Ontology of Experimental Variables as an Extension of Infrastructure for Behavioral Research Data FAIRification*

Alena Begler[1[0000-0003-4375-1106]], Grigoriy Anufriev[2[0000-0002-1915-6211]], Irina Leshcheva[1[0000-0003-0132-3605]]

[1] St. Petersburg State University, St. Petersburg 199034, Russia
[2] Higher School of Economics, Moscow 101000, Russia
alena.begler@gmail.com

This is pre-publication version. Please find published version here: https://doi.org/10.1007/978-3-030-98876-0_24.

Paper citation: Begler, A., Anufriev, G., Leshcheva, I. (2022). Ontology of Experimental Variables as an Extension of Infrastructure for Behavioral Research Data FAIRification. In: Garoufallou, E., Ovalle-Perandones, MA., Vlachidis, A. (eds) Metadata and Semantic Research. MTSR 2021. Communications in Computer and Information Science, vol 1537. Springer, Cham. https://doi.org/10.1007/978-3-030-98876-0_24

Abstract. Data sharing is becoming a common practice in behavioral research. Thousands of experimental datasets can be found in open repositories; however, most of them cannot be properly reused due to lack of documentation. We present a structured review of ontologies for experimental research data with a description of 16 ontologies that we divided into three groups according to their approach to variable descriptions: general data description with no attention to variables, scientific research description with either abstract representation of variables or focus on their measurement, and domain-specific ontologies with classes for biological and cognitive fields. The structured resources review can be found at https://doi.org/10.17632/xw288mx2ws.1. We propose an Empirion ontology that provides a variables description that makes it possible to integrate variables from different datasets. To do this, the ontology inherits three-level variable description and enriches it with (1) connections with information about the variable's measurements, and (2) typology of variables based on their role in the experiment. The ontology source code together with supportive materials can be found at our GitHub repository: https://github.com/jimijimiyo/empirion.

Keywords: Ontology, Conceptual Modeling, FAIR Data, Research Data, Experimental Data Integration.

---

* Acknowledgments: The reported study was funded by RFBR, project number 20-07-00854.

1 Introduction

Thousands of social science datasets obtained by different research groups can be found in open repositories; however, the quality of data sharing is considerably low. For example, dataset files are not always accompanied by the metadata and license files necessary for data understanding and reuse. Poor data documentation increases the perceived effort of data reuse (Yoon & Kim, 2017), reduces data consumer satisfaction (Faniel, Frank & Yakel, 2019) and might even prevent the reuse of a dataset that has already been explored (Yoon, 2016). Moreover, it increases the time and effort entailed in already time-consuming data preparation work.

The improvement of data reuse is one of the goals of the FAIR principles (Wilkinson, 2016). A set of data FAIRification initiatives has been successfully implemented; however, an important part of data management is still absent. While most developed models focus on a dataset as a whole, there is a need for variable-level metadata schemas (Dumontier, 2019). Detailed variable description makes it possible not only to interpret a dataset but also to integrate several datasets obtained in similar research and test preliminary hypotheses. And the use of semantic technologies is recommended for the implementation of FAIRification initiatives (Jacobsen et al., 2020).

Our contribution is twofold. Firstly, we present a structured review of ontologies for experimental research data (Section 2). Secondly, we propose Empirion ontology that describes the variables common to behavioral research datasets in a way that allows for their integration (Section 3).

2 Structured review of ontologies for experimental data

To find related ontologies, we performed a structured resources review in accordance with Campos et al. (2020) methodology. We used the terms “ontology” and “vocabulary” and their synonyms to find literature on structured resources and the terms “experimental variable(s)” and “experimental data(set)” with synonyms to indicate domain field. The query string together with the raw materials can be found at https://doi.org/10.17632/xw288mx2ws.1.

Overall, 69 publications were retrieved by Google Scholar to the query string. Of these, 9 publications mentioned structured resources. These publications were checked for the papers that cited them (using Google Scholar), resulting in 88 more papers that were analyzed using the same procedure, which added another 9 relevant papers. The reference lists of the 9 papers from the initial search (a total of 78 publications) were also analyzed and yielded in 14 more relevant papers.

The resulting 32 papers were analyzed for structured resources selection. We walked through the papers and checked all the references and footnotes regarding structured resources (ontologies, thesauri, metadata schemas, and so on). We did not include unstructured resources (like university websites or guidelines). If we encountered an unstructured resource that could contain structured resources (like a data repository that might use some ontology), we searched for such structured resources. Nor did we include meta-models of metadata such as OWL or METS that provide language for metadata schemas. If the webpage in the source paper was inaccessible, we tried to find a new website.

Of the resulting 97 structured resources, 10 addressed experimental datasets or variables description and were available in OWL or RDF. Additionally, 6 more resources were found beyond the structured data analysis and were also included in the resulting list of structured resources. Thus, we found 16 ontological resources that describe experimental variables or datasets. We divided them into three groups according to their scopes.

2.1 Ontologies for general data description

The majority of generic ontologies for data description are not concerned with the description of dataset parts such as variables: IAO defines a dataset as an information entity that describes some part of reality, DCAT denotes that a single dataset can have several forms, cube focuses on dataset structure as a set of observations that have some attributes and dimensions, and the VoID approach appears inapplicable to experimental data representation, as such data are rarely published as RDF datasets.

- **Information Artifact Ontology (IAO)** (Smith & Ceusters, 2015) is a mid-level ontology that extends Basic Formal Ontology (BFO) (Arp, Smith & Spear, 2015) for the description of information entities like *information content entity* and *data set*.
- **Data Catalog Vocabulary (DCAT)** (Albertoni et al., 2020) is a W3C Recommendation for the description of data catalogs. It provides classes for the representation of any digital asset, including *Dataset* and its physical representations (*Distribution*).
- **RDF Data Cube Vocabulary (cube)** (Cyganiak & Reynolds, 2014) is a W3C recommendation for the publication of multi-dimensional data in RDF format. It contains entities for the description of datasets (*DataSet*), which consist of observations (*Observation*), their different parts (e.g. *Slice*) and structures (*DataStructureDefinition*).
- **Vocabulary of Interlinked Datasets (VoID)** (Alexander, Cyganiak, Hausenblas & Zhao, 2011) describes a dataset (*dataset*) as linked data with a particular RDF serialization (*technical feature*).

2.2 Ontologies for scientific research description

The ontologies in this group follow two different approaches to variable description: by their role in the experiment (EXPO and SIO) and by the notion of measurements (OoEVV and disco). The other ontologies in this group (EXACT2, HAScO, ro) do not provide concepts for variables description.

- **An ontology of scientific experiments (EXPO)** (Soldatova & King, 2006) describes scientific experiments. The ontology provides classes for *ScientificExperiment* characteristics (*ExperimentalResults*, *ExperimentalGoal*, *ExperimentalDesign*, etc.) and contains a typology of variables with respect to their role in the experiment:

it distinguishes *TargetVariable*, *ObservableVariable*, *InferableVariable*, *ExperimentalFactor* and *CalculableVariable*. The ontology uses SUMO as upper-level ontology and is designed to be further extended with domain-specific ontologies.
- **DDI-RDF Discovery Vocabulary (disco)** (Bosch, Gregory, Cyganiak & Wackerow, 2013) for the description of survey datasets. It represents datasets at two levels: data content (*LogicalDataSet*, subclass of *dcat:Dataset*), which is organized using *Variables* and physical data (*qb:DataFile*, subclass of *dcat:Distribution*). It provides three-level representation: any variable has its instance in a tabular dataset (*Variable*), its representation with particular values such as a list of codes (*Representation*) and a dataset-independent description of what it is (*RepresentedVariable*). Each variable is also a *skos:Concept*, and *Representation* is either a *skos:ConceptScheme* (for variables represented with codelists) or *rdfs:Datatype* (for variables with numeric values). The variable should be connected to a dataset (*LogicalDataSet*) or *Study*.
- **Experiment Action Ontology (EXACT2)** (Soldatova et al., 2014) is dedicated to biomedical protocols description. Its core classes are *Experimental action*, *Descriptor of experimental action*, *Experimental procedure*, *Experimental protocol*. Thus, it makes it possible to describe an experiment as a process and its properties. It is compliant with upper-level BFO ontology and uses several domain-specific OBO Foundry ontologies as extensions of the corresponding classes.
- **Human-Aware Science Ontology (HAScO)** (Pinheiro et al., 2018) for scientific data annotation, created for HADatAc platform[1] support. It describes the *Data Acquisition* process, such as empirical experiment, which is described with a *Data Acquisition Schema* and conducted using some instruments (*Deployment*). The ontology addresses datasets only through SIO that HAScO uses as upper-level ontology.
- **Ontology of Experimental Variables and Values (OoEVV)** (Burns & Turner, 2013) was created for the description of experimental variables in neuroimaging experimental studies. It includes classes for the *ExperimentalVariable* contained in the *OoevvElementSet* and for the information on how they were measured: *MeasurementScale* with subclasses presenting types of scales and *MeasurementValue* with subclasses presenting the types of values that the variable can take. The ontology is aligned with BFO and IAO and uses CogPO for domain-related concepts.
- **Research Objects (ro)** (Belhajjame et al., 2015) is an ontology for the description of research objects in the research workflow. It is part of a family of ontologies that supports tools for research objects management such as myExperiment portal[2]. The ro ontology extends Annotation Ontology and ORE and represents any *ResearchObject* without further differentiation.
- **Semanticscience Integrated Ontology (SIO)** (Dumontier et al., 2014) is an upper-level ontology that was created for research data integration in the biomedical field. Similarly to the BFO and IAO, it contains a *dataset* as a type of *information content entity* and as a collection of *data items*. It distinguishes *dataset* from *scientific data* and captures information about a dataset's origins as a result of some *purposeful*

[1] http://www.hadatac.org/
[2] https://www.myexperiment.org/

process, for example – an *experiment*. In the SIO, dataset is understood as a *computational entity* while *variable* is understood as a *mathematical entity* that is divided into *control*, *independent* and *dependent variable*.

2.3 Domain-specific ontologies for experimental data description

These ontologies extend BFO and IAO with a description of experimental procedures and, partially – experimental data. They contain classes that can be used for variables description in the biomedical (EFO, OBI, SP) and cognitive (CogPO, cogat) domains, CogPO additionally contains a relatively general typology of experimental variables.

- **Experimental Factor Ontology (EFO)** (Malone et al., 2010) is a huge ontology describing biological samples. It uses BFO as upper ontology; it is compliant with OBI and uses several other domain ontologies as extensions. It was designed as an application ontology for the description of genomics data and provides tools for this.
- **Ontology of Biomedical Investigation (OBI)** (Bandrowski et al., 2016) is a community standard that provides entities for biomedical research description, including the research processes and the entities participating in them. It is part of OBO Foundry (Smith et al., 2007), uses BFO and IAO as upper and mid-level ontologies together with other more domain-specific OBO ontologies.
- **SMART Protocols (SP) Ontology** (Giraldo, García & Corcho, 2014) describes a research process. It uses BFO1 as upper ontology and extends IAO classes with domain-specific subclasses.
- **Cognitive Paradigm Ontology (CogPO)** (Turner & Laird, 2012) was created to describe human behavior experiments. It collaborates with several cognitive science projects[3] and uses BFO as upper ontology and extends *obi:planned process* with *Behavioral Experimental Paradigm* and *Behavioral Experimental Paradigm Condition*. It also contains relatively universal classes for behavioral experiment design such as *Instruction*, *Response*, *Stimulus Modality*.
- **Cognitive Atlas Ontology (cogat)** (Poldrack et al., 2011) is a domain-specific ontology for cognitive science experiments that was created for Cognitive Atlas portal[4] support. It extends some CogPO classes with domain-specific concepts.

2.4 Part of the infrastructure for variables description

The majority of the reviewed ontologies are interconnected and form infrastructure clusters. For example, cogat extends some CogPO classes that in turn extend *obi:planned process*, which is a subclass of *bfo:process*.

[3] http://www.cogpo.org/
[4] https://www.cognitiveatlas.org/

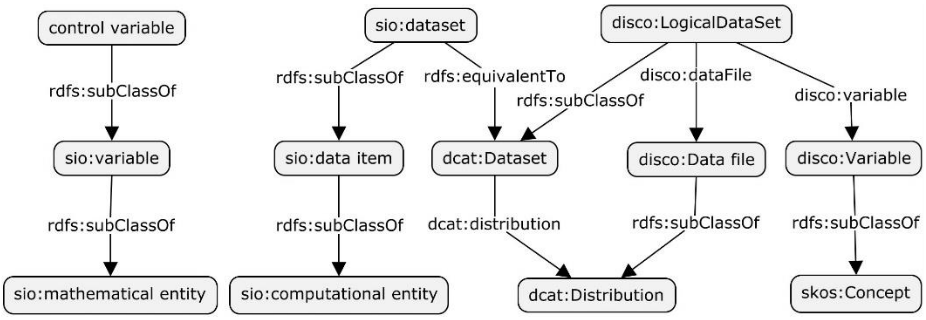


Fig. 1. Part of the infrastructure for experimental variables description.

Some of the ontologies form part of the infrastructure providing a sizable set of variables characteristics (Fig. 1); however, not all of the ontologies that mention variables are connected: while disco and SIO provide connection with DCAT for the description of a variable as part of a dataset, CogPO and EXPO do not describe variables this way, using other upper-level ontologies instead (BFO and SUMO, respectively). The OoEVV sees variables as part of a dataset but isn't connected to other models. Thus, despite the fact that existing ontologies provide a set of useful variables characteristics (different abstraction levels in disco, role in SIO and EXPO, type in CogPO, and measurements in OoEVV), to connect variables from different datasets, these characteristics should be provided simultaneously, and this is not always possible when a bunch of models must be used. The Empirion ontology is aimed at filling this gap.

3 Empirion ontology

During the structured resources review, we looked at the ontological landscape for open research data and identified an issue preventing research data integration: lack of a variables description allowing for the integration of variables from different datasets into a single one. In the Empirion ontology, we combined existing models for variable description and added several connections to provide such an opportunity.

3.1 Development methodology

Ontology development followed Scenario 3 of the NeOn methodology (Suárez-Figueroa, Gómez-Pérez & Fernández-López, 2012): Reusing ontological resources.

The main ontology **requirement** is the ability to provide terminology for the integration of datasets obtained in behavioral experiments. Researchers in this field might be interested in answering questions such as the following: "How do the parameters of a certain type of stimulus influence reaction time?" or "In what context was a particular concept studied?" In other words, there are some hypotheses related to experimental parameters that can be investigated with the re-analysis of existing datasets.

For the **conceptualization** of ontology vocabulary, we applied a combination of top-down and bottom-up modeling styles (Uschold & Grüninger, 1996). We started with an analysis of the datasets we found in open research repositories and derived a list of preliminary concepts. This allowed us to derive groups of variables that included most of the identified terms. Then, we matched these groups with candidate ontologies for **reuse:** the already-existing ontologies that were identified at the structured resource review stage and contained variable descriptions (see Section 2). Finally, to formulate upper-level classes, we rethought metadata classification into embedded, associated and physical (Duval, Hodgins, Autodesk, Sutton & Weibel, 2002).

We **formalized** the Empirion ontology with Ontology Web Language using Protégé ontology editor (Musen, 2015) and **populated** it with the help of a set of Python modules (Leshcheva & Begler, 2020); this stage is ongoing.

3.2 Ontology structure

To describe variables in a way allowing for the integration of variables from different datasets, two aspects of variables representation should be presented in an ontology simultaneously: their connection to the information necessary for their interpretation (i.e. how the variables are represented in a dataset), and their role in the experiment (i.e., what they are from the experiment design point of view). In this section, we briefly describe the main design decisions for Empirion ontology; its full structure can be found at our GitHub repository: https://github.com/jimijimiyo/empirion.

By the structure of their representation, variables can be divided into three types based on their values: (1) values described by the list, for example, correctness of answer as correct and incorrect; (2) values measured in certain units, for example, reaction time in seconds or milliseconds; (3) string values (for open questions) or dimensionless values (for marks). This typology is not directly reflected in the ontology's structure, but it shapes the population process: for the first type of variables, a list of values should be created and linked to reference values; for the second type, a connection to measurement unit should be added; the third type is a simple string.

To provide such information, the Empirion ontology inherits disco's three-level variable description and enriches it by connecting variables to the information necessary for their interpretation and physical representation (Fig. 2):

- *disco:Variable* is dedicated for the variables as they are in the dataset. Individuals of this class are variables as they are in a dataset.
- *disco:Representation* connects variable representation with *MeasurementUnit* or value list (corresponding individuals in the *RepresentationValue* class should be created). It makes it possible to make a comparison of two variables reflecting the same experimental aspect with different instruments. For example, if in the one experiment reaction time was measured in milliseconds, and in the other in seconds, this information will be explicitly presented in the ontology.
- *disco:RepresentedVariable* is a most abstract variable representation that is not depended on variables exact representation in the dataset and is connected with abstract representations of all known variable values: particular cases of the variable values in the particular dataset contained in the *RepresentationValue* class are connected to

the corresponding values of *RepresentedVariableValue* class with *owl:sameAs*, thereby providing an opportunity for the interpretation of values in different datasets.

By their role in the experiment, variables can be dependent (reaction time), independent (stimulus characteristics) or characteristics of something in the experiment (like subject age). This difference is reflected in the Empirion ontology by the *disco:Variable* subclasses structure (part of it can be seen in Fig. 4), including:

- *cogpo:Stimulus* and *cogpo:Response* for presented stimulus (independent variable) and participant's reactions (dependent variable).
- *cogpo:StimulusRole* and *cogpo:ResponseRole* for the role of the variable in the experiment. For example, a stimulus may be a prime – to which any reaction is anticipated, or a target – to which a reaction is supposed.
- *StimulusCharacteristic* and *ResponseCharacteristic* for different modifications of responses and stimuli, respectively.
- *SubjectCharacteristic* for information about the participant.
- *ExperimentPartCharacteristic* for information about different experiment parts: the simplest part of any experiment is the trial (the simplest part of the experiment reflecting a single record comprised of a participant's stimulus and reaction), while some experiments have a more complex design with blocks representing different sets of trials inside the experiment.

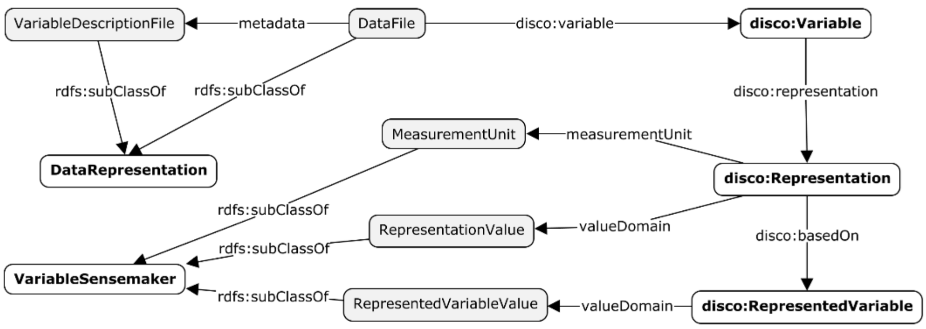
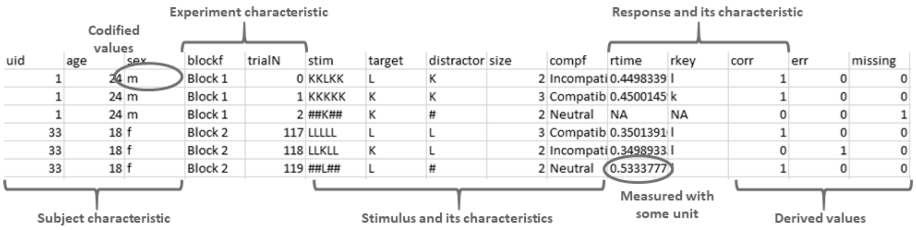


Fig. 2. Empirion ontology main classes and properties. Upper-level classes in white figures.

3.3 Usage scenario

Imagine that we're interested in re-analyzing several datasets obtained in the flanker task experiment (Eriksen & Eriksen, 1974). This task is common for modern experimental studies of perception, as it models the simplest form of cognitive conflict. During the experiment, the participant sits in front of a computer screen and presses a button corresponding to a central sign in a stimulus. The stimuli in the task are rows of objects (for example, letters), while the flanker objects might either be the same as the central objects or different. If all of the objects in a row are the same, the stimulus is called congruent; if they're different – it's called incongruent. For example, the stimulus "< < < < <" is congruent (as the target symbol "<" is the same as the flanker symbols), and the stimulus “< < > < <” is incongruent (as the target symbol “>” is different from the distractor stimulus). The result of the participant’s performance in an experiment is a record of stimuli, responses and their parameters. Such datasets are often presented in tabular (.csv or .xlsx) format and can be found in open repositories. An example of such a dataset is depicted in Fig. 3.

Variables in the exemplar datasets have different types and their values can be either codified or accompanied by measurement units. The variables also differ by type, and they should instantiate different subclasses of *disco:Variable* class. An example of such instantiation of two real datasets is shown in Fig. 4. Both datasets were found with a “flanker task” keyword search at figshare (Dataset 1[5]) and OSF (Dataset 2[6]) research repositories. For the sake of readability, the column order was changed.



uid	age	sex	blockf	trialN	stim	target	distractor	size	compf	rtime	rkey	corr	err	missing
1	24	m	Block 1	0	KKLKK	L	K	2	Incompati	0.4498339	l	1	0	0
1	24	m	Block 1	1	KKKKK	K	K	3	Compatib	0.4500145	k	1	0	0
1	24	m	Block 1	2	##K##	K	#	2	Neutral	NA	NA	0	0	1
33	18	f	Block 2	117	LLLLL	L	L	3	Compatib	0.3501391	l	1	0	0
33	18	f	Block 2	118	LLKLL	K	L	2	Incompati	0.3498933	l	0	1	0
33	18	f	Block 2	119	##L##	L	#	2	Neutral	0.5333777		1	0	0

Fig. 3. Example of a dataset obtained in a cognitive study experiment.

In addition to the depicted mapping, the variables should be connected to the information necessary for their interpretation. For example, the variable *Condition* in Dataset 1 and variable *flanker_type* in Dataset 2 provides the same information about stimulus congruence but with different codifications. To say in the ontology that “incongruent and noncongruent means the same stimulus condition; namely – its noncongruence,” we applied a three-level variables description (Fig. 5): we preserved both codification vocabularies and for every item created an *owl:sameAs* connection with reference vocabulary.

---

5 https://figshare.com/articles/dataset/The_Role_of_Sustained_Attention_in_the_Production_of_Conjoined_Noun_Phrases/1517601

6 https://osf.io/pzxq6/

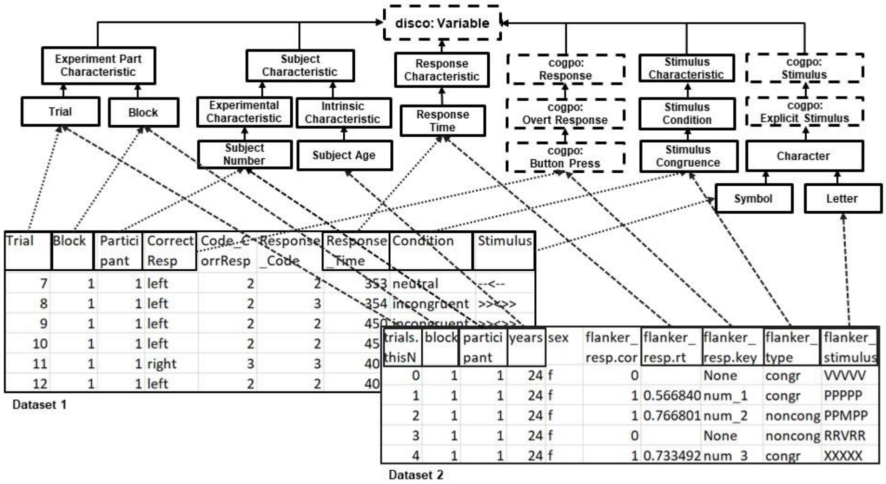


Fig. 4. Mapping of two datasets to the Empirion ontology. Imported classes are in the dashed rectangles. Black arrows represent class-subclass relations. Dashed and dotted arrows represent instantiation.

The proposed approach makes it possible to fully describe a variable from the experimental study dataset because it allows for the preservation of both its meaning in the experiment (with subclasses of *disco:Variable*) and the information necessary for its values interpretation (with three-level representation).

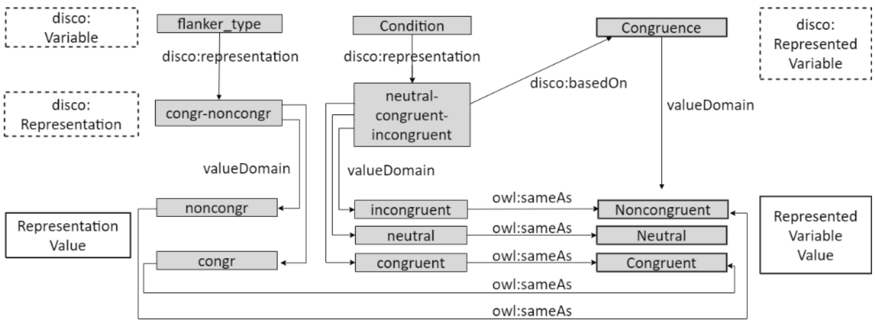


Fig. 5. Representation of variable values in Empirion ontology. Imported classes are in the dashed rectangles. Instances are filled with grey.

4 Discussion and conclusion

Empirion ontology builds on efforts towards experimental variables description for the better reusability of datasets. While existing ontologies are focused either on variable measurements or on the high-level variable-type definition, Empirion connects these approaches and adds some details (see Table 1 for comparison).

Table 1. Comparison of Empirion with existing ontologies

Existing ontology	Variable's description	Empirion difference
EXPO	By role (e.g., target variable)	For now, Empirion does not classify variables by role. It is partially superseded by type classification. Additions on raw and inferred variables differentiation are to be added.
SIO	By role (e.g., control variable)	
disco	By abstraction level (e.g., variable, represented variable)	Empirion borrows this approach with three additions: (1) direct connection with measurements and list values; (2) types of variables added; (3) variables are not *skos:Concept* but individuals of classes for the corresponding types.
OoEVV	By measurement (e.g., measurement scale and value)	In the Empirion ontology, measurements of variables are presented not by scales but by units of measurement and lists of values.
CogPO	By type (e.g., response, stimulus modality)	These types were reused in Empirion and formed an extension of *disco:Variable* class.

For now, Empirion ontology has not been integrated with any upper-level ontology. Despite the fact that many research ontologies use BFO as an upper-level ontology, we decided not to stick to it at this time. Our main concern was that our reuse of some ontologies that are already BFO-IAO mapped might prevent integration. For example, CogPO classifies *Stimulus* as an *object aggregate*. This is, in our opinion, a questionable decision, as *object aggregate* is a subclass of BFO *material entity* (i.e. physical object) and it is not often the case for *Stimulus* as the majority of them are presented on the computer screen.

References

1. Albertoni, R., Browning, D., Cox, S., Gonzalez-Beltran, A., Perego, A. & Winstanley, P. (2020). Data Catalog Vocabulary (DCAT) - Version 2. Retrieved from https://www.w3.org/TR/vocab-dcat-2/
2. Alexander, K., Cyganiak, R., Hausenblas, M. & Zhao, J. (2011). Describing Linked Datasets with the VoID Vocabulary. Retrieved from https://www.w3.org/TR/void/
3. Arp, R., Smith, B. & Spear, A. D. (2015). *Building Ontologies with Basic Formal Ontology*. MIT press.
4. Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., … Zheng, J. (2016). The Ontology for Biomedical Investigations. *PloS One*, *11*(4), e0154556.
5. Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., … Goble, C. (2015). Using a suite of ontologies for preserving workflow-centric research objects. *Journal of Web Semantics*, *32*, 16–42.
6. Bosch, T., Gregory, A., Cyganiak, R. & Wackerow, J. (2013). DDI-RDF discovery vocabulary: A metadata vocabulary for documenting research and survey data. *Proceedings*

of the WWW2013 Workshop on Linked Data on the Web (LDOW2013), 996. Retrieved from http://ceur-ws.org/Vol-996/papers/ldow2013-paper-12.pdf

7. Burns, G. A. P. C. & Turner, J. A. (2013). Modeling Functional Magnetic Resonance Imaging (fMRI) Experimental Variables in the Ontology of Experimental Variables and Values (OoEVV). *NeuroImage*, *82*, 662–670.
8. Campos, P. M. C., Reginato, C. C., Almeida, J. P. A., Barcellos, M. P., de Almeida Falbo, R., Silva Souza, V. E. & Guizzardi, G. (2020). Finding reusable structured resources for the integration of environmental research data. *Environmental Modelling & Software*, *104813*(August), 104813.
9. Cyganiak, R. & Reynolds, D. (2014). The RDF Data Cube Vocabulary. Retrieved from https://www.w3.org/TR/vocab-data-cube/
10. Dumontier, M. (2019). *Accelerating biomedical discovery with an Internet of FAIR data and services*.
11. Dumontier, M., Baker, C. J. O., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., … Hoehndorf, R. (2014). The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. *Journal of Biomedical Semantics*, *5*(1), 1–11.
12. Duval, E., Hodgins, W., Autodesk, S. F., Sutton, S. & Weibel, S. L. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*, *8*(4), 15.
13. Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.
14. Faniel, I. M., Frank, R. D. & Yakel, E. (2019). Context from the data reuser's point of view. *Journal of Documentation*, *75*(6), 1274–1297.
15. Giraldo, O., García, A. & Corcho, O. (2014). SMART protocols: SeMAntic representation for experimental protocols. *Proceedings of the 4th Workshop on Linked Science 2014 - Making Sense Out of Data (LISC2014) Co-Located with the 13th International Semantic Web Conference (ISWC 2014), 1282*, 36–47. CEUR Workshop Proceedings.
16. Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., … Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, *2*(1–2), 10–29.
17. Leshcheva, I. & Begler, A. (2020). A method of semi-automated ontology population from multiple semi-structured data sources. *Journal of Information Science*, 0165551520950243.
18. Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., … Parkinson, H. (2010). Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* , *26*(8), 1112–1118.
19. Musen, M. A. (2015). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, *1*(4), 4–12.
20. Pinheiro, P., Santos, H., Liang, Z., Liu, Y., Rashid, S. M., McGuinness, D. L. & Bax, M. P. (2018). HADatAc: A framework for scientific data integration using ontologies. *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks Co-Located with the 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th to 12th, 2018, 2180*. Retrieved from http://ceur-ws.org/Vol-2180/paper-49.pdf
21. Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., … Bilder, R. M. (2011). The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, *5*(September), 1–11.
22. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., … Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, *25*(11), 1251–1255.

23. Smith, B. & Ceusters, W. (2015). Aboutness: Towards foundations for the information artifact ontology. *Proceedings of the International Conference on Biomedical Ontology*, *1515*, 1–5.
24. Soldatova, L. N. & King, R. D. (2006). An ontology of scientific experiments. *Journal of the Royal Society, Interface / the Royal Society*, *3*(11), 795–803.
25. Soldatova, L. N., Nadis, D., King, R. D., Basu, P. S., Haddi, E., … Rudkin, B. B. (2014). EXACT2: The semantics of biomedical protocols. *BMC Bioinformatics*, *15*(14), S5.
26. Suárez-Figueroa, M. C., Gómez-Pérez, A. & Fernández-López, M. (2012). The NeOn Methodology for Ontology Engineering. In M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta & A. Gangemi (Eds.), *Ontology Engineering in a Networked World* (pp. 9–34). Berlin, Heidelberg: Springer Berlin Heidelberg.
27. Turner, J. A. & Laird, A.R. (2012). The cognitive paradigm ontology: Design and application. *Neuroinformatics*, *10*(1), 57–66.
28. Uschold, M. & Grüninger, M. (1996). Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, *11*(2), 93–136.
29. Wilkinson, M. D. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.
30. Yoon, A. (2016). Red flags in data: Learning from failed data reuse experiences. *Proceedings of the Association for Information Science and Technology*, *53*, 1–6.
31. Yoon, A. & Kim, Y. (2017). Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms and data repositories. *Library & Information Science Research*, *39*(3), 224–233.