IMPLEMENTATION STORIES

FAIRSFAIR
Fostering Fair Data Practices in Europe

IMPLEMENTATION STORIES

Developing the SSHOC reference ontology

Authors
Ricarda Braukmann (DANS)

From interviews with
Athina Kritsotaki, Eleni Tsouloucha and Chrysoula Bekiari from FORTH - Institute for Computer Science[1] - who coordinated the development of the SSHOC Reference Ontology

Introduction

The SSHOC Reference Ontology (SSHOCro) is developed in the context of the Social Sciences and Humanities Open Cloud (SSHOC)[2] project. The SSHOC project aims to support the discipline-specific part of the European Open Science Cloud (EOSC)[3] focused on the Social Sciences and Humanities (SSH). The SSHOCro[4] was developed within Task 4.7 Modelling the SSHOC Life Cycle as part of SSHOC Work Package 4 focusing on innovations in data production. The goal of the SSHOCro is to establish a common framework for organising knowledge around all steps in the research data life cycle within the SSH domain. The model - described in Resource Description Framework (RDF) schema - is event-based and aims to capture all the relevant scientific activities of the data lifecycle including the tools, datasets and services used at each phase. SSHOCro is based on the CIDOC-CRM[5] , an ontology for information integration in the field of cultural heritage. The work on SSHOCro was coordinated by FORTH, in particular by Athina Kritsotaki, Eleni Tsouloucha, Chrysoula Bekiari and Maria Theodoridou, hereafter referred to as "the authors" of the SSHOCro. Various SSHOC partners and stakeholders were consulted in SSHOCro's development, including data archives and repositories, research infrastructures and data catalogues. The work on SSHOCro started in March 2019 and a final model will be presented at the end of the SSHOC project in March 2022. The ontology will however be further developed and refined after the project.

FAIRsFAIR recommendation

"Describe research outputs using agreed terminologies and metadata standards to make data FAIR"

FAIRsFAIR Recommendations on practice to support FAIR principles

1. Website FORTH: https://www.ics.forth.gr/about-ics
2. More information on SSHOC: https://www.sshopencloud.eu/ [accessed 10th november 2021]
3. Gotz et al (2020) ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC, https://doi.org/10.5281/zenodo.3675080.
4. The SSHOCro beta version is described in SSHOC deliverable D4.18 https://doi.org/10.5281/zenodo.3744860
5. More information on CIDOC-CRM: http://www.cidoc-crm.org/ [accessed 10th november 2021]

Development of the ontology

The goal of the SSHOCro

The goal of the SSHOC Reference Ontology (SSHOCro) is to create a semantic interoperability framework for the description of the data lifecycle in the SSH domain[6]. The SSHOCro can be used to document each stage of the research work, trying to capture the tools and services used at each point in the data lifecycle. The ontology was developed to be used across different fields within the SSH domain to provide a common standard allowing information to be shared and exchanged in a standardised manner.

Initial development of the ontology

The authors of the SSHOCro started their work in 2019 by collecting an overview of existing workflows and metadata standards used in the SSH domain. A literature review was conducted assessing papers that describe methods used to document the various stages of the research work[7]. From the literature, a set of replication studies[8] were selected as use cases to be documented using the SSHOCro model.

In addition to the literature research, a workshop[9] was organised to consult the broader SSH community and find representative workflows.

The information gathered through this desk and community research was analysed and the initial model of the SSHOCro was developed. The basis for the SSHOCro was the Conceptual Reference Model from the Center for Intercultural Documentation (CIDOC-CRM)[5], the ISO standard ontology for Cultural Heritage data.

The SSHOCro model is described in RDF and the beta version RDF schema has been published in March 2020[10], while a revised version will be made available at the beginning of 2022.

Mapping to SSH metadata standards

After its initial release, the SSHOCro model was tested and revised with the help of various SSHOC partners and other stakeholders in the SSH domain. In particular, repositories working with social science and humanities data were consulted, including the Finnish Social Science Data Archive (FSD), The Ethnic and Migrant Minority (EMM) Survey Registry, and LINDAT CLARIN repository. Based on this consultation, two metadata documentation standards frequently used in the SSH were selected to map to the SSHOCro as a proof of concept: DDI (Data Documentation Initiative) and CMDI[11] (Component MetaData Infrastructure). The two metadata schemas were integrated and harmonised with the SSHOCro using use cases from the EMM/FSD repositories (DDI) and LINDAT (CMDI) . The tool used for the transformation of the data of the selected use cases was the X3ML (3M) Toolkit[12].

The updated model was further harmonised with outputs from other SSHOC tasks, including Task 3.5 which worked with metadata and data format interoperability issues and built an interoperability hub. The SSHOCro was also assessed and refined against the ontology used in the SSHOC Open Marketplace.

6. More background is available in SSHOC deliverable D4.18 https://doi.org/10.5281/zenodo.3744860
7. The literature on workflows can be found in the Bibliographic References section - D4.18, https://doi.org/10.5281/zenodo.3744860
8. In particular, work by Camerer and colleagues (2018) who evaluated the replicability of social science experiments was consulted https://osf.io/pfdyw/
9. The workshop was realised as part of SSHOC Task 4.7 and is described in an internal milestone document.
10. The RDF schema is published in Zenodo (https://doi.org/10.5281/zenodo.3744925) as as supplement to D4.18 (https://doi.org/10.5281/zenodo.3744860)
11. More information about the mapping between DDI, CMDI and SSHOCro can be found in deliverable D4.19 https://doi.org/10.5281/zenodo.4457496
12. More information about the X3ML (3M) Toolkit is available here: https://www.ics.forth.gr/isl/x3ml-toolkit [accessed 10th november 2021]

Development of the ontology

Ongoing process

The development of the SSHOCro is still ongoing and the authors plan to collect additional use cases to validate the SSHOCro. Envisioned use cases include national galleries, social data networks, as well as other stakeholders in the SSH domain. As part of the SSHOC project, a revised model of the SSHOCro will be presented at the end of the SSHOC project in March 2022, yet the ontology can be further developed after the end of the SSHOC project.

Challenges encountered

The biggest challenge

The biggest challenge that was experienced by the authors of SSHOCro was that the model is based around an event-based workflow approach that is used to describe the various steps of the research cycle. The available information in existing metadata, however, often did not reflect this workflow structure. Most research papers and documentation of datasets are based on the final results rather than on the processes that created them.

Another challenge is to move from the initial model towards a standard that is implemented and used in real-life research examples across the SSH domain. The SSHOCro model has been presented at various occasions and co-developed with stakeholders from the SSHOC community. Next steps include further validation and use of the ontology to describe research projects so the SSHOCro can be updated and further refined in the future.

SSHOCro future steps

Harmonise the data

The SSHOCro was developed to harmonise the documentation and organisation of knowledge about data generated in the social sciences and humanities. It provides a standardised model in an RDF format to capture provenance information as well as describing the tools and conditions related to a given data object at each step of the life cycle.

The development of the SSHOCro model is a first step towards a more standardised and organised description of knowledge that facilitates data sharing and replication of research results within the SSH domain. The SSHOCro is currently released as a beta version which is being revised with feedback from the community.

At the end of the SSHOC project in the beginning of 2022, an updated version will be published which can be further used and refined in the future.

Further information

Bekiari,C., Kritsotaki, A., & Tsouloucha, E.. (2020). SSHOC D4.18 SSHOC Reference Ontology (beta version) (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3744861

Tsouloucha,E, Kritsotaki,A., Bekiari,C. & Theodoridou,M. (2021). D4.19 Mapping of two indicative selected standards to the SSHOCro (v1.0). Zenodo. https://doi.org/10.5281/zenodo.4457496

Tsouloucha,E, Kritsotaki,A., Bekiari,C. & Theodoridou,M. (2021). D4.19 linked files. https://doi.org/10.5281/zenodo.5106461

Kleemola,M., et al., (2021). D3.6 Report on SSHOC format interoperability solution services, including new software. https://doi.org/10.5281/zenodo.5561604

Bekiari,C. et al., (2020). MS 20 Selection of SSH metadata standards for mapping to SSHOCro (v1.0). Zenodo. https://doi.org/10.5281/zenodo.4582018

About FAIRsFAIR Implementation Stories

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report Turning FAIR into Reality and the FAIRsFAIR Recommendations on practice to support FAIR principles. FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

FAIRsFAIR - Fostering FAIR Data Practices in Europe

European Commission Grant Agreement No. 831558
H2020-INFRAEOSC-2018-4
www.fairsfair.eu - support@fairsfair.eu

Acknowledgements

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

Disclaimer

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

Copyright rests with the authors