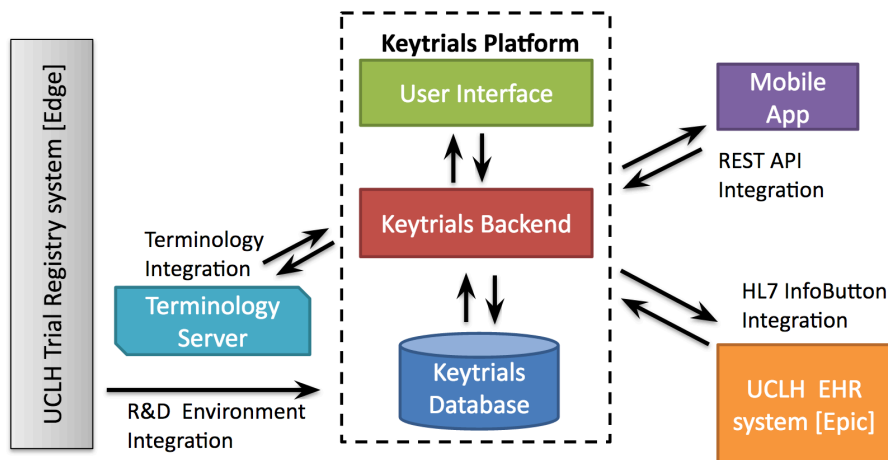


Fig. 1. Overview of Keytrials Platform, showing its integration with local Trials Registry and hospital EHR System.

¹ REST – Representational State Transfer

² API – Application Programming Interface

³ R&D – Research and Development

⁴ However, since morbidity and mortality information in hospital systems has traditionally been

² ~~Applied as a standard for reporting to the World Health Organisation (WHO), aspects~~ of the clinically relevant information (e.g. diagnosis, age, gender, interventions, etc.) tend to

For the purposes of this paper, three aspects of Keytrials are of interest – R&D³ Environment Integration, Terminology Integration and HL7 InfoButton Integration. Together, these three functionalities allow local (or remote) trials to be accessible for trial-matching within the EHR, at the point of care.

2.1 R&D Environment Integration

This functionality allows Keytrials to import existing trials from a clinical trials registry. Within UCLH, existing trials are held in a local trial management system (Edge), which acts as the primary source of trials. However, Keytrials also allows existing trials to be imported from remote registries like ClinicalTrials.gov.

2.2 Terminology Integration

This functionality allows Keytrials to access a centralized `terminology server` that provides search (lookup) functionality for healthcare terminologies like SNOMED CT. Keytrials uses both the terminological content (e.g. *concept ids, descriptions*, etc.) and the semantic relationships within SNOMED CT. For example, when users can search for disease conditions they can search for matches using the preferred terms (*small cell lung cancer*) or synonyms (*oat cell carcinoma of lung*). Both return the exact trials, since the terminology server resolves them to the same SNOMED CT concept. Keytrials also uses the underlying semantics of SNOMED CT as part of returning matches. For example, if a user searches for `*Plasma Cell Neoplasm*`, it will also bring back `*Multiple Myeloma*` even though there is no textual match between *Plasma Cell Neoplasm* and *Multiple Myeloma*. It does this because in SNOMED CT, *Multiple myeloma* is defined as a type of *Plasma Cell Neoplasm* - which makes results more intuitive to our clinician users. A longer discussion of how SNOMED CT as a standard is implemented in our workflow is discussed in section 3.4.

2.3 UCLH EHR System

UCL Hospitals (UCLH) have recently implemented Epic as their EHR system across all clinical specialties. As part of this roll out, UCLH decided to adopt SNOMED CT as the reference terminology for their EHR, in keeping with the national requirements in the UK. However, instead of natively using SNOMED CT to populate their diagnosis, UCLH procured a 3rd party content provider that provides an interface terminology system for clinicians to use. This is mapped to ICD 10 [12] and SNOMED CT. However, Epic does not currently support the transmission of SNOMED CT concept ID via the Infobutton interface. Instead it can only provide the ICD 10 code. So when Keytrials interfaces with Epic, it receives ICD codes instead of SNOMED CT codes. Keytrials then uses the `*terminology server*` to translate these ICD codes into their SNOMED CT equivalents as needed.

³ R&D – Research and Development

3 Standards based Integration with EHR

3.1 SNOMED CT Annotation of Trials

Trials that have been imported into Keytrials have both structured (defined) data elements (e.g. *status*, *open date*, *closing date*, etc.) and unstructured elements (e.g. *eligibility criteria*, *description/summary of trial*, etc.). In order to match suitable trials with existing patient details (e.g. *age*, *disease conditions*, *gender*), it is often the eligibility criteria of a trial that are of most relevance. However, most of this information is provided as *free-text* in trial, which is not coded to any *standardised* medical vocabulary/terminology. As described above, the EHR itself is coded in either ICD or SNOMED CT – leading to situation where trial-matching will require the clinical trials to also be *coded* using the same coding system. As part of the project, we use Bio-YODIE [13], a *Natural Language Processing* (NLP) engine to annotate clinical trials with their corresponding disease conditions. The results of this NLP process are clinical trials with associated disease conditions coded in SNOMED CT. These *annotated trials* are then stored in Keytrials, making them available for subsequent queries.

3.2 HL7 InfoButton interface with EHR

Context-dependent *infobuttons* have been proposed & used for displaying contextually relevant knowledge resources within EHRs [14]. This approach for integrating online knowledge resources with EHRs has been standardized by HL7 as the InfoButton standard [15]. The InfoButton standard allows systems (e.g. EHR systems) to request information from *knowledge resources* using a standardised *reference model* which can be expressed as a series of URL (Uniform Resource Locator) query parameters and values. These requests can then be sent to the *knowledge resource* using Hyper Text Transfer Protocol (HTTP) technologies. A limited subset of these InfoButton standardised URL parameters are shown in table 1 below.

Table 1. Selected subset of InfoButton URL parameters relevant for clinical trials

URL parameter name	Description	Code systems
Main search criteria	The main clinical concept of interest in a knowledge request (e.g., a medication, a laboratory test result, a problem)	ICD, SNOMED-CT
Gender	The patient's gender	HL7 administrative gender
Age	The patient's age as a value and a unit	Not Applicable
Task Context	The action the user is performing in a clinical information system when a knowledge request is triggered (e.g., order entry, laboratory results review, problem list review)	HL7 Act Code

HL7 InfoButton has been used to varying degrees of success in EHR systems for clinical decision support, medication alerts and for allowing access to online references [17]. It has more recently also been used to integrate genomic resources within EHRs to mixed success [18]. However, since its inclusion in the 'meaningful use' certification in the US [19], major EHR vendors support its use out of the box. Within our project, Epic the EHR system in use in UCLH supports InfoButton based requests, making it quite attractive as a way for accessing trial information held in Keytrials. This in effect, turns Keytrials into a knowledge resource for clinical trials and allows us to use InfoButton URL queries to access trials appropriate for a patient.

3.3 InfoButton Queries for Clinical Trials in Keytrials

Using the URL query parameters specified in the standard, it is possible to create a InfoButton request to a knowledge resource as below:

<https://locationofresourcehere.com?age.v.u=a&age.v.v=78>

The above request specifies that the value of 'age' as '78'. A slightly more realistic query being sent to a test server for Keytrials would look like:

<https://uat.keytrials.com/#/trial?age.v.u=a&age.v.v=78&ageGroup.v.c=D000368&mainSearchCriteria.v.c=C34>

This translates to a query for all matching trials suitable for a patient of *age 78 years* and an ICD-10 diagnosis of '*Lung Cancer*' (C34).

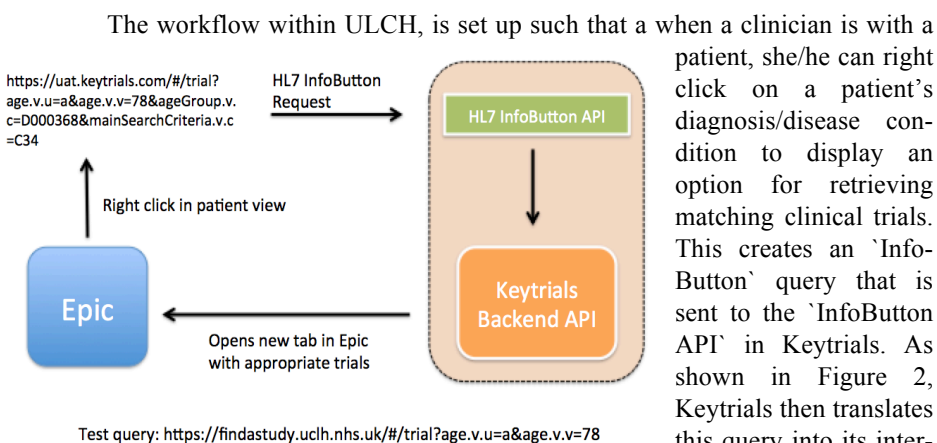


Fig. 2. Overview of InfoButton based integration between Keytrials and the EHR system.

figure the EHR system to display these matching trials in a separate built-in browser tab. This allows the physician in effect to perform trial-matching at the point of care, directly from the EHR.

The next section describes how this InfoButton query with the ICD-10 diagnosis code is translated into the semantic equivalent of *'all descendants of Lung cancer'* using SNOMED CT via the terminology server.

3.4 Semantic Search of Trials using SNOMED CT (via a terminology server)

In the above section we explained how the criteria for finding a clinical trial are passed to Keytrials platform using HL7 InfoButton request parameters. One notable part of these InfoButton request parameters is the *'mainSearchCriteria.v.c'* parameter, which represents the *'code'* in the coding system used for identifying concepts of interest (e.g. disease diagnosis, procedures, etc.). In Epic, ICD-10 is used to code diagnosis. For example in Figure 2, the fragment after this parameter with value C34 is the ICD-10 code for *'Lung Cancer'*. When this InfoButton query is sent to Keytrials, it parses this query and extracts the code *'C34'*. Since this request parameter is known to contain a *'code'*, this is sent to the Terminology Server for lookup. Within the Terminology Server, this code is associated with ICD-10 and we use a *'cross-map'* to go from ICD-10 to SNOMED CT. This transform from ICD-10 to SNOMED CT and associated issues are described in section 4.2.

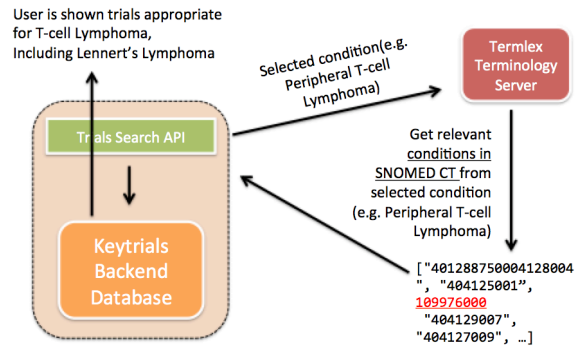


Fig. 3. Use of Terminology server in Keytrials – to retrieve all types/descendants of T-cell Lymphoma, including Lennert's Lymphoma (shown in red).

Once we find an equivalent SNOMED CT concept for an ICD-10 code, we perform a *'semantic expansion'* based on the meaning of this SNOMED CT concept. For example, when the query is for *'T-cell Lymphoma'*, we know that in most cases the user is expecting trials for all types of *'T-cell Lymphomas'*. Our terminology server calculates this *'semantic expansion'* (*transitive closure*) on the fly and returns all transitive sub-types (descendants) for that concept. We refer to this *'semantic expansion during search'* as *'semantic search'*. This *'semantic search'* based on SNOMED CT has the added benefit of picking up concepts that would otherwise have been missed by *'text-based'* search alone. For example, in Figure 3, we are able to include trials for *'Lennert's Lymphoma'* as part of *'T-cell Lymphoma'* trials, since in SNOMED CT it is declared as a sub-type of *'T-cell Lymphoma'*. Any *'text-based'* search for *'T-cell Lymphomas'* would have likely missed *'Lennert's Lymphoma'* as it does not have the token *'T-cell'* in it.

4 Discussion

Since the ability to support queries is based on the HL7 InfoButton and SNOMED CT standards, we believe our approach should be adoptable by other investigators. We believe that within the UK we are the first project to adopt InfoButton and SNOMED CT standards for accessing clinical trials from an EHR system. This approach however was not without issues given how clinical trials and clinical medicine do not often support the same standards. We share some of our experiences in this section. These challenges can be separated into trials related and EHR related issues.

4.1 Issues with Clinical Trials data

We have previously mentioned how existing large registries of trials have issues in staying up to date with trials that are on-going and open for recruitment. This continues to be a problem even in smaller registries. In our project, we were forced to build a batch import integration between the local trial management system and Keytrials. This batch import is currently run weekly to ensure that Keytrials is kept in sync with the updates to local trial registry. We however recognise that creating integrations for multiple local trial registry systems will be expensive as every system will likely have its own internal representation. Standards based interchange format would help simplify this task. While CDISC-ODM [20] exists, it is tied to the operational workflow of running clinical trials as opposed to specifying the data standards for trials. In the future term, we believe that HL7 FHIR might evolve to become a standardised representation for clinical trials [21, 22]. However, this current specification of a `ResearchStudy` is still in early stages of development [23].

A further issue with making clinical trials accessible to EHR systems is the inability to explicitly specify eligibility criteria (*inclusion, exclusion criteria, disease conditions, interventions, etc.*) as structured/coded entities. While FHIR seems to allow this level of specification in the future, a vast number of existing studies are free-text based, limiting the ability to automatically match trials to coded diagnosis, age or other information in EHR systems. This limitation can be overcome using NLP as adopted within our project and other initiatives [24 - 27]. However, this approach of post-processing and annotating trials could be avoided if clinical trials registries could facilitate the coding of eligibility criteria at the time of trial registration.

4.2 Issues with EHR data

Similar to the state of clinical trials ecosystem, the landscape in EHRs is still riddled with large amounts of un-coded and unstructured free-text information⁴. While having

⁴ However, since morbidity and mortality information in hospital systems has traditionally been coded using ICD for statutory reporting to the World Health Organisation (WHO), aspects of the clinically relevant information (e.g. diagnosis, age, gender, interventions, etc.) tend to be coded more commonly.

ICD-10 used for coding diagnosis provides a slightly better starting point for integrating EHRs with clinical trials, often the level of granularity required by researchers and physicians interested in research is not provided by ICD as it was primarily designed for statistical reporting.

SNOMED CT is starting to see adoption across the globe and in the UK, but within our project we note that Epic did not support SNOMED CT natively. This meant that when we had to integrate our EHR (coded using ICD-10) with clinical trials (annotated using SNOMED CT via NLP), we were forced to use ICD codes as part of the `mainSearchCriteria` attribute in InfoButton to send diagnosis codes to Keytrials. This required a workaround within Keytrials, where all ICD codes passed via InfoButton were then processed by the `terminology server` to convert them into corresponding SNOMED CT codes. As knowledgeable readers will note, going from ICD-10 to SNOMED CT will often result in a *lossy* transform, as SNOMED CT is often more granular/specific than ICD. This *lossy transform* and incorrect use of the semantics of SNOMED CT while perhaps not immediately relevant for trial-matching is likely to become more important when automated trial-matching becomes more prevalent. We believe that with greater adoption of SNOMED CT, we will likely see native use of this standard in EHR systems in the future so these *lossy* transforms can be avoided.

While not immediately part of the EHR issues, we also noted within our project that the use of SNOMED CT presented interesting challenges. For example, in SNOMED CT searching for `adenocarcinoma` might present two exact matches – one of them being a *morphological abnormality* and the other being a *disorder* making it confusing for users as to which match to select. This can easily be addressed by ensuring that only relevant SNOMED CT hierarchies are included by default during search – in this case only including `clinical findings` hierarchy from SNOMED CT. However, it should also be noted that even within `clinical finding` hierarchy, exactly named matches could sometimes appear. For example, searching for `fatigue` might return a *symptom* and a *disorder*, both of which are part of the `clinical finding` hierarchy. Needless to say, like all clinical information systems using a terminology, a degree of clinical assurance is required to improve usability.

However, on the whole using a combination of SNOMED CT and InfoButton has provided a degree of assurance and flexibility within our project. We believe that as clinical trials registries and EHR systems continue to mature, standards based integration will continue to become more prevalent and a lot more plug-n-play.

5 Conclusions

In this paper we shared our experience of using existing healthcare standards SNOMED CT and HL7 InfoButton to make data in a clinical trials accessible to EHR systems. While InfoButton has been used with mixed results in other domains, it has

not been previously been used to access clinical trials in the UK. One major challenge in making clinical trials discoverable and connecting them to EHRs is the lack of standardisation of trial eligibility criteria – with most being just un-coded, free-text content. This is a barrier for matching patients to eligible trials, even if relevant information (*coded diagnosis, age, gender, etc.*) is already available within the patient record in the EHR system. We used an NLP approach to annotate eligibility criteria (e.g. disease conditions) in SNOMED CT, thereby allowing us to use a fuller range of InfoButton query parameters to match trials to patients directly from the EHR system.

Since our approach is based on international standards, we believe it could serve as a means of creating reusable integrations between clinical trial registries with EHR systems. However, the lack of standardisation of clinical trials might mean that significant effort is required to integrate a clinical trials registry needs to a HL7 Info-button compliant EHR system. We note that a standardised specification of clinical trials could make this integration less onerous. However, existing standards for clinical trials do not yet specify this level of detail (CDISC-ODM) and others are not yet sufficiently mature to meet this need (HL7 FHIR). A similar, albeit slightly different problem exists within EHR systems where relevant information is coded but in ICD-10, which does not always provide the level of detail required for clinical research. However the increasing adoption of SNOMED CT in this space will likely solve that issue, even if SNOMED CT itself comes with its own set of challenges. We hope that as standards for clinical trials and EHRs mature and become more widely adopted, it will be possible to make clinical trials discoverable at the point of care in EHR systems using a plug-n-play model.

Acknowledgements The authors would like to acknowledge that UCLH BRC (Biomedical Research Centre) and CRIU (Clinical Research Informatics Unit) funded development of the Keytrials platform.

References

1. Campbell EG, Weissman JS, Moy E, Blumenthal D. Status of clinical research in academic health centers: views from the research leadership. *JAMA*. 2001;286(7):800–6
2. Rindfleisch TC, Brutlag DL. Directions for clinical research and genomic research into the next decade: implications for informatics. *J Am Med Inform Assoc*. 1998;5(5):404–11
3. Siminoff LA, Zhang A, Colabianchi N, Sturm CM, Shen Q. Factors that predict the referral of breast cancer patients onto clinical trials by their surgeons and medical oncologists. *J Clin Oncol*. 2000;18(6):1203–11
4. Embi PJ, Jain A, Clark J, Harris CM. Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. *AMIA Annu Symp Proc*. 2005;2005:231–235.
5. Stahl DC, Evans RM Jr, Afrin LB, DeTeresa RM, Ko D, Mitchell K. Web services-based access to local clinical trial databases: a standards initiative of the Association of American Cancer Institutes. *AMIA Annu Symp Proc*. 2003;2003:624–628.
6. Clinical Trials Gov Site, <https://clinicaltrials.gov> last accessed 2019/09/21
7. UK Clinical Trials Gateway, <https://bepartofresearch.nihr.ac.uk/> last accessed 2019/09/21

8. Manheimer E, Anderson D. Survey of public information about ongoing clinical trials funded by industry evaluation of completeness and accessibility. *BMJ* 2002 Sep 7; 325
9. Embi PJ, Jain A, Clark J, Harris CM. Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. *AMIA Annu Symp Proc.* 2005;2005:231–235.
10. Huang, Z., Ten Teije, A., & Van Harmelen, F. (2013). SemanticCT: a semantically-enabled system for clinical trials. In *Process Support and Knowledge Representation in Health Care* (pp. 11-25). Springer, Cham.
11. Stearns, Michael Q., et al. "SNOMED clinical terms: overview of the development process and project status." Proceedings of the AMIA Symposium. American Medical Informatics Association, 2001.
12. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1992.
13. Gorrell, Genevieve, Xingyi Song, and Angus Roberts. "Bio-YODIE: A Named Entity Linking System for Biomedical Text." *arXiv preprint arXiv:1811.04860* (2018).
14. Cimino JJ, Elhanan G, Zeng Q. Supporting infobuttons with terminological knowledge. In: Proc AMIA annu fall symp; 1997. p. 528–32.
15. Context-aware knowledge retrieval (infobutton) product brief. HL7 International Wiki Site. https://wiki.hl7.org/index.php?title=Product_Infobutton, accessed 2019/09/21.
16. Retrieval, HL7 Context-Aware Knowledge. "Knowledge Request URL-Based Implementation." *Health Level Seven Int'l specification, Jan* (2010).
17. Del Fiol, Guilherme, et al. "Implementations of the HL7 Context-Aware Knowledge Retrieval ("Infobutton") Standard: challenges, strengths, limitations, and uptake." *Journal of biomedical informatics* 45.4 (2012): 726-735.
18. Heale, Bret SE, et al. "Integrating genomic resources with electronic health records using the HL7 Infobutton standard." *Applied clinical informatics* 7.03 (2016): 817-831.
19. US Department of Health and Human Services. "2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications." (2015): 1-159.
20. Kuchinke, Wolfgang, et al. "CDISC standard-based electronic archiving of clinical trials." *Methods of information in medicine* 48.05 (2009): 408-413.
21. Bender, Duane, and Kamran Sartipi. "HL7 FHIR: An Agile and RESTful approach to healthcare information exchange." *Proceedings of the 26th IEEE international symposium on computer-based medical systems.* IEEE, 2013.
22. Leroux, Hugo, Alejandro Metke-Jimenez, and Michael J. Lawley. "Towards achieving semantic interoperability of clinical study data with FHIR." *Journal of biomedical semantics* 8.1 (2017): 41.
23. HL7 FHIR Research Study Specification, <https://www.hl7.org/fhir/researchstudy.html>, accessed 2019/09/21
24. Wu, Menghua, et al. "Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials. gov." *Pacific Symposium on Biocomputing Co-Chairs.* 2014.
25. Pffnner, Pascal B., et al. "ClinicalTrials. gov as a data source for semi-automated point-of-care trial eligibility screening." *PloS one* 9.10 (2014): e111055.
26. Deleger, Louise, et al. "Building gold standard corpora for medical natural language processing tasks." *AMIA Annual Symposium Proceedings.* Vol. 2012. American Medical Informatics Association, 2012.