

Reuse of Design Pattern Measurements for Health Data

Núria Queralt-Rosinach¹, Mark Wilkinson², Rajaram Kaliyaperumal¹,
César H. Bernabé¹, Qinqin Long¹, Michel Dumontier³, Paul N. Schofield⁴ and
Marco Roos¹

¹Leiden University Medical Center, Eindhovenweg 20, 2333 ZC Leiden, The Netherlands

²Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

³Institute of Data Science, Paul-Henri Spaaklaan 1, Maastricht University, Maastricht 6229EN, The Netherlands

⁴University of Cambridge, Downing Street, Cambridge CB2 3DY, United Kingdom

Abstract

Research using health data is challenged by its heterogeneous nature, description and storage. The COVID-19 outbreak made clear that rapid analysis of observations such as clinical measurements across a large number of healthcare providers can have enormous health benefits. This has brought into focus the need for a common model of quantitative health data that enables data exchange and federated computational analysis. The application of ontologies, Semantic Web technologies and the FAIR principles is an approach used by different life science research projects, such as the European Joint Programme on Rare Diseases, to make data and metadata machine readable and thereby reduce the barriers for data sharing and analytics and harness health data for discovery. Here, we show the reuse of a pattern for measurements to model diverse health data, to demonstrate and raise visibility of the usefulness of this pattern for biomedical research.

Keywords

Health data, Design pattern, Ontology, FAIR

1. Motivation

To enable informed healthcare decisions, hospitalised patients are characterised by different health data such as travel history, comorbidities, and medications, and are monitored by clinical measurements. Observational measurements provide insights into disease which range from diagnosis and prognosis for individual patients to epidemiological understanding of the disease in a population. The COVID-19 outbreak made clear that rapid analysis of observations across a large number of healthcare providers can have enormous health benefits. This has brought into focus the need for a common model of quantitative health data that enables data exchange and federated computational analysis.


During the last virtual BioHackathon 2020 COVID-19, we created a minimal formal model for COVID clinical observations using Semantic Web standards for quantitative traits, based on quantitative information in the COVID-19 WHO RAPID Case Report Form. The model

FOIS 2021 Demonstrations, held at FOIS 2021 - 12th International Conference on Formal Ontology in Information Systems, September 13-17, 2021, Bolzano, Italy

✉ n.queralt1_rosinach@lumc.nl (N. Queralt-Rosinach)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

describes clinical measurements to express quantities, their units, and the assay to obtain the measurement [1]. The application of ontologies, Semantic Web technologies [2] and the FAIR principles [3] is an approach used by different life science research projects, such as the European Joint Programme on Rare Diseases (EJP RD)¹, to make data and metadata machine readable and thereby reduce the barriers for data sharing and analytics and harness health data for discovery. Here, we show the reuse of the same design pattern for measurements to model health data for three different applications: 1) observations in patient registries; 2) lab measurements in hospitals; and 3) epidemiological measures in outbreaks.

2. The SIO Design Pattern Measurements

The Semantic Science Integrated Ontology (SIO) is an upper-middle level ontology that is commonly used to represent biomedical Linked Data [4]. SIO is an OWL ontology that provides a simple, integrated ontology of types and relations for rich description of objects, processes, and their attributes. It follows a worldview that primarily differentiates objects from processes: objects are entities that occupy space (in their mass or energy), persist in time, and maintain their identity even as they gain or lose parts. It also provides different Design Patterns (DP) such as the DP Measurements, which overlaps with our minimal data model for quantitative traits. The SIO DP Measurements [5] is a process-centric pattern that essentially relies on three concepts: entity, quantity and measuring process (Figure 1). Quantities have specific values that should be specified using the 'has value' datatype property and the datatype. Units can be specified using the Unit Ontology with the 'has unit' object property. Quantities are the result, i.e. the output of a measurement process and can be time-indexed to a time point or time interval. The measuring process specifies that the output of the process is the creation of a quantity. Entities can be described in terms of their quantified attributes. SIO also enables us to specify which qualities, capabilities or roles are involved in a particular process, so as to more richly describe the key components for that process to occur.

3. Applications

We reuse the SIO DP Measurements for three different applications, to represent:

3.1. Observations in Patient Registries

We apply the SIO DP Measurements to model patient observational health data. Patient registries are organized systems that use observational methods to collect data, including longitudinal data, on a population defined by a particular disease, condition, or exposure. In the Rare Disease (RD) domain, they constitute key tools to pool data to achieve a sufficient sample size for epidemiological and/or clinical research. The EJP RD is building a FAIR federated ecosystem to enable efficient RD research. To increase interoperability among the enormously fragmented data from RD patients contained in hundreds of registries across Europe, the EJP RD dedicates effort to build semantic data models for a set of common data elements defined for RD patient

¹European Joint Programme Rare Diseases (EJP RD) <https://www.ejprarediseases.org/>

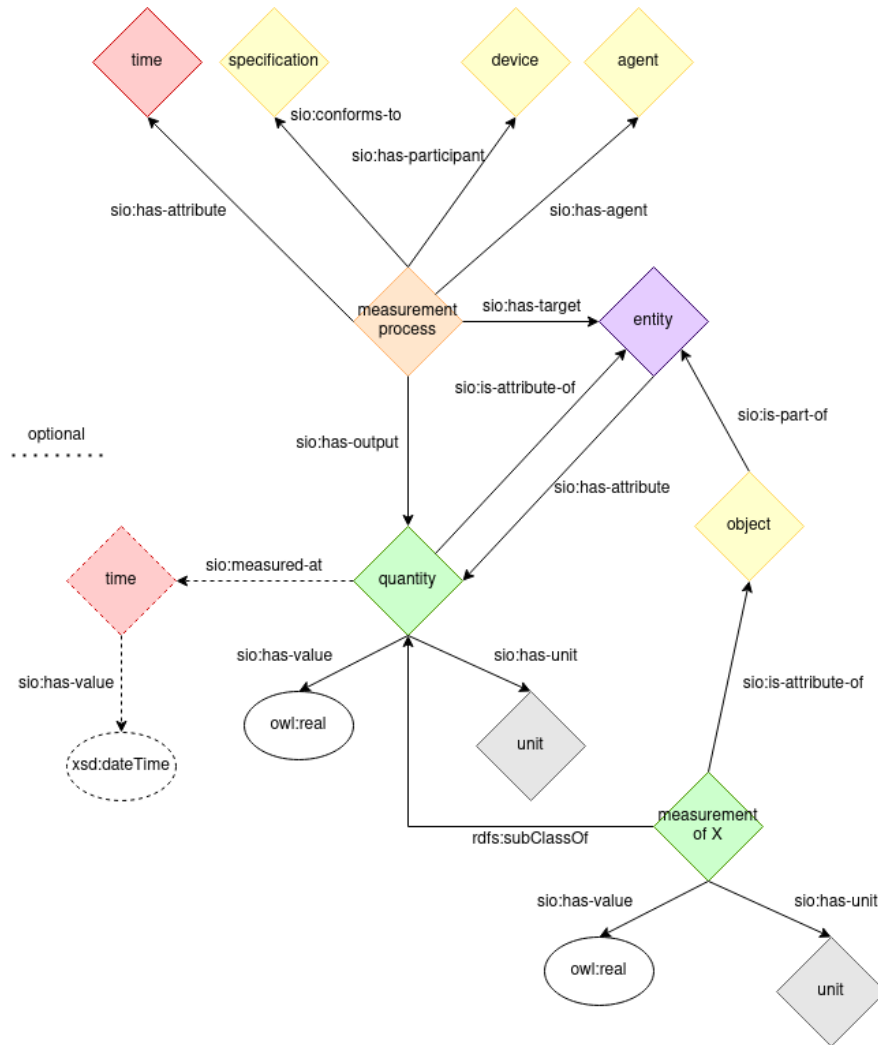


Figure 1: The SIO design pattern for measurements.

registries by the European Joint Research Centre². The SIO DP Measurements pattern is reused to provide the core foundation to build these semantic models to uniformly represent the observations collected in patient registries [6] (see the EJP RD core model for these semantic models, which is based on the SIO design pattern, in Figure 2). The modelling objective is to represent every observation as the result of some measurement process with patients, clinicians, and machines as participants. Application of the model will facilitate efficient, automated use of registries to identify new pathways for treatment, develop clinical research tools, and recruit potential participants for clinical trials.

²https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en

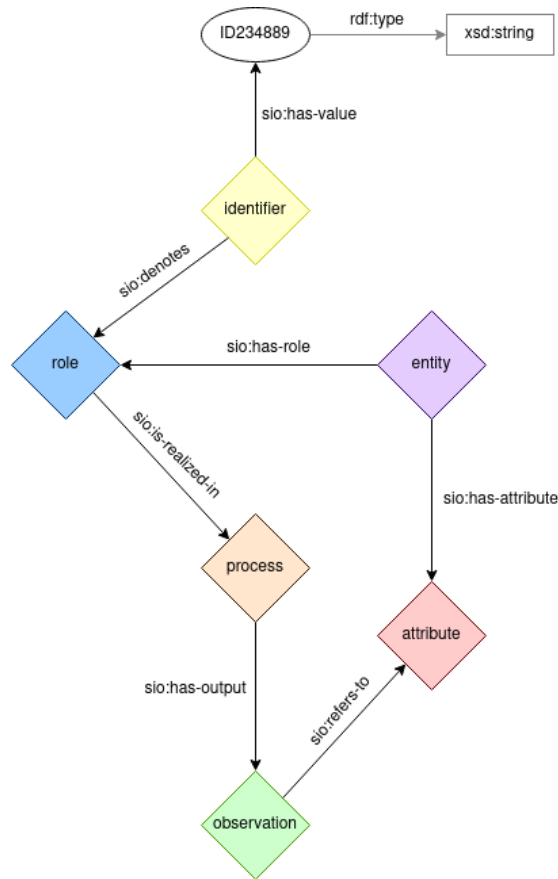


Figure 2: The EJP RD core model based on the SIO design pattern.

3.2. Laboratory Measurements in Hospitals

We apply the SIO DP Measurements to model patient quantitative health data. The worldwide COVID-19 pandemic stressed the need to have patient data available and accessible for gaining new insights timely and efficiently, not only within the hospital, but also across hospitals and countries. Clinicians monitor biomolecular concentrations, other physiological signs, and symptoms manifested in different organ systems of their patients at different points in time and collect multi-omics data that need to be integrated for computational analysis. These lab measurements are very valuable data because they give intrinsic information about the underlying biological mechanism and patient disease trajectory that could be used to make informed and tailored therapeutic decisions. The life science community has been developing different ontologies to represent molecular biology, clinical measures, and disease phenotypes. Based on the SIO DP Measurements and the EJP RD core model we are establishing an ontological linking model of heterogeneous data such as immunoresponse-related lab measurements [7] using OWL ontologies from the Open Biological and Biomedical Ontologies (OBO) Foundry, SIO, and other Semantic Web standards with the aim of making clinical data amenable for

analysis with Linked Open Data and further ‘ontologised’ Linked Data from other hospitals.

3.2.1. Integration into GA4GH Phenopackets Standard

Phenopackets is an exchange standard for the description of aberrant phenotypes of human subjects in relation to DNA sequence data, which is amenable for genomic research. Based on a minimal overlapping model of the SIO DP Measurements, we implemented the ‘measurement’ Phenopackets extension in v2 to characterize clinical measurements³.

3.3. Epidemiological Measures in Outbreaks

We apply the SIO DP Measurement to model quantitative epidemiological data. One year ago, the novel COVID-19 infectious disease emerged and spread, causing high mortality and morbidity rates worldwide. In the OBO Foundry [8], there are more than one hundred ontologies to share and analyse large-scale datasets for biological and biomedical sciences. However, this pandemic revealed that we lack tools for an efficient and timely exchange of this epidemiological data which is necessary to assess the impact of disease outbreaks, the efficacy of mitigating interventions and to provide a rapid response. In this work we reused the SIO DP Measurements to develop an OBO ontology [9]. We aligned the SIO DP Measurements to the OBO principles, and mapped classes and relations to OBO ontologies’ terms [10]. With the development of this OBO ontology we provide a compatible logical model for quantities that enables researchers to represent and share machine readable epidemiology surveillance data that can interoperate with other biomedical ontologies in the OBO Foundry for rapid analysis, modelling and response.

4. Demonstration

We used the SIO DP Measurements to develop ontological models amenable for analysis and the development of computer applications, such as semantic similarity, semantic mining, machine learning or feature embedding, reasoning and biomedical predictions. In this dynamic demonstration, we will show how to design semantic models using the SIO DP Measurements to represent three different health data sets. The aim is to make attendees gain understanding of the rationale underlying this SIO design pattern. Therefore, we will model some instances together, such as observations in patient registries, lab measurements, and epidemiological variables.

5. Discussion and Conclusion

Data harmonization based on DP enables efficient research. For example, it allows querying of heterogeneous data that were modelled using the same pattern. In Semantic Web applications, this feature is an opportunity to build SPARQL queries with a simple canonical graph pattern, thus not only improving interoperability of FAIR data, but also reusability. Furthermore, this harmonized representation of data at patient and population levels may also bring the opportunity

³<https://phenopacket-schema.readthedocs.io/en/v2/measurement.html>

to design an axiom pattern to link epidemiological data with additional clinical data. This may help to represent computable cohorts for precision medicine and raise the exciting opportunity to apply formal reasoning for knowledge discovery. While there are several ontologies and design patterns that capture measurements and are applied in similar contexts, e.g. LOINC ⁴ in clinical contexts, the Clinical Measurement Ontology ⁵ in some model organisms and a schema for the description of phenotypes [11], here we demonstrated that reusing the same design pattern for measurements can represent heterogeneous health data and can be applied in diverse contexts from clinical measurements in hospitals to elements in patient registries and measures in epidemiological studies for outbreak monitoring. Remaining challenges for cross-institutional analysis are for instance preserving patient data-privacy and safety. However, these challenges are not blockers for making data interoperable, i.e. they can be addressed in parallel. In summary, the application of the SIO DP Measurements resulted in three diverse biomedical applications: 1) the semantic harmonization of observational real world patient data; 2) the development of a semantic model for data integration within the hospital; and 3) the development of an OBO ontology for monitoring outbreaks. With the demonstration of the SIO DP Measurement, we aim to raise visibility and foster understanding on how to use it for health data modelling and integration. Future steps are the application of building ontology-based knowledge graphs and exploit harmonized patient data by federated query and analysis.

Acknowledgments

This initiative is supported by funding from the European Union's Horizon 2020 research and innovation program under the EJP RD COFUND-EJP N° 825575. We would also like to thank to the EJP RD, the GO FAIR VODAN, and the ZonMW Health Holland under the Trusted World of Corona, for supporting the research on FAIR data that was reused here. We would like to acknowledge that work in the BEAT-COVID project was partly funded by the Wake Up To Corona crowdfunding initiated by the Leiden University Fund (LUF).

References

- [1] N. Queralt-Rosinach, S. M. Bello, R. Hoehndorf, C. Weiland, P. Rocca-Serra, P. N. Schofield, Modeling quantitative traits for covid-19 case reports, medRxiv (2020). URL: <https://www.medrxiv.org/content/early/2020/06/20/2020.06.18.20135103>. doi:10.1101/2020.06.18.20135103. arXiv:<https://www.medrxiv.org/content/early/2020/06/20/2020.06.18.20135103.full.pdf>
- [2] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web., Scientific American 284 (2001) 34–43.
- [3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific data 3 (2016).

⁴<https://loinc.org/>

⁵<http://www.obofoundry.org/ontology/cmo.html>

- [4] M. Dumontier, et al, The semanticscience integrated ontology (sio) for biomedical research and knowledge discovery, *Journal of Biomedical Semantics* 5 (2014). doi:10.1186/2041-1480-5-14.
- [5] Sio dp measurements homepage, 2014. URL: <https://github.com/MaastrichtU-IDS/semanticscience/wiki/DP-Measurements>.
- [6] R. Kaliyaperumal, M. D. Wilkinson, P. Alarcón Moreno, N. Benis, R. Cornet, B. dos Santos Vieira, M. Dumontier, C. H. Bernabé, A. Jacobsen, C. M. A. Le Cornec, M. P. Godoy, N. Queralt-Rosinach, L. J. Schultze Kool, M. A. Swertz, P. van Damme, K. J. van der Velde, N. van Lin, S. Zhang, M. Roos, Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data, *medRxiv* (2021). URL: <https://www.medrxiv.org/content/early/2021/07/30/2021.07.27.21261169>. doi:10.1101/2021.07.27.21261169. arXiv:<https://www.medrxiv.org/content/early/2021/07/30/2021.07.27.21261169.full.pdf>.
- [7] Lumc lab measurement model graph, 2021. URL: https://github.com/NuriaQueralt/beat-covid/blob/master/fair-data-model/cytokine/model-triples/lab_measurement_semantic_model.png.
- [8] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, S. Lewis, The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration, *Nature Biotechnology* 25 (2007) 1251–1255. doi:doi : 10.1038/nbt1346.
- [9] The cemo ontology owl file, 2021. URL: <https://github.com/NuriaQueralt/covid19-epidemiology-ontology/blob/main/owl/cemo.owl>.
- [10] Cemo model graph, 2021. URL: https://github.com/NuriaQueralt/covid19-epidemiology-ontology/blob/main/images/covid19_epidemiology_model.png.
- [11] G. Gkoutos, E. Green, A. Mallon, et al., Using ontologies to describe mouse phenotypes., *Genome Biol* 6 (2005). doi:10.1186/gb-2004-6-1-r8.