Beyond reproduction, experiments want to be understood

Jérôme Euzenat
Jerome.Euzenat@inria.fr
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France
Grenoble, France

ABSTRACT

The content of experiments must be semantically described. This topic has already been largely covered. However, some neglected benefits of such an approach provide more arguments in favour of scientific knowledge graphs. Beyond being searchable through flat metadata, a knowledge graph of experiment descriptions may be able to provide answers to scientific and methodological questions. This includes identifying non experimented conditions or retrieving specific techniques used in experiments. In turn, this is useful for researchers as this information can be used for repurposing experiments, checking claimed results or performing meta-analyses.

CCS CONCEPTS

• **Information systems** → *Web searching and information discovery*; *Specialized information retrieval*; • **Social and professional topics**;

KEYWORDS

e-science, scientific knowledge graphs, semantic experiment description, semantic technologies

Reference: Jérôme Euzenat. 2022. Beyond reproduction, experiments want to be understood. In: *Companion Proceedings of the Web Conference, April 25–29, 2022, Virtual Event, Lyon, France*, pp. 774–778. ACM, New York, NY, USA. DOI: 10.1145/3487553.3524676

1 INTRODUCTION

The movement towards open science promotes, beyond article publication, the availability of research artifacts. The benefits of these efforts are well understood, especially for their contribution to accountability and reproducibility.

We concentrate here on experimental sciences and the need to describe experiments. This covers fields as diverse as experimental psychology, microarray sieving, multi-agent social simulation or ethological studies to name a few. Experiments are diverse and do not all fit in the same mold, even within such fields. It is thus not expected to design a common template, unless very general, hence imprecise.

Valuable efforts support paper publication, data publication and especially experimental data publication. Available open data repositories provide storage, integrity, identification and accessibility. Proper metadata allows to index them through specific aggregated indexes and find them through search engines. This definitely contributes to provide Findable, Accessible, Interoperable and Reusable (FAIR) data. But experiments are not restricted to data.

Here, I do not concentrate on the necessary format or infrastructure for making this happening. These annotations may consist in specific graphs [1] or fragmented nanopublications [6, 12]. Many efforts have been dedicated to these aspects which can be exploited for experiment descriptions. I plead for having experiments described with semantic web technologies, independently from their format. This is not restricted to describing how data is processed, which is already quite well managed, e.g. through electronic notebooks, but also how data is obtained before being analysed [11, 14].

Although this information is usually provided in research papers, it is often left aside in metadata, because the most urging tasks are to safeguard, timestamp, identify and eventually replay the experiments. However, this is a very important issue in the long run.

There are already many individual reasons to describe ones' experiments more formally:

accountability Laboratory logbooks may be used for attribution and anteriority, as well as checking authors accountability to their claims. Automatic search through a well-documented corpus of their electronic records is useful.

reproducibility Research papers are assumed to describe experiments so that they can be repeated. As valuable would be an experiment description from which it is possible to run the experiment again and to compare the results to those which were already found.

intelligibility They can be used in order to understand better how an experiment was performed precisely, what was the type of reagent used or how were some values measured.

The term *individual* has been used above because it concerns one individual's experiments, or the collection of the experiments in one laboratory. This already allows experimenters to quickly search from their experimental results. But too often, these experiments are taken in isolation. Beyond reproduction, reuse, reanalysing and simply being aware of experiment results is useful. This should allow other experimenters to build on each other results more quickly and accurately. So far, this is achieved through reading papers.

Consider the following example: one carries out an extensive experiment to determine if a population (e.g. people, animals, cells, agents) bearing a particular feature (e.g. gene, habit, context, program) is more efficient with respect to some measure (e.g. fighting a disease, reproducing, achieving a task, solving a problem). Having completed the experiment, a paper is submitted to a prestigious journal. But the reviews come back stating that such an experiment has already been performed returning significantly different results. Would have it been better to be able to find this experiment

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

beforehand? Nowadays, this is achieved by screening and reading papers. It is not very easy to be sure that the literature has been thoroughly explored and that the really relevant experiments have been found. Less dramatic, having a formal description of the experimental settings allows to quickly determine what the difference between them are, either to borrow one superior technique or to promote a new one. Finally, a precise description should allow to better understand why the results of so similar experiments are different.

The point of this position paper is not to merely call for describing experiments semantically. It is to try to convince why, beyond documenting resources and contributing reproducibility, beyond putting paper content in a scientific knowledge graph [1], semantically describing experiments themselves can be extremely useful in modern science. Taking advantage of the flexibility, extendibility and semantics of semantic technologies in order to make these a more practical tool should allow, in particular, sophisticated queries, meta-analyses over experiment results, data reuse, experiment repurposing, and more automated science.

In the following, we first present a brief state of the art of current efforts for publishing experiment descriptions. We then provide characteristics of scientific experiment knowledge graphs. Finally we consider the benefits of such scientific knowledge graphs, through querying and more broadly applications interpreting descriptions. We then discuss some ways which should ease their adoption.

2 STATE OF THE ART

Semantic eScience [2] exploits semantic web technologies to provide an interoperable and machine-interpretable infrastructure for scientific enquiry. Within the past years, it has been a continuing source of attention.

2.1 Metadata

Necessary technologies are already in place for data and metadata publication and identification, effectively making it largely Findable, Accessible, Interoperable, Reusable (FAIR) data [26].

On the storage and publication side, popular resources such as Zenodo, FigShare, OSF storage and various dataverses offer long-term storage and identification of data. Stored data is associated with metadata. The used metadata schemes, or the part understood by repositories, are easily converted in a myriad of formats and schemata. However, they are rather poor.

These metadata are pushed and consumed by a variety of indexes such as re3data, DataCite or Google dataset search [21]. The latter offers search among data sets described with schema.org dataset and DCAT.

Along these generalist repositories and search indexes, there exist many disciplinary ones that provide more specialised metadata. There are now indexes of these resources such as FAIRsharing.org, with proper metadata search [23].

Many efforts are already dedicated or related to representing experiments. There exists more or less standalone software such as eLabFTW [15] to computerise laboratory notebooks and to make them more easily accessible. But the reproducibility requirements pushes towards more systematic efforts to describe experiments [13]. Concerning computational experiments, of particular interest is the COMSES Computational Model Library [22] which tracks mostly agent-based models in life and social sciences. It aims at reproducing computational simulations.

However, not all experiments are computational by nature, and experiment descriptions should go beyond data manipulation. A natural way to implement more semantic experiment descriptions is to use semantic web technologies (RDF, OWL, SPARQL).

2.2 Content

Experiment descriptions should not merely contain metadata about experiments, but relations between objects. An experiment may be a process in order to collect data to test a hypothesis. All elements should be described: the process, the hypothesis, the resulting data, the experimental conditions and the way to assess the hypothesis against data.

None of this has to be made in isolation: such descriptions must aggregate many readily available resources. Relations between such elements —which process produced which data, what hypothesis is it supposed to support, in which paper has it been published— can be kept track of through provenance assertions [12, 19]. They must be linked to repositories of authors, publications which are already well covered [5].

Concerning the objects of scientific statements, many resources express these with semantic web technologies. In life and health sciences, the Gene ontology, OBO Foundry, bioportal, and Bio2RDF have been here for long and continuously improved [7].

Methodological aspects of research have not been left aside. The Open Research Knowledge Graph project (ORKG) [1, 18] ambitions to represent all that can be found in a research paper: this is wider than experiments, but experiments have reproducibility requirements which lead their descriptions to go beyond what is available in papers. Some work have proposed the expression of experiments [24], of hypotheses [9] or claims [3]. On the experimental side: the researchobject project has provided a way to describe protocols [14] as well as SMART protocols [11]. Attempts are also made to describe evaluation methods [25].

The state of the art allows for expressing formally broad aspects of scientific knowledge; not only the results of scientific enquiry but the whole process that led to establish results. Of course, this work should be extended, but it is already a solid basis.

3 BEYOND METADATA CATALOGUES: SCIENTIFIC KNOWLEDGE GRAPHS

There are four actors in the dissemination of such experiments descriptions:

experimenters who provide experiment description (metadata) and experimental data.
repositories which provide storage and identification services.
catalogues (or indexes) which provide description aggregation and search.
users who will use data described in these catalogues.

All these actors are providing a necessary service, but users have a very limited opportunity to access relevant information for fulfilling their needs. They may be scientists as well as evaluators, integrity investigators, etc. They may have very different needs (dataset look up, similar experiments, similar hypotheses) and goals (reproducibility, repurposability, enquiry).

Hence, we need *scientific knowledge graphs* in which to search, navigate, query and manipulate experiment descriptions. These may be specific to one laboratory, to a field or very general. These scientific knowledge graphs should offer search and query features to many aggregated experiment descriptions. Instead of full-text or faceted search capabilities, they should offer full query evaluation with adequate ontology interpretation.

A scientific knowledge graph uses semantic (web) technologies to provide access to many, interrelated, experiment descriptions. The benefit of semantic technologies are worth recalling in this context:

versatility Any aspect of experiments should be expressible whatever the domain it is about. The variety of available vocabularies and ontologies is useful for that purpose.

openness It should always be possible to extend the vocabulary or express unexpected information without being restricted to a particular format.

expressiveness It should be possible to use formal ontologies and exploit these ontologies to acurately answer queries.

readiness It should be supported by actual software for expressing, querying and manipulating experiments.

For a single hypothesis, there may be different experiments designed to test it. An experiment design may (actually should) have been processed several times. And the results of an individual experiment may be analysed several times, with different techniques. Each of these actions has precursors (the design is a precursor of the experiment) and may have different performers, creation date, etc.

This calls for a modular description of experiments alike the FRBR model [17]: distinguishing the design of an experiment, its actual performance (and eventually associated results) and the analysis of its outcome. The fact that all runs share features of their implementations and all implementations are testing the same hypothesis should lead to share information across experiments.

Scientists could find an intellectual satisfaction of properly describing experiments. But practically, such scientific knowledge graphs should bring them important benefits in querying and understanding experiments.

4 QUERYING

Scientific knowledge expressed with semantic web technologies would clearly facilitate searching results [16]. Hence we first quickly present obvious types of queries that a scientific knowledge graph should be able to answer. These are classified in three categories.

4.1 Search

Search corresponds to what could be extracted from a tabular data base using a single table. Although it is very useful, full-text search is left aside. It can answer queries like:

- "Which experiments use a specific dataset?",
- "Which experiments have been performed but not analysed?",
- "Which experiments simulate (type) population evolution (keyword) of *E. coli* (organism)?",
- "Which experiments have been invalidated?".

4.2 Advanced search

Advanced search aims at making a difference between what can be obtained from a single table containing one experiment per line (simple search) and what can be obtained from a data graph. In particular, it takes advantage of the relations between well-identified objects, such as:

- "Which experiment designs test a particular hypothesis?",
- "Which experiment designs are derived from another specific design?",
- "Which experiments repeat or reproduce a specific experiment?",
- "Which experiments have a protocol using tumor tissue (product) as a sample?" [11],
- "Which hypotheses have not been confirmed since a precise software release?".

4.3 Scientific search

Scientific search is search informed by science. It could take advantage of knowledge about the protocol or knowledge about the objects.

- "Which experiments of a design have yield different outcomes (accepting/rejecting hypothesis)?",
- "Which conditions differ between two experiments implementations? between two designs?",
- "Which experiment designs test the growth (modality) of a specific measure?",
- "Which experiments in mammals (general category) have reported cumulative cultural evolution (general effect)?",
- "Which experiments provide measures to control for equal motivation (general condition)?".

All these queries have been designed to report experiments, but the other variables of the queries are equally useful.

Some benefits of these queries are not brought by what these repository were intended to (reporting the conditions under which an experiment has been performed), but serendipitously finding secondary, but recorded elements.

5 BEYOND QUERYING, UNDERSTANDING

Querying descriptions is the first natural benefit of providing experiment descriptions. Moreover, they should provide computers with a better grasp at what is described. Not pretending that they actually understand what is described, this give them leverage to go beyond querying. We provide below examples of what is possible.

Beyond reproducibility, which already requires a precise description of performed experiments, machine-interpretable descriptions would allow easier experiment adaptation, modification, comparison, or checking.

5.1 Consistency checking

If experiments are described in a meaningful way, it should be possible to test beforehand, or statically, that they make sense. Classical tests are that every process element is provided with proper input, can be processed (e.g. compatible software configurations) and will generate expected output. Such tests can be performed at the design level and at the experiment level.

Tools may be developed to help (debugging and) peer-reviewing experiment design. They may also help to identify missing information in descriptions and so doing facilitate off-line reproducibility.

This would be a very important support for experimenters, as well as pre-registration reviewers [4]. Going further, this could help checking beforehand that the setting indeed tests the hypothesis.

5.2 Repeating, reproducing

Checking the consistency of descriptions statically should be very important independently from any actual attempt at reproducing experiments. However, static analysis can test that a description make sense, not what sense it makes. The latter is achieved through performing the experiment.

This is, in principle, far easier for computational experiments. Providing accurate descriptions of computer-based experiments can allow a computer to reproduce them. For that purpose, the semantic descriptions must also be either operational or translatable to an executable format.

This is the goal of adopting executable workflow languages for describing protocols. They provide both executability and better intelligibility. However, it should be paid attention that executable descriptions indeed correspond to what is intended.

This would contribute to one of the main goals of FAIR and open research: improving reproducibility.

5.3 Result checking and data reuse

Once an experiment has delivered data, they have to be analysed. Such an analysis may be part of the experiment protocol so that experiment processing can go as far as providing the outcome: tell if the experiment supports or rejects the hypothesis and why.

However, it should also gain at being described separately, as this would allow to replace one implementation of a standard data processing procedures by another making it less dependent on data analysis packages.

In particular, it may be useful to apply, post-hoc, different analyses to the same data. But it may also be useful to apply the same analysis on different (or updated) data. For instance, data analysis workflows can be exploited for continuous reevaluation of hypotheses by updating data analyses when new data is available [10].

5.4 Meta-analysis and comparison

When experiment results are properly registered with individual experiment descriptions, it should be easier to take advantage of query capabilities to perform meta-analyses against knowledge graphs. Comparing experiments and their results may allow to:

- identify experiments that may be considered the same (reproduction),
- identify experiments that can be considered variations and tell how they differ (comparison),
- identify experiments whose results contradict one another,
- suggest conditions that impact some results (this may need further data processing), or
- predict averse effects from results [20].

This may help organising experiment descriptions (provide an additional overlay on them) or extracting important information for meta-analysis papers. ORKG already goes in this direction by providing comparison services informed by the research activity [18].

5.5 Repurposing

Repurposability is the ability to quickly change parameters of one experiment and to set-up a new experiment by assembling pieces of existing experiments.

Repurposing, both from the initial experimenter and others, is one important benefit from experiment descriptions. This is specially the case when the description allows to actually run the experiment. In such a case, modifying the initial design or experiment description leads to a new, already described and ready to process, experiment.

It may also be used to properly reproduce an experiment using different techniques.

Here again, automatic tooling may be used to check variation validity or to suggest variations.

5.6 Automated science: the grand thing

The longer term benefits of such scientific experiment knowledge graphs and better machine understanding is more automated science. Experiment descriptions could lead to formal scientific collaboratories [8] in which scientists exchange, modify and discuss experiments. Machine learning could help cleaning and mining result as well as suggesting interesting research questions.

6 REQUIREMENTS

Ingredients for scientific knowledge graphs are available. What is missing is that scientific practices embrace semantic web technologies. Requiring scientists to provide immediate work for delayed benefits is not easy. Although, they are already used to different degrees of formality imposed by their communities. This is visible in notebook holding and additional requirements for publishing papers.

A large part of experiment descriptions are still encountered in published papers (with increasing formality). Thus, many efforts are devoted to extract these from texts [1]. Reducing the costs of providing extensive experiment descriptions can come from more integration in work habits and new applications providing benefits. Here are some suggestions:

automation can be provided by various software used for building or recording the experiments. For instance, eLabFTW [15] is able to connect to laboratory supply databases: identifying exactly what product 'item' was used in an experiment. This contributes to accountability, but can also help providing rich data. Extracting this data already facilitate providing it to publication platforms such as Zenodo.

incrementality The main point is to be able to provide as much data as one wants, not to force providing a complete description. The openness and extensibility of semantic technologies allow to overcome incomplete metadata and to make the best out of it.

reuse by referring to other components (design, authors, papers, genes) it is easier to describe experiments. An FRBR-inspired modular design would also facilitate component reuse.

7 CONCLUSION

I simply claimed that precisely describing experiments semantically and connecting them in a scientific knowledge graph have many benefits. Illustrations of these benefits have been provided. I consider that the largest benefits of the semantic description of experiments are for the experimenters themselves in their actual experimental activity.

Infrastructure for storing and retrieving experiment data is available. Semantic web technologies and ontologies describing research objects and processes have been proposed. Describing experiments formally has attracted attention already. Hence, this paper is an attempt at drawing more attention to experiments.

ACKNOWLEDGMENTS

This work has been partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003).

REFERENCES

[1] Sören Auer, Viktor Kovtun, Manuel Prinz, Anna Kasprzik, Markus Stocker, and Maria-Esther Vidal. 2018. Towards a Knowledge Graph for Science. In *Proc. 8th International Conference on Web Intelligence, Mining and Semantics (WIMS), Novi Sad (RS)*. ACM press, 1:1–1:6. https://doi.org/10.1145/3227609.3227689

[2] Boyan Brodaric and Mark Gahegan. 2010. Ontology use for semantic e-Science. *Semantic Web* 1, 1-2 (2010), 149–153. https://doi.org/10.3233/SW-2010-0021

[3] Cristina-Iulia Bucur, Tobias Kuhn, Davide Ceolin, and Jacco van Ossenbruggen. 2021. Expressing High-Level Scientific Claims with Formal Semantics. In *Proc. 11th International Conference on Knowledge Capture (K-Cap), Online*. ACM press, 233–240. https://doi.org/10.1145/3460210.3493561

[4] Christopher Chambers, Birte Forstmann, and Andrew Pruszynski. 2018. Science in flux: Registered reports and beyond at the *European Journal of Neuroscience*. *European Journal of Neuroscience* 49, 1 (2018), 4–5. https://doi.org/10.1111/ejn.14319

[5] Paolo Ciccarese, David Shotton, Silvio Peroni, and Tim Clark. 2014. CiTO + SWAN: The web semantics of bibliographic records, citations, evidence and discourse relationships. *Semantic Web* 5, 4 (2014), 295–311. https://doi.org/10.3233/SW-130098

[6] Tim Clark, Paolo Ciccarese, and Carole Goble. 2014. Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. *Journal of biomedical semantics* 5, 28 (2014). https://doi.org/10.1186/2041-1480-5-28

[7] Michel Dumontier. 2010. Building an effective Semantic Web for health care and the life sciences. *Semantic Web* 1, 1-2 (2010), 131–135. https://doi.org/10.3233/SW-2010-0018

[8] Jérôme Euzenat. 1995. Building consensual knowledge bases: context and architecture. In *Towards very large knowledge bases*, Nicolaas Mars (Ed.). IOS press, Amsterdam (NL), 143–155.

[9] Daniel Garijo, Yolanda Gil, and Varun Ratnakar. 2017. The DISK Hypothesis Ontology: Capturing Hypothesis Evolution for Automated Discovery. In *Proc. Workshops and Tutorials of the 9th International Conference on Knowledge Capture (K-CAP), Austin (TX US) (CEUR Workshop Proceedings, Vol. 2065)*, Ilaria Tiddi, Giuseppe Rizzo, and Óscar Corcho (Eds.). CEUR-WS.org, 40–46. http://ceur-ws.org/Vol-2065/paper08.pdf

[10] Yolanda Gil, Daniel Garijo, Varun Ratnakar, Rajiv Mayani, Ravali Adusumilli, Hunter Boyce, Arunima Srivastava, and Parag Mallick. 2017. Towards Continuous Scientific Data Analysis and Hypothesis Evolution. In *Proc. 31st Conference on Artificial Intelligence (AAAI), San Francisco (CA US)*. AAAI Press, 4406–4414.

[11] Olga Giraldo, Alexander Garcia, Federico Lopez, and Oscar Corcho. 2017. Using semantics for representing experimental protocols. *Journal of Biomedical Semantics* 8 (2017), 52. https://doi.org/10.1186/s13326-017-0160-y

[12] Paul Groth, Andrew Gibson, and Jan Velterop. 2010. The anatomy of a nanopublication. *Information services & use* 30, 1-2 (2010), 51–56.

[13] Yusuf Hannun. 2021. Build a registry of results that students can replicate. *Nature* 600 (2021), 751. https://doi.org/10.1038/d41586-021-03707-9

[14] Kristina Hettne, Harish Dharuri, Jun Zhao, Katherine Wolstencroft, Khalid Belhajjame, Stian Soiland-Reyes, Eleni Mina, Mark Thompson, Don Cruickshank, Lourdes Verdes Montene gro, Julian Garrido, David de Roure, Oscar Corcho, Graham Klyne, Reinout van Schouwen, Peter 't Hoen, Sean Bechhofer, Carole Goble, and Marco Roos. 2014. Structuring research methods and data with the research object model: genomics workflows as a case study. *Journal of biomedical semantics* 5, 1 (2014), 41. https://doi.org/10.1186/2041-1480-5-41

[15] Michael Hewera, Daniel Hänggi, Björn Gerlach, and Ulf Dietrich Kahlert. 2021. eLabFTW as an Open Science tool to improve the quality and translation of preclinical research. *F1000Research* 10 (2021), 292. https://doi.org/10.12688/f1000research.52157.3

[16] Wei Hu, Honglei Qiu, JiaCheng Huang, and Michel Dumontier. 2017. BioSearch: a semantic search engine for Bio2RDF. *Database* 2017 (2017), bax059. https://doi.org/10.1093/database/bax059

[17] IFLA Study Group on the Functional Requirements for Bibliographic Records. 2009. *Functional Requirements for Bibliographic Records: Final Report*. Report 19. IFLA. https://www.w3.org/TR/prov-o/

[18] Mohamad Yaser Jaradeh, Allard Oelen, Kheir Eddine Farfar, Manuel Prinz, Jennifer D'Souza, Gábor Kismihók, Markus Stocker, and Sören Auer. 2019. Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In *Proc. 10th International Conference on Knowledge Capture (K-Cap), Marina Del Rey (US)*. 243–246. https://doi.org/10.1145/3360901.3364435

[19] Timothy Lebo, Satya Sahoo, and Deborah McGuinness (eds.). 2013. *PROV-O: The PROV Ontology*. Recommendation. W3C. https://www.w3.org/TR/prov-o/

[20] Adeeb Noor, Abdullah Assiri, Serkan Ayvaz, Connor Clark, and Michel Dumontier. 2017. Drug-drug interaction discovery and demystification using Semantic Web technologies. *Journal of the American Medical Informatics Association* 24, 3 (2017), 556–564. https://doi.org/10.1093/jamia/ocw128

[21] Natasha Noy. 2018. Making it easier to discover datasets. https://www.blog.google/products/search/making-it-easier-discover-datasets/

[22] Nathan Rollins, Michael Barton, Sean Bergin, Marco Janssen, and Allen Lee. 2014. A Computational Model Library for publishing model documentation and code. *Environmental modelling and software* 61, 4 (2014), 59–64. https://doi.org/10.1016/j.envsoft.2014.06.022

[23] Susanna-Assunta Sansone, Peter McQuilton, Philippe Rocca-Serra, Alejandra Gonzalez-Beltran, Massimiliano Izzo, Allyson Lister, and Milo Thurston. 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nature biotechnology* 37, 4 (2019), 358–367. https://doi.org/10.1038/s41587-019-0080-8

[24] Larisa Soldatova and Ross King. 2006. An ontology of scientific experiments. *Journal of the Royal Society Interface* 3, 11 (2006), 795–803. https://doi.org/10.1098/rsif.2006.0134

[25] Markus Stocker, Manuel Prinz, Fatemeh Rostami, and Tibor Kempf. 2018. Towards research infrastructures that curate scientific information: A use case in life sciences. In *Proc. 13th International Conference on Data Integration in the Life Sciences (DILS), Hannover (DE) (Lecture Notes in Computer Science, Vol. 11371)*. Springer, 61–74. https://doi.org/10.1007/978-3-030-06016-9_6

[26] Mark Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip Bourne, Jildau Bouwman, Anthony Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair Gray, Paul Groth, Carole Goble, Jeffrey Grethe, Jaap Heringa, Peter 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott Lusher, Maryann Martone, Albert Mons, Abel Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (2016), 18. https://doi.org/10.1038/sdata.2016.18