

CRIS systems integration as a case study for the Metadata Schema and Crosswalk Registry

Joonas Kesäniemi^{1, 2}, Tommi Suominen¹, Dragan Ivanović³, and
Jan Dvořák⁴

¹CSC - IT Center for Science, Espoo, Finland

²Aalto University, Espoo, Finland

³University of Novi Sad, Novi Sad, Serbia

⁴Charles University, Prague, Czech Republic

February 2024

The [Metadata and Schema Crosswalk Registry](#) (MSCR) is being developed in the [FAIRCORE4EOSC project](#). The MSCR allows registered users and communities to create, register and version schemas and crosswalks. The published content can be searched, browsed and downloaded without restrictions. The MSCR also provides a mechanism to facilitate the transformation of metadata records from one format to another via registered crosswalks. The key objective of the MSCR is to make the crosswalks/mappings more FAIR. FAIR refers to the four now well-established principles of Findability, Accessibility, Interoperability and Reusability [Wilkinson et al., 2016]. The MSCR provides communities and individual researchers with the possibility to manage their metadata schema and/or relevant metadata schema crosswalks.

In the FAIRCORE4EOSC project, we are applying the MSCR to improve the interoperability of Current Research Information Systems (CRISs). One example of the data flows being implemented is the harvesting of metadata records from the Finnish national CRIS research.fi to the international OpenAIRE e-infrastructure. Instead of creating a direct mapping between source and target system's data models, the goal is use the Common European Research Information Format (CERIF) [Jörg, 2010, Jeffery et al., 2014] as the common core conceptual model. CERIF is a research information exchange format, which is an international standard for the exchange of information between CRISs or between CRISs and other information systems, e.g. OA Repositories, Research Data systems and others, both on a local, national or international level. However, in its [currently released](#) form, CERIF is tightly coupled with the relational database schema, which limits its global applicability. CERIF has aged, more in its format than its content, as it does not provide us with an Linked Data

enabled entities for creating crosswalks. Preferably PIDS or at least URIs are needed in order to point to specific attributes and classes in the crosswalks.

euroCRIS has been working towards a new refactored CERIF since 2021 [Dvořák and Ivanović, 2023] and this work has been continued in the context of FAIRCORE4EOSC project. The modernized CERIF2 comes with a new modeling notation, modular structure and adaptation for use in APIs and as part of Linked Open Data. In the beginning of 2024, the work has focused on extending the CERIF core with more use case specific modules. For the purpose of FAIRCORE4EOSC project, at least Scholarly Publication module and Dataset module should be realized.

Related work by the Research Data Alliance (RDA) Scientific Knowledge Graphs - Interoperability Framework Working Group (SKG-IF WG) is aiming to develop a framework (including model and schema) for information exchange between Scientific Knowledge Graphs (SKGs). OpenAIRE is providing the EOSC Research Discovery Graph (RDGraph), which is an instance of an SKG. The SKG Interoperability Framework release v1.0 of the SKG-IF provides a first release of the JSON based serialization of the model.

The MSCR will be used to create crosswalks from research.fi to CERIF2 and from CERIF2 to SKG-IF Schema. It will also facilitate the operationalisation of the crosswalks i.e transformation of metadata records from the national CRIS to the international RDGraph using CERIF as the common core interchange format. We are working on models and tools that would support mappings in semantic, structural and syntactic levels [Ouksel and Sheth, 1999]. This would also include the utilization of semantic mappings between vocabularies referenced in the schemas associated with the crosswalk, e.g. mapping between values used to represent field of science and similar classification schemes. All information required for transformations can be maintained within the MSCR.

The set of interoperable of systems can be extended by creating crosswalks from e.g. the Flemish FRIS [Dengis and Bal, 2022], Greek Helix, VIVO to CERIF (and vice versa). By piping the results of crosswalk operationalizations together it is technically possible for these systems to export and import metadata to and from any target system (with an pre-existing CERIF crosswalk) without the need to produce all the one-to-one mappings.

We will present the current status of the CERIF2 related work and discuss the key issues discovered while documenting mappings between research.fi API outputs and the more conceptual CERIF2 core model. Finally, we intend to demonstrate the use of MSCR by first turning crosswalk documentation from spreadsheets into MSCR mappings and then transforming sample metadata records from one format to another.

Acknowledgements

The work described in this extended abstract has taken place in the context of the FAIRCORE4EOSC-project. FAIRCORE4EOSC has received funding from the EU’s Horizon Europe research and innovation programme under Grant

Agreement no. 101057264.

References

- P. B. Dengis and I. D. Bal. Monitoring open science policy using a regional cris – the flanders case with fris. *Procedia Computer Science*, 211:101–110, 1 2022. ISSN 1877-0509. doi: 10.1016/J.PROCS.2022.10.181.
- J. Dvořák and D. Ivanović. Refactoring cerif and mapping it to research. fi. 2023.
- K. Jeffery, N. Houssos, B. Jörg, and A. Asserson. Research information management: the cerif approach. *International Journal of Metadata, Semantics and Ontologies*, 9:5, 2014. ISSN 1744-2621. doi: 10.1504/IJMSO.2014.059142. URL <http://www.inderscience.com/link.php?id=59142>.
- B. Jörg. Cerif: The common european research information format model. *Data Science Journal*, 9:CRIS24–CRIS31, 7 2010. ISSN 1683-1470. doi: 10.2481/dsj.CRIS4. URL http://ec.europa.eu/research/era/index_en.html<http://datascience.codata.org/articles/abstract/10.2481/dsj.CRIS4/>.
- A. M. Ouksel and A. Sheth. Semantic interoperability in global information systems. *SIGMOD Rec.*, 28(1):5–12, mar 1999. ISSN 0163-5808. doi: 10.1145/309844.309849. URL <https://doi.org/10.1145/309844.309849>.
- 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, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. ’t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 3 2016. ISSN 2052-4463. doi: 10.1038/sdata.2016.18.