Use of Ontologies in Research: An Open Science Solution

Guadalupe Cañas Herrera[1]

s1848151

Science and the public[2]

21 June 2017

Leiden University

[1]canasherrera@lorentz.leidenuniv.nl

[2]Number of Words: 5301

Use of Ontologies in Research: An Open Science Solution

Introduction

Summer of 2015 gave me the opportunity of researching in a real environment for the very first time. At the beginning, I worked at Brown University, within the Astrophysics group. When I came back, I joined my former home, University of Cantabria, concretely the High Energy Physics department. I was hired for an innovation project: "the development of an ontology focused on the Z^0 boson decay into two muons". Probably, for those who are not familiar with laws of particle physics, the previous statement may be confusing. Translated into "everyday language", my task was to design and produce an organized framework of knowledge in the form of a scheme, within the context of Open Science, able to describe the decay process of a fundamental particle into lighter products. Open Science was actively defended by several of my professors, who showed the importance of sharing data results, methods and knowledge. Although I must admit that when I was offered the position I did not understand what the project was about, nor its final aim, I was excited to take part in it.

The first step in my journey was to understand why Open Science was required in research, and, in fact, this discovery was made practically on the first day, when I had to start learning how analysis of the data recorded by the *Compact Muon Solenoid* (CMS), one of the multiple detectors at the *Large Hadron Collider* (LHC), was performed by the group of High Energy Physics at University of Cantabria. I found out that they did not have a homogeneous *workflow of analysis*, and even though CMS data was open access, the extension of the files set difficulties for people less qualified than the current particle physicists. Nevertheless, this was not the end. I came across with similar issues during the whole project, until I finally accepted what the truly problem was: we were preaching Open Science, but we were not fully operating under the Open Science philosophy.

Therefore, after working for almost four months in the design of an Open Science tool, I feel myself confident enough in order to expose what ontologies have to offer for current problems in science, arguing how they may even be suitable solutions. First, I will introduce the reader in the concepts of Open Science and ontologies, explaining their main features. Then, I will show examples of successful ontologies in different fields, comparing their characteristics with other kind of Open Science tools. Finally, I will study the disadvantages this semantic tool may face during its implementation in the future. Along the lines of this essay, and when needed, I will express my experience in applying Open Science statements in real scientific research, recognising that although the implementation of an Open Science philosophy will not be easy, it is extremely essential.

Open Science Framework: what it is and why it is necessary

Numerous publications have surfaced in the mainstream media expressing disappointment with science, even arguing that "Science is broken" (King & Aschwanden, 2016). Belief in science is under threat by bad quality research, the existence of replicability issues across various fields and false published scientific results. This debate also takes part within the scientific community itself, giving as a result the distrust of citizens with respect to science.

In general, it is widely argued by the scientific community and others that the problems of the scientific method in modern practice may be solved by Open Science. Open Science can be understood as a philosophy that defends the transparency in experimental methodology, observation, and collection of data, as well as the public availability and re-usability of scientific data, public accessibility and transparency of scientific communication, everything fomented by web-based tools dedicated to facilitate scientific collaboration (Gezelter, 2009). In conclusion, it is considered to reconceptualise scientific output as being the entire research circle, rather than just results reported in papers. Working under Open Science rules will first require to introduce a general and standard filed guide of how to perform Open Science research cycle, and, secondly, to provide a framework of knowledge where we can implement the Open Science paradigm (Fecher & Friesike, 2016).

The applications of Open Science have demonstrated to solve some of the problems facing the scientific community. For instance, the PYTHON written package ASTROPY focused on providing daily used astronomy programming tools in a user-friendly environment, has improved usability, interoperability, and collaboration among astronomers worldwide by means of simply using open code (Astropy Collaboration et al., 2013). On the other hand, STATCHECK, a public available script written in the open-source software R, has drawn attention to the extensive misuse of p-values whilst also providing researchers and peer-reviewers with a tool to check results and correct mistakes (Epskamp & Nuijten, 2017).

From these examples we can extract how transparency in the reporting of results reduces the chances researchers may have to engage in questionable research practices or scientific misconduct. In fact, (Sijtsma, 2016) cites as the main cause of questionable research practices the lack of transparency in result publication.

Extrapolating from this statement, one could argue that a suitable solution is extending the same level of transparency to the whole research cycle rather than just the results, giving essential feedback in order to reduce inaccuracies.

For instance, instead of thinking in Open Science solutions destined to checking the presence of uncertainties and fraud at the last stages of the research, it would be more efficient to set up tools at earlier stages of the scientific method. For example, a potential Open Science solution is public pre-registration (J.E. & C.A., 2009), whereby the original research plan is registered, allowing for peers to check how the researchers have deviated from their original intentions. This mechanism may also induce certain homogenization of the method within different scientific areas. Engaging in strategies such as public pre-registration requires a non-negligible amount of effort. This is worth it however, as open data also means better data. For instance, it has been found that researchers who are willing to share data related to their publications are associated with better quality evidence. On the other hand, reluctance to share data is associated with weaker evidence and reporting errors of statistical results more frequently (Wicherts, 2016). This is not only true for individual researchers but also for journals; a greater transparency has been associated with higher quality peer-reviews.

Thus, a final argument for promoting more rigorous, higher quality research would be to design an Open Science tool able to intercede in most stages of the ongoing research, not only in earlier nor final phases. Precisely, this kind of Open Science tool would also reduce the current "publish or perish" pressure on researchers, which is another of the key causes of questionable research practices and scientific misconduct, as it will remark the importance of quality over quantity (Sijtsma, 2016). Moreover, it would shift the optimal research strategy of producing novel and underpowered studies (Higginson & Munafò, 2016), which has led to the proliferation of bad science, to a strategy that benefits the entirety of the scientific community. Importantly, the benefits cannot only be found at the community level: individual scientists stand to benefit as well. All other variables being equal, there is evidence that research has a greater impact if the data associated with it has been publicly available (Piwowar, Day, & Fridsma, 2007). This is unsurprising, given that sharing data improves research efficiency dramatically by reducing duplicated efforts and by increasing the opportunities to collaborate in the research process.

Overall, these are powerful arguments supporting the adoption of an Open Science philosophy, in which scientists should ideally work with an Open Science tool suitable to take part in any of the stages of the research production. As far I understand, one of this needed tools may be ontologies.

Ontologies: the needed tool to bond Open Science

According to Tom Gruber (Gruber, 2009), an ontology is a formal specification of a shared conceptualisation. This definition is widely spread, although its meaning may be complex and abstruse for somebody who has just started to learn about this topic. It can be understood as a set of concepts in a domain of knowledge linked by their relations using a common vocabulary. Therefore, an ontology is basically a model that describes a concrete knowledge field and introduces the possibility of organizing and planning science research. This scheme can be understood both by people and machines, easing communication between them, which at the same time allows sharing knowledge.

Learning the concept: technical specifications

Before going any further in analysing what ontologies have to offer, it is essential to understand some concepts in Computer Science related with this tool. Basically, computer resources, which are any physical or virtual components of limited availability within a computer or information management system, are identified through Uniform Resources Identifiers, or URIs, according to the Semantic Web and the World Wide Web Consortium. To organize these URIs, it is used the Resource Description Framework (RDF), that allows also to describe relations between URIs in the form of triples. The triplets follow the structure of

$$\text{Subject} + \text{verb} + \text{complement},$$

where the subject is a URI and the complement may be another URI or information related to the subject as data. In this last case, that data associated to the subject is known as metadata. The organization of these triplets may be applied in a common way in order to share knowledge. In this case, the verb in the triplet structure is obtained from a common language. The most well-known are RDF-s and OWLs languages. Furthermore, OWL introduces the possibility of organizing the triplets through logical closures so that we can explain conditional situations. Therefore, RDF together with RDF-s and OWL, provide the perfect mechanism to design an ontology in a field of knowledge[3].

The parts of an ontology

As it has been already mentioned, an ontology is a formal explicit description of concepts in a domain of discourse in the form of **classes**, **properties** of each class describing various features and attributes of the concept, and restrictions on properties known as **facets**. An ontology together with a set of **individual** instances of classes constitutes a **knowledge base**.

[3]The author of this essay has decided to include technical specifications with the aim of showing how ontologies work, in exactly the same way the author learnt it from the very first time.

For those who may be familiar with Object-Oriented (OO) programming, the concept of ontology is straightforward. Therefore, individuals can be thought to be as instances that inherit their main properties from a fatherly class, which also include another properties that identify only that individual. This analogy with OO programming is explained in Table 1.

	Object-Oriented Programming	**Ontology**
General Part	Class	Class
Individual Part	Object	Individual
Properties	Attributes	Data Properties / Object Properties

Table 1

Visual explanation of the similarities between Object-Oriented Programming and ontology's structure.

Individuals are related with other individuals or metadata through properties that are classified as *object properties*, where an individual from a class is related with the individual of another class, and *data properties*, where an individual from a class is related to a string, or numerical values, which do not come from a fatherly-class (see Figure 1).

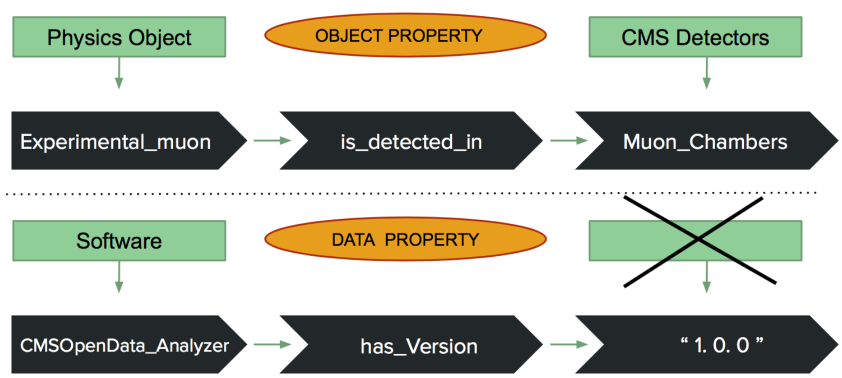


Figure 1. Object (above) and Data (below) properties diagram attached to ontology individuals.

Why ontologies? Current used examples

Most of the time, researchers work with already-defined steps in procedure. Data analysis processes in research methods generally consist of a sequence of time ordered steps over data suitable to be modelled using scientific workflows. Consequently, a scientific workflow is a description of a process for accomplishing a scientific objective. If we develop an ontology able to express the logic between concepts and steps in the process, we will obtain an ideal way of implementing a workflow thanks to a general semantic framework; that is, a collection of common vocabulary and expressions in a knowledge field, for instance, particle physics (see example below). This is precisely what an ontology does: it explains from the beginning (conception of the research project by means of theoretical statements), until the end (publication of the research and setting open access to the data analysis tools and results), the several steps a scientist should pursue (Belhajjame et al., 2014).

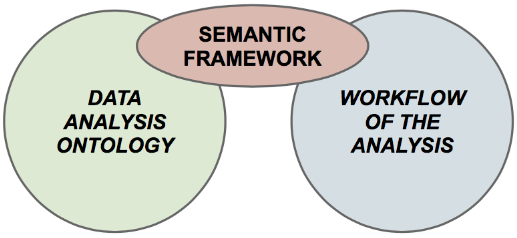


Figure 2. Schematic view of an ideal research process: the scientific workflow procedure should be described by means of a common language (semantic framework) and depicted by designing a collaborative ontology.

A self-made ontology example in High Energy Physics (HEP)

As it was already mentioned during the introduction, I was the responsible designer and creator of an ontology in HEP destined to describe research workflow of scientist collaborating at the CMS experiment (de Lucas, 2016). Hereinafter, the reader may discover, with a real example, how a concrete ontology works.

The design of the CMS Open Data Ontology, made it open available by (Cañas Herrera, 2015), was carried out using a graphical tool called Protégé. This software is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies. It was developed by Stanford University and has become one of the most famous ontologies' application designer (Musen & the Protégé Team, n.d.).

Originally, the ontology was designed to be useful to HEP students at undergraduate and initial graduate stages; however, it evolved to a more ambitious project that expected to comprehend not only the CMS data analysis workflow but the Standard Model itself. Thus, the Ontology was designed from top to bottom, meaning that firstly, the most important or general parts (classes) containing information about High Energy Physics were conceived. Secondly, those classes were fulfilled with extra-information and details. The Ontology structure according to its main classes is presented in Figure 3.

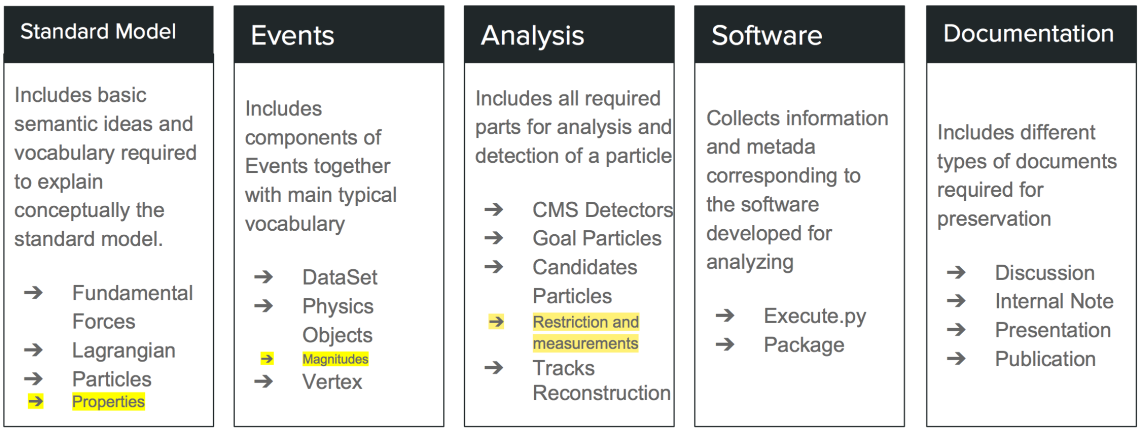


Figure 3. Ontology Class Diagram showing the main classes and their subclasses. **Standard Model**: includes information about the elementary particles, the fundamental forces and the interactions between them. **Events**: describes what the data type and the preservation policy are. **Analysis**: describes the CMS detectors and particles processes. **Software**: provides information about the software used for the analysis. **Documentation**: used to register information not only about the software but the internal notes of the group, as well as required modifications of the workflow process.

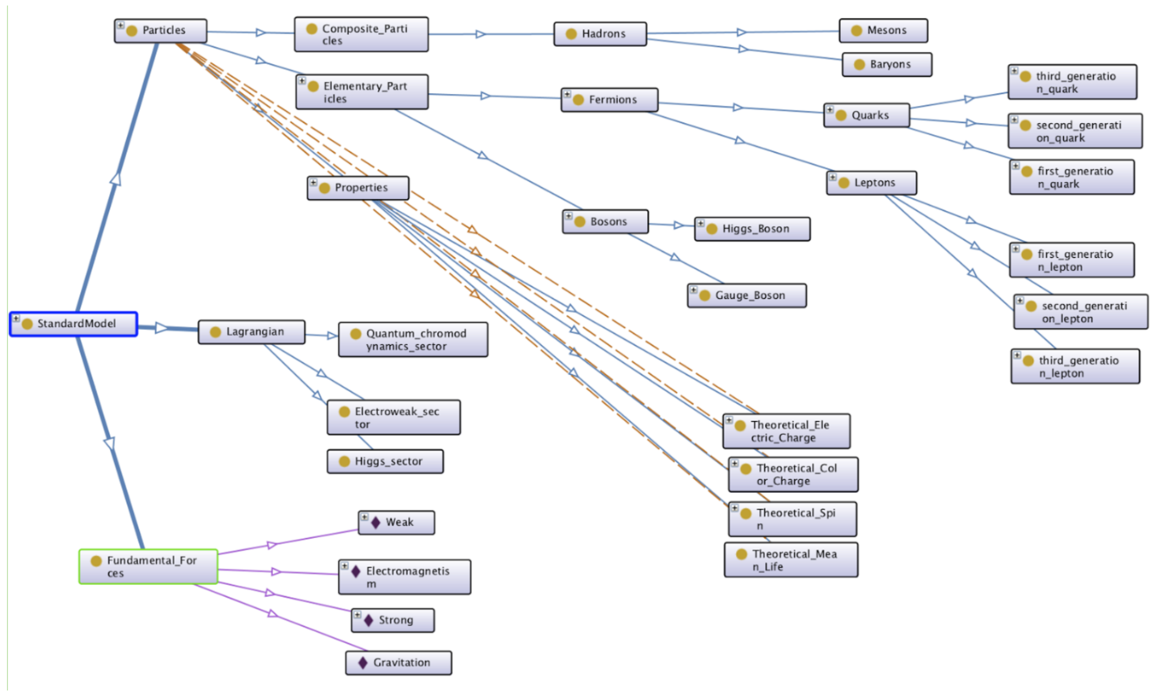
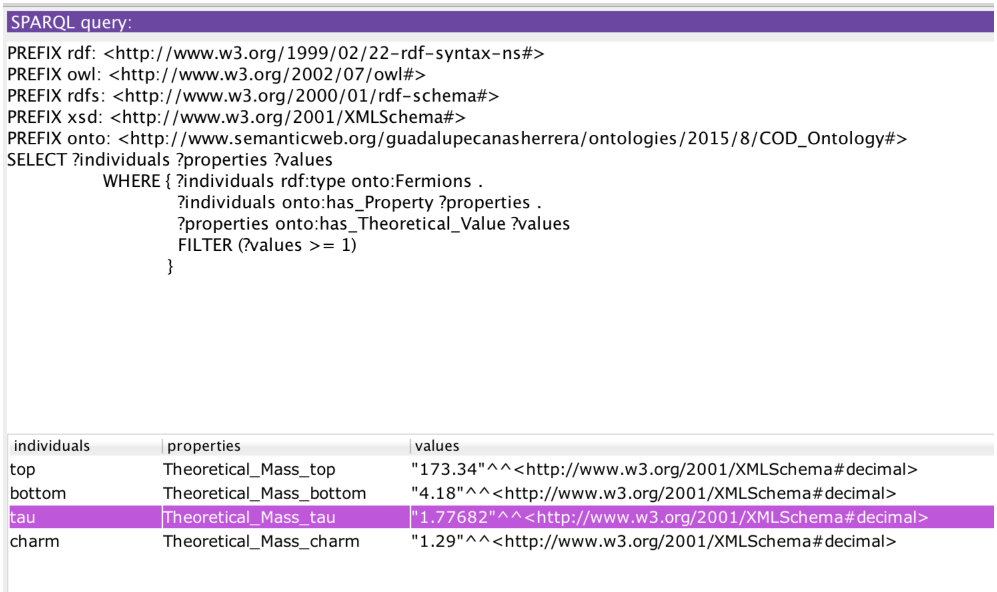


Figure 4. Screen Capture of Protege App at the OntoGraph tab view for the High Energy Physics Ontology's class *Standard Model*. The plot shows the subclasses as well as individuals and metadata.

Without going into further detail regarding the fine structure of the ontology itself, which is far from the scope of this paper, the reader may have already realized that the idea of having a "set of instructions" that records and controls every step of the research is undoubtedly useful, as long as the user knows how to interact with this schema. Ontologies, which are understood not only my humans but also by machines, are suitable to be asked and return complex answers. In this way, we could ask the ontology, for instance, who was the person in charge of writing part of the analysis code and when was the last modification made, which part of the analysis is related with a concrete theoretical statement, whether we are fulfilling the ethical closures and publication rules, or even in which part of the workflow we are at that moment.

Usually, the queries are written in SPARQL, a language designed to ask precisely RDF documents and obtain information from them. In Figure 5, the reader can find an example of a SPARQL query asking the CMS Open Data ontology about the theoretical values of fermions masses.



SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX onto: <http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/8/COD_Ontology#>
SELECT ?individuals ?properties ?values
            WHERE { ?individuals rdf:type onto:Fermions .
                    ?individuals onto:has_Property ?properties .
                    ?properties onto:has_Theoretical_Value ?values
                    FILTER (?values >= 1)
                  }
```

individuals	properties	values
top	Theoretical_Mass_top	"173.34"^^<http://www.w3.org/2001/XMLSchema#decimal>
bottom	Theoretical_Mass_bottom	"4.18"^^<http://www.w3.org/2001/XMLSchema#decimal>
tau	Theoretical_Mass_tau	"1.77682"^^<http://www.w3.org/2001/XMLSchema#decimal>
charm	Theoretical_Mass_charm	"1.29"^^<http://www.w3.org/2001/XMLSchema#decimal>

Figure 5. Example of SPARQL Query filtering possible Fermions' theoretical values.

In autumn 2015, this ontology was introduced to the whole CMS collaboration during the Open Data Initiative Workshop, where the idea was warmly received.

Now, this ontology is being used by University of Cantabria to introduce prospective PhD student to the current analysis of data within the CMS collaboration. Moreover, together with open access tools of Open Data analysis provided by the CMS collaboration itself, the ontology can show undergraduate students (or even senior high school students) how the analysis of particle physics looks like, incrementing not only their knowledge but other skills such as organization of the scientific work and an ethic research philosophy.

Further examples within academia

The use of ontologies is getting widely spread in the scientific community, providing future researchers willing to embrace open science with useful examples for future implementation.

For instance, the Semantic Web for Earth and Environmental Terminology or SWEET are a set of ontologies written in the OWL ontology language publicly available by the Jet Propulsion Laboratory. At the beginning, they were conceived to be prototype for improving the discovery and use of Earth science data, through software understanding of the semantics. Later, they were used with the purpose of organizing research projects about new spaceships behaviour under different climate conditions. Currently, the set of ontologies evolved to include the necessary vocabulary and research framework to guide NASA missions about Global Change (Huang., 2010). Within this domain of knowledge, NASA can work with several partners in the development of common projects in which the scientific workflow will be the same.

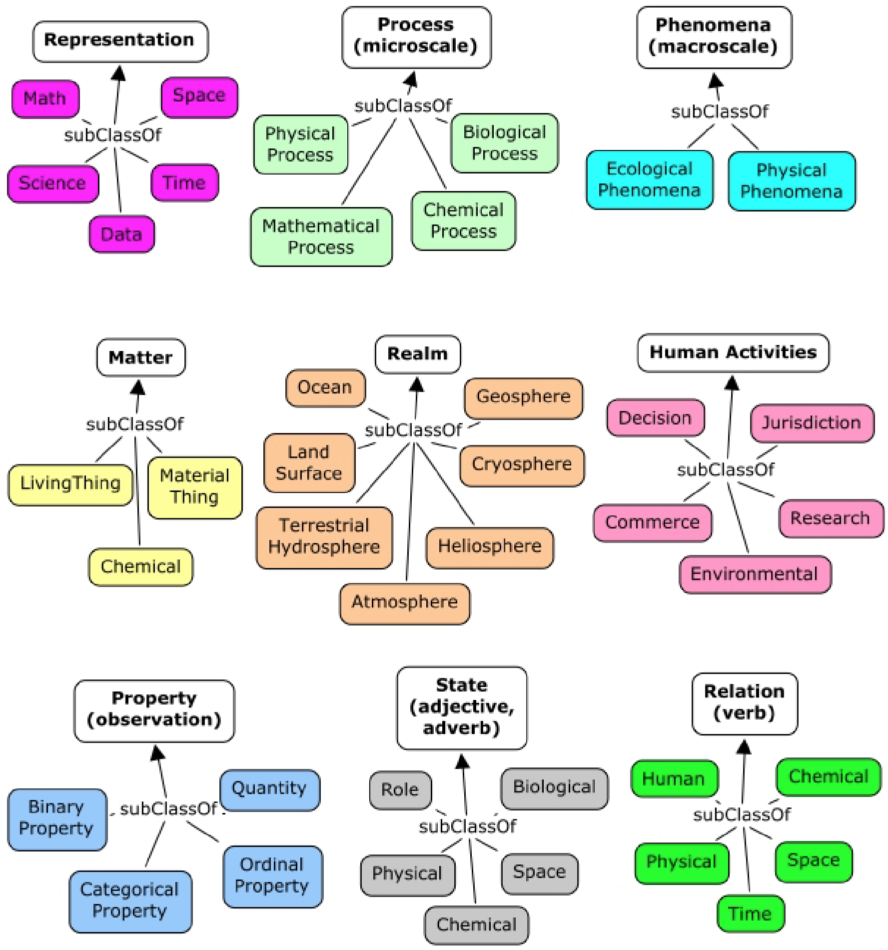


Figure 6. Graphical visualization of some classes and subclasses presented in SWEET ontologies.

Furthermore, ontologies are getting more powerful in natural sciences, concretely in biomedicine, where research ethics need to be delicate applied. These research ethics committees need to face overabundance of information, where bureaucracy together with changing rules and requirements spend up a considerable amount of time. Thus, it seems natural to have an Open Science tool shared by committees around the world in which researchers can share opinions and results. It will provide several benefits such as assistance to those seeking clarity or consensus about how to apply ethical principles in similar situations. Moreover, it may foster the detection of inconsistencies in use of terminologies in different countries by generating standardized forms. These forms will be used, in a compatible way, across investigations, providing a standardized vocabulary for submissions to regulatory agencies, or even improving the collection of data from past studies (Koepsell, Arp, Fostel, & Smith, 2009).

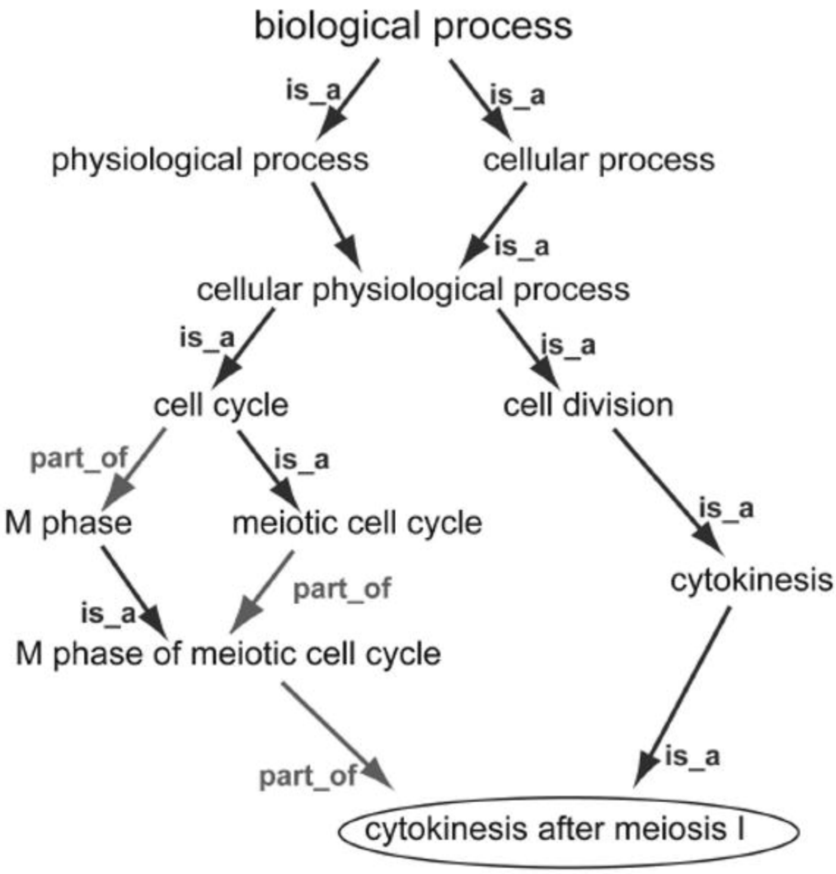


Figure 7. Graphical visualization of part of a Cell Domain Ontology suitable to be included in a biomedical ontology (Koepsell, Arp, Fostel, & Smith, 2009).

Currently, there is a worldwide consensus about the need of an ontology that may offer controlled vocabulary for the ethics of human research. This is due to the fact that biomedical investigations being usually complex processes, frequently involving numerous parties frequently in geographically distinct areas and often taking place over lengthy spans of time. Drugs, devices, and procedures intended to treat or cure disabilities, diseases, or defects must be tested under a variety of conditions. Eventually, human subjects may be used, and there may be significant risks involved in the clinical trial. Because of lapses in scientific and ethical judgements by researchers, human subjects have sometimes suffered unnecessarily. Thus, a standardised tool able to organise and administrate biomedical research together with the required ethics statements that the investigation should fulfil would eventually save time, and consequently, even human lives.

Further examples outside academia

It is worth to mention that ontologies also play an important role far from the scientific environment, such as in communication and baking. For instance, a remarkable example of the use of ontologies in media comes from the British Broadcasting Corporation (BBC). They created a set of ontologies, which are continuously being incremented according to current business requirements, in order give a comprehensive understanding of which content the user likes, why and how. BBC uses ontologies to describe the world around us, content the BBC creates, and the management, storage and sharing of these data within the Linked Data Platform, which also follows the principles of Open Science. Ontologies play an essential role within the BBC's Linked Data policy: they are the scheme-like base to connect content together through those topics BBC publish about (BBC, 2017).

The most interesting part of BBC's ontologies is their open access features, which are embedded within the Open Science philosophy: all ontology versions are provided for members of the public and anyone who wants to get a better understanding of the BBC's resources. Moreover, even though there is already an enormous number of existing semantics framework related with media, BBC creates their vocabulary models. The main reason being that small and controlled models are easier to be changed and adapted in an iterative way: engaging with a large community to clarify the semantics of a widely-adopted model is often not viable when speed is the essence of media communication. In addition, the nuances of each use case might not be appropriately represented in an open vocabulary.

On the other hand, there are ontologies that, still following Open Science principles, should focus in another characteristics not related with speed or viability of adoption. This is the case of the Semantic Bank Compliance ontology tool (Data Models Inc, 2010); it imports, aligns and extends the two dominant reference ontologies: the Financial Industry Business Ontology (FIBO), which is a collaboration between the Enterprise Data Management Council and the Object Management Group, and The Legal Knowledge Interchange Format (LKIF), which is European Union project models legal rules of the kind found in legislation and regulations.

This bank-used ontology defines semantic rules to evaluate federal regulations in finances and economics, later used by Banks or Governments. In this case, the main goal is to help banks to determine regulatory treatment, to compile reports and forms, and to produce capital analysis and reviews of markets. With this purpose, analysts use semantic data structures contained in the ontology to extract information from numerical data.

Preaching Open Science: challenges in implementation

In November 2015, I was cordially invited to expose the results of my ontology design and development in the "CMS Open Data Workshop", aimed to those scientists collaborating within the experiment CMS at LHC to expose their initiatives for a better preservation and sharing of CMS data. During this workshop, several researchers from different universities around the world, as well as third parties such as private companies, mostly showed open data access and analysis tools directed to young students and public. However, just another speaker and I introduced to the audience the need for an ontology to guide CMS data analysis and preserve the data. We both agreed that ontologies were essential for promoting knowledge, developing data reuse and annotation, as well as effectuation of future preservation not only of data but scientific workflows by humans and machines. For instance, one of the obvious outcomes of implementing an ontology that may homogenise High Energy Physics data analysis would be to directly turn out the famous "Particle Physics Data Book" into an ontology itself. As my ontology was still under development and I expected to continue completing it in the incoming months, the coordinator of the workshop suggested that I could compare and cross check my ontology with the other speaker's one. Nevertheless, this speaker refuse to provide with the draft of his ontology, arguing that he preferred not to enter into the "sharing results" practice.

This anecdote manifests the main issue ontology implementation could come across. An ontology is thought to be the skeleton of Open Science statement, the guide you should properly follow in order to research under Open Science directions. Designing an ontology but not sharing the desire of working under Open Science step-lines is a complete non-sense. Therefore, if our final aim is a successful implementation of ontologies in different science fields, researchers should also undoubtedly adopt the Open Science philosophy.

Still, if the reader has carefully gone through the lines above regarding the main features of these tools, a main question must have come up: how can you extract the information that an ontology has to offer if you do not "speak" its language? Evidently, the interaction with ontologies nowadays is tough as it requires the knowledge of non-common complex programming languages. For this reason, lots of effort have been put into the improvement of graphical interfaces that may play the role of instant translators between the user and the ontology (de Lucas, 2016). For this purpose, database management system (DBMS) is currently being built (Vickery, 1997). This is a computer software application that interacts with the user, other applications, and the database itself to capture and analyse data. A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of databases, which would be controlled at the same time by ontologies. Therefore, you can use a DBMS to manage the ontology and extract information from it.

Another problem ontologies still have to overcome is the fact that not all fields of study are simple enough in order to design a self-consistent scheme, for instance non-linear physics systems. Therefore, general implementation of ontologies in science is conditioned to the entanglement of the workflow in certain domains of knowledge. Actually, this is also related with the blurred horizons of uses for an ontology: should it be only used to control the workflow of data analysis? should it contain everything related with the research? should it be included in more advanced and capable ontologies, or should it remain compact? The author of this essay strongly believes that these questions, although fair, must be answered by each research group or experiment, as the common destination, which is the implementation of an Open Science philosophy, is in principle affected by the limits of the ontology itself. Moreover, future feedback in working with ontologies and, in general, under Open Science rules in the future would facilitate giving an answer to these issues which basically appeared because of our current inexperience. As a case in point, from my own experience in designing an ontology, I learnt that it is easy to design the ontology while you are designing an experiment or you are at the earlier stages of a project, as both subjects will be developed at the same time and they can provide feedback to each other.

Furthermore, ethical aspects are a valid concern and the legalistic maze which researchers face concerning the publication of ontologies guiding research projects is a problem that must be resolved. Of particular interest is privacy, not only for researchers but also for commercial enterprises which deal with data. These ethical and legal problems are indeed a limitation, but are far from being an unsolvable problem. Ethics committees have been institutionalised to provide ethical guidance in some aspects of research, such as animal experiments, or the already-mentioned biomedicine sector, which is progressively implementing ontologies. The same approach could be applied for establishing what should be made public.

Finally, others can argue that setting tight instructions that cannot be violated during the research, which is mainly the scope of an ontology, may induce a lack of originality and imagination. They argue that in research you need to get yourself adapted to unexpected events. Yet, an ontology can take care of this type of uncertainties in the exact same way project management tools do so during the analysis of risks and weaknesses of projects. In fact, a final improvement to solve this problem would be the direct link between ontology files and online project management tools such as SMARTSHEET.

What an ontology can make for the general public

University of Cantabria tested in April 2016 my ontology in High Energy Physics with Bachelor students getting into the world of Particle Physics for the very first time. The obtained feedback was highly satisfactory: while in previous semesters the first lectures in data analysis of CMS real events were tedious and complicate to follow by the students, this time they had the possibility to extract proper information and clarify their questions on their own. Some of the feedback we received contained comments such as *"working with an ontology is like if you were studying something new by the first time but the professor directly provide you with good notes and study schemas to learn the content"*.

Seemingly, with this high-end beginning in the use of ontologies, our purpose changed to re-designing the ontology so that it could be used by non-physicists. Rapidly, after some hours of work, we realized that instead of re-designing the ontology what we needed to do was to add more and more information and content. Up to a point, we concluded that if we really wanted to have a user-friendly ontology with the proper graphical environment, we first needed to define a target group (i.e: high school student, people with basic education, elderly population) and complete the ontology by means of trial and error, identifying what the science-educational and informatics gaps were.

Ontologies can also indirectly help with problems between science and the public; in general, Open Science, together with ontologies, may diminish several issues in science that induce the distrust of the public, such as fraud, misinformation or replicability concerns. Ontologies would make malpractices difficult to commit, will reduce inaccuracies in publications and will record necessary information in order to reproduce the experiment in the future.

Indirectly, ontologies can help with problems between science and the public. In general, Open Science, together with ontologies, may diminish several issues in science that induce the distrust of the public, such as fraud, misinformation or replicability concerns. Ontologies would make malpractices difficult to commit, will reduce inaccuracies in publications and will record necessary information in order to reproduce the experiment in the future.

Conclusions

In conclusion, in this essay I have argued that science is facing serious problems due to a biased system which promotes false results at the expense of high quality science.

I believe that by adopting an Open Science philosophy, the scientific establishment can promote a more holistic method of evaluating science, an approach which will lead to better quality output. In order to implement Open Science, we need to set up a guide-line tool able to interact with the researcher in every stage of the research. These Open Science tools necessary to set up this framework of knowledge are Ontologies. Ontologies will set concepts in a domain of knowledge linked by their relations using a common vocabulary. The model will describe a concrete knowledge field by introducing the possibility of organizing and planning science research. The model will need to be understood by both people and machines, easing communication among them.

I have argued how ontologies are currently being implemented in several scientific fields as well as outside academia, where ontologies are used to encourage data sharing and reuse within their communities. Ontologies can be a key-point in the organisation and homogenisation of data and method in scientific work; they can also control the reproducibility of experiments and research data, as they will be the "lab-type" notebook of the whole workflow, and provide the semantic framework for data analyses containing the steps needed to carry out the research. Besides, scientists taking part in the analysis will be able to document, annotate and share their process among the scientific community. Ontologies will also offer better linking policies of scientific journalism, bibliographical resources and control of programming versions. The main obstacle ontologies have to overcome is the fact that, taking into account how much effort is required, they are hard to design and to be implemented. Somehow, they need to be easier to make in order to success in the future.

Moreover, they can be used to improve third parties' understanding of scientific work, not only within the scientific community but to the general public, as long as the target groups are correctly identified and the horizon of the ontology is well fixed. That is, ontologies are essential and unavoidable if we want to create an Open Science philosophy and profit its benefits. Thus, they can help to improve the relationship between science and the public indirectly; nevertheless, ontologies alone will be insufficient to provide with solutions to some public-science issues such as improving the interest in science by lay citizens or to set up a common level in public's science knowledge if the target group is very heterogeneous.

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., … Streicher, O. (2013, October). Astropy: A community Python package for astronomy. *"Astronomy and Astrophysics"*, *558*, A33. doi:10.1051/0004-6361/201322068. arXiv: 1307.6212 [astro-ph.IM]

BBC. (2017, May). *BBC Ontology*. Retrieved from http://www.bbc.co.uk/ontologies

Belhajjame, K., Zhao, J., Garijo, D., Hettne, K., Palma, R., Corcho, Ó., … Goble, C. (2014, January). The Research Object Suite of Ontologies: Sharing and Exchanging Research Data and Methods on the Open Web. *ArXiv e-prints*. arXiv: 1401.4307 [cs.DL]

Cañas Herrera, G. (2015, September). *CMS Open Data Ontology*. Retrieved from https://github.com/gcanasherrera/CmsOpenData_Ontology.git

Data Models Inc, J. (2010, January). *SEMANTIC BANK COMPLIANCE. Ontology Documentation.* Retrieved from http://bankontology.com/ontology-documentation/

de Lucas, J. M. (2016, December). *Open Science?* Retrieved from https://indico.ifca.es/indico/event/432/material/slides/0.pdf

Epskamp, S. & Nuijten, M. B. (2017). *statcheck: Extract statistics from articles and 184 recompute p values.* Retrieved from http://CRAN.R-project.org/package=statcheck

Fecher, B. & Friesike, S. (2016, December). *Open Science: One Term, Five Schools of Thought. 187 Opening Science.* Retrieved from http://book.openingscience.org/basics_%20188%20background/open_science_one_term_five_schools_of_thought.html

Gezelter, D. (2009, July). *What, exactly, is Open Science?* Retrieved from http://openscience.org/what-exactly-is-open-science/

Gruber, T. (2009, December). *Ontology.* Retrieved from http://tomgruber.org/writing/ontology-definition-2007.htm

Higginson, A. D. & Munafò, M. R. (2016, November). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology*, *14* (11), 1–14. doi:10.1371/journal.pbio.2000995

Huang., T. (2010, April). *SWEET*. Retrieved from https://sweet.jpl.nasa.gov

J.E., G. & C.A., C. (2009, July). *The promise of pre-registration in psychological research.* Retrieved from https://www.apa.org/science/about/psa/2015/08/pre-registration.aspx

King, R. & Aschwanden, C. (2016, December). *Science Isn't Broken.* Retrieved from http://fivethirtyeight.com/features/science-isnt-broken/

Koepsell, D., Arp, R., Fostel, J., & Smith, B. (2009). Creating a Controlled Vocabulary for the Ethics of Human Research: Towards a Biomedical Ethics Ontology. *Journal of Empirical Research on Human Research Ethics*, *4* (1), 43–58.

Musen, M. A. & the Protégé Team. (n.d.). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, (1(4)), 4–12. Retrieved from http://doi.org/10.1145/2757001.2757003

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007, March). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, *2* (3), 1–5. doi:10.1371/journal.pone.0000308

Sijtsma, K. (2016). Playing with data—or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, *81* (1), 1–15. doi:10.1007/s11336-015-9446-0

Vickery, B. (1997). Ontologies. *Journal of Information Science*, *23* (4), 277–286.

Wicherts, J. M. (2016, January). Peer review quality and transparency of the peer-review process in open access and subscription journals. *PLOS ONE*, *11* (1), 1–19. doi:10.1371/journal.pone.0147913