

Workshop machine-actionable Software Management Plans

Organizer: Semantic Technologies team at ZB MED Information Centre for Life Sciences

Place: Cologne

Date: 2023.05.31

Participants/Authors of this report: Olga Giraldo ¹[0000-0003-2978-8922], João Cardoso ²[0000-0003-0057-8788], Eva Martin del Pico ³[0000-0001-8324-2897], Alban Gaignard ⁴[0000-0002-3597-8557], Lukas Geist ¹, Yves Vincent Grossmann ⁵[0000-0002-2880-8947], Fotis Psomopoulos ⁶[0000-0002-0222-4273], Elli Papadopoulou ⁷[0000-0002-0893-8509], Dhvani Solanki ¹, Leyla Jael Castro ¹[0000-0003-3986-0510]

¹ ZB MED Information Centre for Life Sciences

² RDA DMP Common Standards Working Group

³ BSC-CNS

⁴ CNRS

⁵ Max Planck Digital Library

⁶ Centre for Research and Technology Hellas

⁷ ATHENA Research Center / OpenAIRE

Introduction

The [Semantic Technologies team](#) at [ZB MED](#) initiated a project to add machine-actionability to the [ELIXIR Software Management Plans \(SMPs\)](#) [1] in December 2022. The initial phase of the project, funded by RDA/EOSC Future, concluded in May 2023 with a workshop where experts working with SMPs and machine-actionable Data Management Plans met together. The project will continue under the umbrella of the [NFDI4DataScience](#) consortium.

The purpose of the workshop was validating the metadata schema support maSMPs [2, 3], improving its alignment with existing SMPs, and identifying gaps wrt ELIXIR SMPs and [RDMO SMPs](#) [4] and Research Data Alliance (RDA) maDMPs [5, 6]. The workshop counted with the participation of 10 people, from [ELIXIR Software Development Best Practices Group](#), [Bioschemas](#) [7], [RDMO](#) SMPs, [RDA DMP Common Standards Working Group](#), [ARGOS](#) and [ZB MED](#). In addition to the resources represented by participants, we also include [Codemeta](#) [8] in the analysis. Here we present a report of what happened and what was achieved.

Presentations

We had five presentations as follows:

1. The ELIXIR Software Management Plans (file P1_The ELIXIR Software Management Plan.pdf)
2. Sustainable and FAIR Software in Research - A RDMO Catalogue for Software Management Plans (file P2_RDMO_SMP.pdf)

3. RDA's Approach to Machine Actionable Data Management Plans (P3_maDMP.pdf)
4. machine-actionable Software Management Plans (file P4_maSMP.pdf)
5. (more) Findable bioinformatics softwares with Bioschemas (file P5_Bioschemas.pdf)

Ontology validation

ELIXIR and RDMO

The complete set of properties from software source code and software release were analyzed. The first part of the analysis was focused on identifying which properties proposed in our metadata schema are covering the questions specified in the ELIXIR and RDMO SMP models. The coverage level has three possible values as follows: i) “yes” when the property fully covers one or more questions, ii) “partially” when the property covers either part of the questions or is related to them, iii) “not” when the property does not correspond to any question (however it could represent a possible improvement in the questionnaire if added). The reviewed properties from software source code is available in [Table 1](#) (see file T1_maSMP-SoftwareSourceCode-ELIXIR-RDMO.tsv). The reviewed properties from software release are available in [Table 2](#) (see file T2_maSMP-SoftwareRelease-ELIXIR-RDMO.tsv). Main outcomes of this stage are listed below:

- Identification of commonalities. It was possible to identify, from the ELIXIR and RDMO questionnaires, similar questions related to a specific property.
 - Example, “What programming languages are you using in your project?” (from ELIXIR), and “Which programming language(s) do you plan to use?” (from RDMO) were linked to the property programmingLanguage.
- Identification of questions not covered by our metadata schema. It was possible to identify, from the same questionnaires, a subset of requirements difficult to represent in the proposed metadata schema.
 - Example, “*How do you capture the environment?*” (from ELIXIR), and “*How is software documentation created?*” (from RDMO). A list of questions not covered by our metadata schema is detailed in [Table 3](#) (see file T3_maSMP_NotCovered.tsv).

Bioschemas y Codemeta

The second part of the analysis was focused on identifying which properties proposed in our metadata schema are covering terminology from the [ComputationalTool Profile](#) in Bioschemas and terminology from [Codemeta](#). Software source code alignment was only done against Codemeta because the terminology proposed at the ComputationalTool Profile in Bioschemas is suitable just for software release. The Software source code alignment to Codemeta, is available in [Table 4](#) (see file T4_maSMP-SoftwareSourceCode-Codemeta.tsv). The Software release alignment to Codemeta and Bioschemas, is available in [Table 5](#) (see file T5_maSMP-SoftwareRelease-Codemeta-Bioschemas.tsv). The coverage level has three possible values as follows: i) “yes” when the maSMP property is the same as the one used in the other vocabulary,, ii) “partially” when the maSMP property corresponds to a property with a

different name in the other vocabularies,, iii) “not” when the maSMP property does not have a corresponding property in the other vocabularies.

RDA maDMP

An initial draft mapping the maDMP, and corresponding ontology DCSO, to schema.org was created, see [Table 6](#) (see file T6_maDMP-DCSO-schemaorg.tsv). This draft could be a starting point for the maSMP metadata schemas to include additional elements related to the actual plan, the project, the funders and so on.

Conclusions

This workshop represented a first step to achieve an alignment between the different parties involved in the existing SMP models.

An enrichment of our metadata schemas was achieved. A new version of our metadata schema in the form of an ontology was obtained in order to increase the coverage of questions proposed by ELIXIR and /or RDMO. The ontology is available at Zenodo [9] and GitHub [10] while documentation is provided via [GitHub pages](#).

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101017536 and is part of the Research Data Alliance and European Open Science Cloud Future call 2022.

This project has received funding from NFDI4DataScience project funded by Deutsche Forschungsgemeinschaft DFG, project no. 460234259.

References

1. Alves, R., Bampalikis, D., Castro, L., Fernández, J. M., Harrow, J., Kuzak, M., ... Via, A. (2021, October 25). ELIXIR Software Management Plan for Life Sciences. <https://doi.org/10.37044/osf.io/k8znb>
2. Giraldo, O., Geist, L., Quiñones, N., Solanki, D., Rebholz-Schuhmann, D., & Castro, L. J. (2023). Machine-actionable Software Management Plan Ontology (maSMP Ontology). Zenodo. <https://doi.org/10.5281/zenodo.7976401>
3. Giraldo, O., Geist, L., Quiñones, N., Solanki, D., Alves, R., Bampalikis, D., Fernandez, J. M., Martin Del Pico, E., Psomopoulos, F. E., Via, A., Rebholz-Schuhmann, D., & Castro, L. J. (2023). A metadata schema for machine-actionable Software Management Plans [Application/pdf]. DaMaLOS 2023. <https://doi.org/10.4126/FRL01-006444988>
4. Klar, J., Michaelis, O., Engelhardt, C., Enke, H., Frenzel, J., Hausen, D., Jagusch, G., Kramer, C., Lindstädt, B., Ludwig, J., Heike, N., Straka, J., Strötgen, R., Ulrich, R., Wedlich-Zachodin, K., & Wuttke, U. (2023). Research Data Management Organizer (RDMO) [Python]. <https://doi.org/10.5281/zenodo.596581>

5. Miksa, T., Walk, P., & Neish, P. (2023). RDA DMP Common Standard for Machine-actionable Data Management Plans (1.1). <https://doi.org/10.15497/rda00039>
6. Cardoso, J., Castro, L. J., Ekaputra, F. J., Jacquemot, M. C., Suchánek, M., Miksa, T., & Borbinha, J. (2022). DCSO: Towards an ontology for machine-actionable data management plans. *Journal of Biomedical Semantics*, 13(1), 21. <https://doi.org/10.1186/s13326-022-00274-4>
7. Gray, A.J.G, Goble, C.A. and Jimenez, R., 2017. Bioschemas: From Potato Salad to Protein Annotation. In *International Semantic Web Conference (Posters, Demos & Industry Tracks)*.
8. Matthew B. Jones, Carl Boettiger, Abby Cabunoc Mayes, Arfon Smith, Peter Slaughter, Kyle Niemeyer, Yolanda Gil, Martin Fenner, Krzysztof Nowak, Mark Hahnel, Luke Coy, Alice Allen, Mercè Crosas, Ashley Sands, Neil Chue Hong, Patricia Cruse, Daniel S. Katz, Carole Goble. 2017. CodeMeta: an exchange schema for software metadata. Version 2.0. KNB Data Repository. <https://doi.org/10.5063/schema/codemeta-2.0>
9. Giraldo Olga, Geist Lukas, Quiñones Nelson, Solanki Dhvani, Rebholz-Schuhmann Dietrich, & Castro Leyla Jael. (2023). machine-actionable Software Management Plan Ontology (maSMP Ontology) (1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.8089518>
10. Giraldo O, Geist L, Quiñones, Lukas, Solanki, Dhvani, Rebholz-Schuhmann D, Castro LJ. maSMPs - Ontology and Software. 2023. Available: <https://github.com/zbmed-semantic/maSMPs>