A metadata analysis for machine-actionable Software Management Plans

Olga Giraldo [1], Renato Alves [2], Dimitrios Bampalikis [3], José M. Fernández [4], Eva Martin del Pico [4], Fotis Posomopoulos [5], Nelson Quiñones [1], Dhwani Solanki [1], Allegra Via [6] and Leyla Jael Castro [1]

[1] *ZB MED Information Centre for Life Sciences, Cologne, Germany*
[2] *European Molecular Biology Laboratory, Heidelberg, Germany*
[3] *National Bioinformatics Infrastructure Sweden, Uppsala, Sweden*
[4] *Barcelona Supercomputing Center, Barcelona, Spain*
[5] *Centre for Research and Technology Hellas, Thessaloniki, Greece*
[6] *Institute of Molecular Biology and Pathology - National Research Council, Rome, Italy*

Abstract

Data Management Plans (DMPs) describe the data management lifecycle for the data corresponding to a research project, including activities from collection to preservation. Machine-actionable DMPs improve text-based DMPs by adding a semantic layer representing the most common elements relevant to DMPs, from datasets to funders. Similar to DMPs, Software Management Plans (SMPs) follow the software management lifecycle. The ELIXIR SMP was developed to support life science researchers and making it easier to follow research software good practices aligned to the findable, accessible, interoperable and reusable principles for research software. Currently, the ELIXIR SMP is a questionnaire-based document that would benefit from a machine-actionable approach. Here, we present a preliminary metadata analysis including possible types and properties from Schema.org that could be used to model machine-actionable SMPs.

Keywords

Research software management plans, machine-actionability, metadata analysis
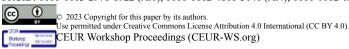
1. Metadata analysis

As a first step for the machine-actionable version of the ELIXIR Software Management Plan (maSMP) [1], we have created an initial mapping from its questions to types and properties from Schema.org [2]. We summarize our first draft in Table 1.

Table 1

Possible types and properties from schema.org supporting maSMPs

Section	Question	schema.org type and properties
Accessibility and license	How can the software be accessed by third parties?	SoftwareSourceCode (codeRepository, conditionsOfAccess), CreativeWork (url)
	Does your software have a license?	CreativeWork (license)

Documentation	What type of documentation is available?	SoftwareSourceCode (workExample, workTranslation) or CreativeWork (isBasedOn)
	Is the purpose of the software stated in the documentation?	Thing (disambiguatingDescription)
	Does the documentation describe how to test, use, build, deploy, install	Thing (disambiguatingDescription)
Testing	What type of testing do you use?	—
	Are sample data and/or parameters provided?	SoftwareApplication (supportingData), SoftwareSourceCode (usageInfo)
Interoperability	Do you use existing and standard input/output formats?	
Versioning	Do you use a version control system?	—
	Do you use Semantic Versioning?	SoftwareApplication (softwareVersion)
Reproducibility	Do you provide releases of your software?	SoftwareApplication (releaseNotes)
	How do you define language-specific dependencies of your software and their version? How do you capture the environment necessary to run the software?	SoftwareApplication (softwareRequirements)
	Do you provide input and output examples that can be used to reproduce the functioning of your software?	SoftwareApplication (supportingData)
	Do you state how to report bugs and/or usability problems by the software user(s)?	—
Recognition	Do you include citation information and ORCID?	CreativeWork (citation)
	Do the releases have a PID?	—
	Is the software registered in a domain-specific registry?	CreativeWork (isPartOf)

We will extend the model and ontology supporting machine-actionable version of Data Management Plans (maDMPs) [3,4] to include the case posed by maSMP. We will further develop the mapping to schema.org and Bioschemas [5].

2. Acknowledgments

3. References

[1] Alves R, Bampalikis D, Castro LJ, González JMF, Harrow J, Kuzak M, et al. ELIXIR Software Management Plan for Life Sciences. BioHackrXiv; 2021. doi:10.37044/osf.io/k8znb

[2] Guha RV, Brickley D, Macbeth S. Schema.org: evolution of structured data on the web. Commun ACM. 2016;59: 44–51. doi:10.1145/2844544

[3] Miksa T, Simms S, Mietchen D, Jones S. Ten principles for machine-actionable data management plans. PLOS Computational Biology. 2019;15: e1006750. doi:10.1371/journal.pcbi.1006750

[4] Cardoso J, Castro LJ, Ekaputra FJ, Jacquemot MC, Suchánek M, Miksa T, et al. DCSO: towards an ontology for machine-actionable data management plans. Journal of Biomedical Semantics. 2022;13: 21. doi:10.1186/s13326-022-00274-4

[5] Gray AJG, Goble C, Jimenez RC. From Potato Salad to Protein Annotation. ISWC Posters and Demo session. Vienna, Austria; 2017. p. 4. Available: http://ceur-ws.org/Vol-1963/paper579.pdf