# The MedRed Ontology for Representing Clinical Data Acquisition Metadata

Jean-Paul Calbimonte[1], Fabien Dubosson[1], Roger Hilfiker[2], Alexandre Cotting[1], and Michael Schumacher[1]

[1]Institute of Information Systems, e-Health Unit
[2]Institute of Health Sciences
University of Applied Sciences and Arts Western Switzerland,
HES-SO Valais-Wallis, Sierre, Switzerland
{name.surname}@hevs.ch

**Abstract.** Electronic Data Capture (EDC) software solutions are progressively being adopted for conducting clinical trials and studies, carried out by biomedical, pharmaceutical and health-care research teams. In this paper we present the MedRed Ontology, whose goal is to represent the metadata of these studies, using well-established standards, and reusing related vocabularies to describe essential aspects, such as validation rules, composability, or provenance. The paper describes the design principles behind the ontology and how it relates to existing models and formats used in the industry. We also reuse well-known vocabularies and W3C recommendations. Furthermore, we have validated the ontology with existing clinical studies in the context of the MedRed project, as well as a collection of metadata of well-known studies. Finally, we have made the ontology available publicly following best practices and vocabulary sharing guidelines.

## 1 Introduction

Clinical research activities require the involvement of heterogeneous individuals of a given population, needed to assess and validate biomedical hypotheses concerning behavior, treatments, interventions and other studies. Clinical trials and other such studies can be complex and span long periods of time, and the data acquisition process requires careful management and accuracy. Although in the past, manually filled forms were the norm for acquiring data in this context, nowadays the use of Electronic Data Capture (EDC) solutions has shown to improve the efficiency of the process, while maintaining quality and accuracy standards [3, 19]. In particular, EDC helps reducing and/or eliminating data transcription and transmission times, providing data validation and input enforcement, or helping scheduling the site visits [5, 7]. Furthermore, EDC provides faster access to data in running studies, which can help to perform live-analytics over the acquired datasets. Due to these benefits, clinical research organizations, pharmaceutical companies, and university hospitals, among others, make use of EDC and related systems such as OpenClinica, REDCap, TrialDB, InForm, Medidata Rave or Datatrak [16].

As an example, consider an osteoarthritis study performed by the Phisiotherapy Lab at HES-SO Valais-Wallis, on the local population. The implementation of this study may include the usage of several *instruments*, such as questionnaires over a selected group of patients, each of which contains several *sections*, *questions*, and *variables* to

be annotated and recorded. The study can be divided in different *arms* where diverse methods are applied for comparison purposes; and furthermore, it can be split in repeated *events* over time, using similar instruments for evolution tracking. Such study could reuse well known and validated instruments, such as the HOOS Hip survey [14], or extend it with additional instruments, sections, and variables.

Given the large number of clinical studies that are performed worldwide, and their complexity, it has become a need to share their results, as well as their structure and metadata. This would enable: validating existing protocols, reusing and refining clinical research instruments, extending previous studies, performing surveys and systematic analytics of clinical trials, etc. However, to achieve this, it is first necessary to tackle the heterogeneity issues regarding the description and representation of these studies. The most used format for representing studies in EDC software, ODM (Operational Data Model) [9], lacks a semantically-rich model able to address the aforementioned challenges, and is therefore insufficient as a foundational model for achieving semantic interoperability for clinical studies and trials.

In this paper we present the MedRed Ontology, a semantically-rich model designed to represent the metadata of clinical studies, including the definition of its constituting instruments, the different steps of each one, their organization in arms and events, as well as the data variables captured using them. Thanks to its integration with existing vocabularies (PROV-O [11] and P-Plan [6]), the MedRed ontology can also capture complex relationships among instruments and studies, including composition, derivation, authoring, and versioning. These features make it possible to track changes of a study across time, or to indicate that a study was designed based on an existing one. MedRed also includes the representation of validation conditions on the clinical instruments, using the SHACL language [8] for representing constraints. The MedRed Ontology has been validated using pilot studies led by the Institute of Health of HES-SO Valais-Wallis, in the context of the MedRed data lifecycle project[1]. It has also been applied to a heterogeneous collection of study metadata descriptions extracted from the REDCap [7] library of health studies and instruments. Finally, MedRed has been made publicly available under standard formats, on a permanent URL, and following ontology publication guidelines.

## 2 Related Work

Ontologies for clinical studies have been developed in recent years, typically focusing on the description of different types of studies, including taxonomies and classifications [17]. The OBO Foundry [18] contains several biomedical ontologies, some of which are related to the description of studies. Examples include the Ontology for Biomedical Investigations, Clinical Measurement Ontology, and the Informed Consent Ontology. However, these are more specific to biomedical document descriptions, measurements, and consent information, respectively. The Bioportal repository also contains relevant ontologies, e.g. Clinical Trials Ontology, which contains a large vocabulary of clinical trial types. Other ontologies in Bioportal (e.g. MESH, SNOMED, HL7) include general references to clinical study concepts, but do not provide detailed descriptions of them.

[1] MedRed Project: `http://w3id.org/medred/project`

Clinical Data capture software are widely used today as a backbone technology for data acquisition in research studies. Professional tools include OpenClinica, REDCap, CancerGrid, InForm, Datatrak, Medidata Rave, etc [4, 7]. Significant efforts have been made to agree on standards for clinical studies, and the ODM (Operational Data Model) [9] proposed by CDISC[2] has been adopted by several regulating bodies and also EDC software tools. Based on XML, ODM serves as a communication interface of clinical study data, but it lacks a semantically-rich model able to capture the different relationships among the different components of a clinical study, as well as linking with other standard vocabularies. Recent works [12] developed approaches for semantic annotation of ODM XML export files, using extensions to the RDF DataCube vocabulary. Other efforts [13] have also tried to achieve semantic integration of clinical data management systems, by integrating ODM and the HL7 FHIR standard. Up to now, the ODM specifications are regarded as the reference for data interchange for these systems, although they lack several features as explained in Section 3. Even if there were some attempts to provide semantic annotations for ODM [3, 10], there is yet no comprehensive ontology that incorporate the aspects covered in this work.

## 3 Design Principles

The MedRed Ontology design is founded on the representation of a generic clinical study, understood as a collection of data acquisition instruments. In the following we present the design principles behind the ontology, namely the structure of the *core model*, and the fundamental features of *composition*, *derivation*, *provenance*, and *validation*.

*Core Model.* According to the ODM model of CDISC [9], a *Study* has a *metadata version* element in which the different definitions of its sub-elements are contained, i.e. a *Form*, *Item*, and *Item Group* definition. These commonly materialize as instrument, question and section definitions, respectively, in a questionnaire-based instrument. Taking this model as a starting point, the MedRed ontology first separates the metadata versioning aspects out of the core model, as this is a cross-cutting consideration. A MedRed *Study* is indeed composed of one or more *Instruments*, each of which has an ordered sequence of steps, modeled as *Item* elements. Different kinds of *Items* exist, such as *Question*, *Information*, or *Operation* items. Different sub-classes of *Instrument* may exist, such as a questionnaire, or case form, etc. *Items* may be grouped in *Sections*, providing a logical and nestable organization to the items of the instrument. Each *Item* identifies its previous item in the sequence, and they may be subject to conditional activation to allow branching logic in a sequence of steps. For each *Item* a corresponding *Variable* can be specified, which represents the data that will be captured (e.g. via a question or form entry). *Variables* are associated to data types, and constraints can be defined upon them, e.g. allowed values, rules, etc. Moreover, a *Study* can be organized in different *Arms*, or branches that focus on a particular characteristic for comparison or testing purposes (e.g. different arms for testing different drugs in parallel). MedRed also allows defining events that can help representing longitudinal studies, where different instruments are used over longer periods of time (e.g. demographics at the beginning of the study, a first set of instruments after 3 months, another set 2 months later, etc.)

[2] CDISC (Clinical Data Interchange Standards Consortium): `http://cdisc.org`

*Composition.* The ability to compose studies and instruments using other items and elements is crucial for the MedRed metadata model. For instance, it is possible to combine different existing instruments from other studies in a new one. Similarly, it is possible to combine questions and items of several instruments to elaborate a new sequence of input items for an instrument. This should allow the reuse of existing metadata and studies that have already been successfully implemented, preventing from reinventing the wheel. A generic model that was created with the purpose of representing a sequence of scientific activities in a plan is the P-Plan ontology [6]. Introducing the basic concepts of *Plan* and *Step*, it allows nesting and constructing different structures of planned items. For this reason, it was chosen as a basis for structuring items and instruments in MedRed, allowing very flexible composition designs.

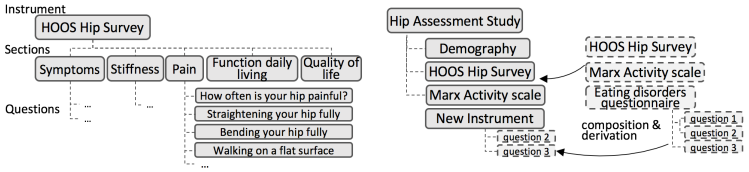


**Fig. 1.** Composition and derivation in MedRed. Left: a sample instrument including its organization in sections and items. Right: a study may incorporate instruments created previously (e.g. Hip survey) or create new instruments reusing items from others (e.g. the eating disorder questionnaire).

*Derivation.* Reusing instruments and items from existing studies also implies that one can be derived from others. One instrument can be amended or extended according to the needs of a different context (e.g. a new study on a different population), by adding new questions or modifying their validation rules, possible values, etc. The representation of this information helps keeping trace of these relationships, as exemplified in Fig 1.

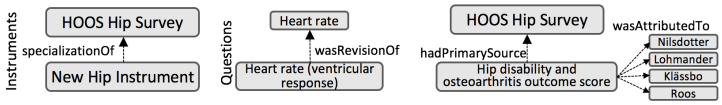


**Fig. 2.** Provenance examples in the MedRed ontology.

*Provenance.* As all studies, instruments, and items can bee seen as traceable resources (or entities according to the PROV model [11]), MedRed allows keeping record of provenance information, including attribution, versioning, authorship, etc. The PROV-O ontology [11] has precisely been defined for this purpose, and as such, we have chosen to align the MedRed core concepts with this model, so that this type of information can be recorded accordingly. For instance, as shown in Figure 2, this allows indicating specialization, revision, source, attribution, and other related information.

*Validation.* In the context of clinical data capture, it is essential to guarantee certain data quality standards, and validation is crucial for defining effective instruments. MedRed opts for reusing existing constraint representation languages in order to incorporate notions of validation into the model. These validation rules should allow flexible definitions, from simple value ranges, to complex pattern matching and combinations of complex rules (e.g. answer to a cholesterol question should be a double value lower than 300 mg/dl.). For this reason, we opted for integrating shape properties, from the SHACL W3C recommendation language [8] for constraints.

## 4 Implementation

Following the design principles stated above, the MedRed ontology was implemented in the OWL language, using the Protégé development environment (19 classes, 12 object and 5 datatype properties). As specified in Section 3, the core model includes the fundamental concepts behind a clinical study: the Study itself, the definition of the Instrument items that compose it, at its inner sub-elements: Section, Item, Operation, as well as other elements as a study Arm and StudyEvent. It has been necessary to cover at least those concepts described in the ODM meta-model to guarantee a minimal compliance with that standard. Furthermore, MedRed goes beyond ODM, as it extends the P-Plan ontology [6] to incorporate nesting and composing of items in a given instrument (Step and MultiStep in P-Plan).

Given that P-Plan extends the PROV-O model, each instrument and item definition is itself a traceable entity, which can be annotated according to the PROV model, including versions, derivative instruments, etc., which are indeed common for studies that evolve with time and that reuse previous instruments. MedRed also aligns to the DDI-RDF vocabulary [2] for describing scientific metadata, as it includes concepts such as Instrument and Questionnaire. Also, for the validation of data acquisition items, MedRed reuses property paths from the SHACL vocabulary [8], which are specifically designed to represent this type of constraints. These dependencies are depicted in Fig. 3.

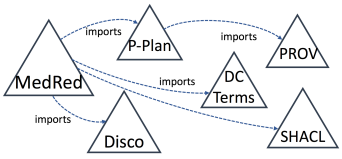


**Fig. 3.** MedRed ontology network: relationship with external vocabularies.

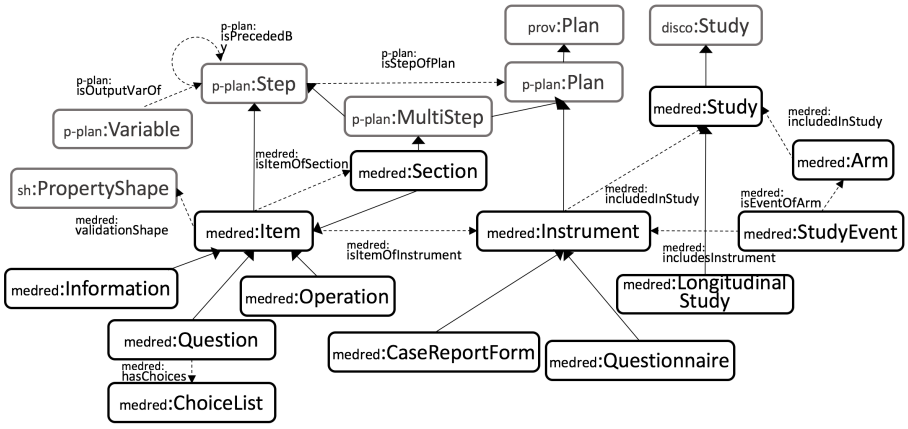


**Fig. 4.** MedRed ontology: an overview of the central concepts.

The central concepts in MedRed (see Fig. 4), as explained above, surround those of an Instrument and Item. Subclasses of these allow for a further specialization of the type of study (e.g. based on questionnaires, entry forms, etc.), or other extensions for more specific uses. The unique identification of each of these items is a fundamental principle for allowing referencing and composing new instruments based on existing ones, therefore meeting the design principles of Section 3. Moreover, the inclusion of the Section concepts allows an unrestricted number of levels and nesting of instrument items, which allows a modular organization of the clinical study.

The salient points of the implementation can be explained through the following examples[3]. The example in Listing 1 shows a 3-month follow-up study definition, including six instruments: one for collecting demographics, another for base line data, 3 monthly questionnaires and a final completion instrument.

```
ex:3MonthFollowUpStudy  a medred:Study ;
  dcterms:description "General health conditions of a patient study ..." ;
  dcterms:identifier  "3MonthFollowUpStudy" ;
  dcterms:title       "3-Month follow up study" ;
  medred:hasInstruments (ex:demographics ex:baseline_data ex:month_1_data ex:month_2_data ex:
       month_3_data ex:completion_data).
```

**Listing 1.** Study following the MedRed ontology.

Each of these instruments can also be fully described, e.g. in terms of their constituent Item elements, as in Listing 2. The instrument is organized in different sections and may include provenance information including authoring, related publications, revisions, etc.

```
ex:expanded_prostate_cancer_index_composite_epic_v2 a  medred:Instrument ;
  medred:items ( ex:epic2200_section1 ex:epic2200_section2 ex:epic2200_section3
                 ex:epic2200_section4 ex:epic2200_section5 ex:epic2200_section6) ;
  dcterms:identifier    "expanded_prostate_cancer_index_composite_epic_2200" ;
  prov:wasAttributedTo  ex:wei, ex:dunn, ex:litwin, ex:sandler;
  prov:generatedAtTime  "2011-07-16T01:52:02Z"^^xsd:dateTime;
  prov:hadPrimarySource ex:epic_article_Urology_56_2000;
  prov:wasRevisionOf    :ex:expanded_prostate_cancer_index_composite_epic_v1;
```

**Listing 2.** An Instrument description using the MedRed ontology

In fact, all components of the study (and instrument) can be annotated with provenance information in order to capture how and when they were defined. In the following examples we omit provenance due to space constraints. In Listing 3 a specific item is described, in this case a question from the previous instrument. The question and its text, the associated variable, and possible display choices, are defined at this point.

```
ex:epic_q48 a medred:Question ;
  medred:isItemofSection ex:section3 ; dcterms:identifier  "epic_q48" ;
  dcterms:title       "14. How often have you had crampy pain in your abdomen, pelvis or rectum?" ;
  medred:choices      ( ex:Morethanonceaday_1 ex:Aboutonceaday_2 ex:Morethanonceaweek_3
                        ex:Aboutonceaweek_4 ex:Rarelyornever_5 ) ;
  pplan:hasOutputVar  ex:epic_q48_var .
```

**Listing 3.** A Question item described with the MedRed ontology

Furthermore, the variable associated to a question (or any Item) can be specified, along with validation rules expressed using SHACL, as in Listing 4. A Cholesterol value is specified, and minimal and maximal values are indicated using a SHACL shape.

```
ex:chol_3 a medred:Question ;
  dcterms:identifier "chol_3" ; dcterms:title "Cholesterol (mg/dL)" ;
  pplan:hasOutputVar ex:chol_3_var ;
  medred:isItemOfSection ex:month_3_datasection1 ;
  medred:validationShape ex:chol_3_shape .
ex:chol_3_shape a sh:PropertyShape ; sh:path medred:dataValue ;
  sh:maxInclusive "300.0" ;  sh:minInclusive "100.0" .
ex:chol_3_var a pplan:Variable ; medred:dataType xsd:double ; medred:varName  "chol_3" .
```

**Listing 4.** Item validation using the MedRed ontology and SHACL.

[3] Prefixes are used as defined in `http://prefix.cc`. medred is used for the MedRed Ontology.

## 5 Exploitation & Discussion

The MedRed Ontology is currently used to represent the metadata of real instruments used in several pilot projects carried out at HES-SO Valais-Wallis, led by the Institute of Health Sciences, and in the scope of the MedRed project. The MedRed project aims at providing an institutional data acquisition platform, mainly targeting clinical data capture. All studies' metadata and their corresponding instruments will be represented in RDF using the ontology, including the entire description of its elements, branching logic, validation, variables, data types, etc. Furthermore, to show the applicability of the ontology to a wider range of clinical data instruments, we have taken a sample of more than thirty instruments from the shared library of REDCap[4], collected by the REDCap project for research purposes from studies all over the world. The full list of instruments used for this experiments can be found in the project source page[5]. A summary, including three of the finished MedRed pilot projects is illustrated in the table of Figure 5. It showcases the heterogeneity of the studies and the features that we covered with the MedRed ontology.

Concerning the availability of the ontology, it has been published through a permanent URI: `http://w3id.org/medred/medred#`, under a CC-BY 4.0 license. The ontology is also referenced through Zenodo, with a DOI assigned to it[6]. The documentation for the ontology has been prepared using the Ontoology [1] framework, and it has also been checked using the OOPS! pitfall scanner service [15]. The latter has only reported minor issues, mainly for the imported ontologies (Oops! report available in the Github repository). The ontology has been made available and discoverable through the Linked Open Vocabularies (LOV) repository , widely used as a reference site for finding vocabularies. Regarding the sustainability of the ontology, it is maintained in an initial phase by the MedRed project. Afterwards, the MedRed platform is expected to function under a business plan similar to that of a Clinical Trial unit, which would consequently guarantee support for the ontology and other related information resources.

| Instrument | Children sleep habits questionnaire | Perceived diabetes self-management Scale | Barriers to diabetes adherence scored | Expanded Prostate Cancer Index Composite | Adolescent Outcomes Questionnaire PR | Agitated Behavior Scale | Apache II Severity of disease Classif. System | Autism Parenting Stress Index | Center for epidemiologic studies depression scale | Charlson Comorbidity Index | Disabilities of the arm, shoulder and Hand | Geriatric depression scale short form | Hamilton Anxiety Scale | Hamilton depression Rating scale | Perceived Stress Scale | Pain Assessment and documentation tool | Multiple sclerosis Quality of Life | Mark activity scale | Interpersonal Reactivity Index | Interpersonal Support Evaluation List | Basic Activities of Daily living scale KATZ | HOOS Hip Survey | Generalized Anxiety disorder 7 item | Subjective Knee evaluation form scored | Fagerstrom Test for cigarette Dependence | FACT-G v4 | Eating disorders examination questionnaire 6.0 | Burn Specific Health scale Brief | BRFSS 2009 module 8: heart attack and stroke | BRFSS 2009 module 13: Adult Asthma history | REDCap Demography | REDCap 3-month follow-up | REDCap Human Cancer Tissue Biobank | REDCap 1-arm longitudinal study | REDCap 2-arm longitudinal study | MedRed pilot: Dynamo | MedRed pilot: WORRK | MedRed pilot: VITA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| topic | sleep habits | diabetes | diabetes | prostate cancer | pediatrics | agitation | disease severity | autism | depression | comorbidity | disabilities | depression | anxiety | depression | stress | pain | multiple sclerosis | activity | social reactivity | social support | activity | hip | anxiety | knee | cigarette | disease severity | eating | burn | heart attack | asthma | demographics | diagnostics | cancer | diagnostics | diagnostics | diabetes | diagnostics | diagnostics |
| # items | 102 | 9 | 24 | 56 | 99 | 20 | 37 | 17 | 21 | 4 | 45 | 16 | 17 | 26 | 14 | 61 | 66 | 4 | 29 | 40 | 6 | 46 | 8 | 24 | 7 | 28 | 33 | 40 | 13 | 15 | 16 | 119 | 198 | 98 | 98 | 23 | 31 | 50 |
| multiple data types | x | x | | x | x | x | x | x | | x | x | | x | x | x | x | x | | | | | x | | x | | | x | | | x | x | x | x | x | x | x | x | x |
| nested items | x | | x | x | x | | | | | | x | | | | x | | x | | | | | x | | x | | x | x | | | | | x | x | x | x | | | |
| validation rules | x | x | | x | x | x | x | x | | x | | | x | x | x | x | | | | | | | | x | | | x | | | x | x | x | x | x | x | x | x | x |
| predefined choices | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| computations | | | x | | | | x | | | | | x | | | | | | | | | | | | x | x | | | | | | x | x | x | x | x | | x | |
| multiple input controls | x | x | | x | x | x | x | x | | x | x | | x | x | x | x | | | | | | | | x | | x | x | | | x | x | x | x | x | x | x | x | x |
| branching logic | | | | | x | | x | | | | x | | | x | | x | | | | | | | | | | x | | | | x | | x | x | x | x | | | |
| matrix grouping | | | | x | x | | | | | | x | | | | x | x | | | | | | | x | x | | | | | | | | x | | x | x | x | | | |
| longitudinal | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | x | x | | | | |
| multiple arms | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | x | | | | |

**Fig. 5.** Summary of the clinical instruments used to showcase the usage of the MedRed ontology.

[4] `https://projectredcap.org/resources/library/`

[5] Instruments used to validate the MedRed ontology: `https://github.com/jpcik/medred`

[6] MedRed Zendoo DOI: `https://doi.org/10.5281/zenodo.819875`

## 6 Conclusion

We presented the MedRed ontology for capturing metadata of clinical studies, following a set of design principles, and extending well-known recommendations. We made it available publicly following best practices and we have shown it fits well for a heterogeneous set of existing instruments. The ontology will be maintained by the MedRed data acquisition project, and in the long term, its growing community.

**Acknowledgements:** MedRed is supported by the Swissuniversities CUS-P2 program.

## References

1. A. Alobaid, D. Garijo, M. Poveda-Villalón, I. S. Pérez, and O. Corcho. Ontoology, a tool for collaborative development of ontologies. In *ICBO*, 2015.
2. T. Bosch, R. Cyganiak, A. Gregory, and J. Wackerow. DDI-RDF discovery vocabulary: A metadata vocabulary for documenting research and survey data. In *LDOW*, 2013.
3. P. Bruland, B. Breil, et al. Interoperability in clinical research: from metadata registries to semantically annotated CDISC ODM. *Stud Health Technol Inform*, 180:564–8, 2012.
4. J. Davies, J. Gibbons, S. Harris, and C. Crichton. The cancergrid experience: metadata-based model-driven engineering for clinical trials. *Science of Computer Prog.*, 89:126–143, 2014.
5. K. El Emam, E. Jonker, M. Sampson, K. Krleža-Jerić, and A. Neisa. The use of electronic data capture tools in clinical trials. *J. medical Internet research*, 11(1):e8, 2009.
6. D. Garijo and Y. Gil. Augmenting PROV with plans in P-PLAN: Scientific processes as linked data. In *Linked Science LISC*, 2012.
7. P. A. Harris, R. Taylor, R. Thielke, J. Payne, et al. Research electronic data capture REDCap—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. of biomedical informatics*, 42(2):377–381, 2009.
8. H. Knublauch and D. Kontokostas. Shapes constraint language (SHACL). *W3C Candidate Recommendation*, page W3C, 2017.
9. W. Kuchinke, J. Aerts, S. Semler, C. Ohmann, et al. CDISC standard-based electronic archiving of clinical trials. *Methods of information in medicine*, 48(5):408–413, 2009.
10. G. B. Laleci, M. Yuksel, and A. Dogac. Providing semantic interoperability between clinical care and clinical research domains. *J. Biomed. and health informatics*, 17(2):356–369, 2013.
11. T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, and J. Zhao. PROV-O: The Prov ontology. *W3C recommendation*, 2013.
12. H. Leroux and L. Lefort. Semantic enrichment of longitudinal clinical study data using the CDISC standards and the semantic statistics vocabularies. *J. biomedical sem*, 6(1):16, 2015.
13. H. Leroux, A. Metke, and M. J. Lawley. ODM on FHIR: Towards achieving semantic interoperability of clinical study data. In *SWAT4LS*, pages 59–68, 2015.
14. A. Nilsdotter et al. Hip disability and osteoarthritis outcome score (hoos)–validity and responsiveness in total hip replacement. *BMC Musculoskeletal Disorders*, 4(1):10, 2003.
15. M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa. Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *IJSWIS*, 10(2):7–34, 2014.
16. J. Shah and et al. Electronic data capture for registries and clinical trials in orthopaedic surgery: open source versus commercial systems. *Clinical Orthopaedics and Related Research®*, 468(10):2664–2671, 2010.
17. I. Sim and et al. The ontology of clinical research (OCRe): an informatics foundation for the science of clinical research. *Journal of biomedical informatics*, 52:78–91, 2014.
18. B. Smith et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251, 2007.
19. T. Souza, R. Kush, and J. P. Evans. Global clinical data interchange standards are here! *Drug discovery today*, 12(3):174–181, 2007.