SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: CA15137 (ENRESSH)
STSM title: Metadata standards in the frame of the ENRESSH-VIRTA information service
STSM start and end date: 01/09/2018 to 15/09/2018
Grantee name: Dragan Ivanovic

PURPOSE OF THE STSM/

The ENRESSH-VIRTA project has been started within the ENRESSH COST action in order to prepare a roadmap for a European database of outputs of SSH research which is one of the goals of the ENRESSH COST action.

The main objective of the proposed STSM is to explore the current technology and data model of ENRESSH-VIRTA, as well as to investigate the potential use of international metadata standards, CERIF in particular in harvesting and exporting publication data in the frame of the European research information service. Thus, the main objective of the proposed STSM corresponds directly to the aims of the ENRESSH COST action, particularly in preparation of a roadmap for a European database of outputs of SSH research.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

- At the beginning of the STSM, Hanna-Mari Puuska and Joonas Nikkanen welcomed Dragan Ivanović and introduced him with working environment, CSC building and the office.
- After that, a couple of meetings with VIRTA team have been organized. Dragan introduced himself and described his background to the members of VIRTA team. Also, VIRTA team members described their background and roles in the development and maintenance of the VIRTA platform. The STSM agenda and aims were discussed. Hanna-Mari presented ENRESSH VIRTA pilot project and its main results. Collected data and its format are analyzed. Joonas Nikkanen and Lauri Hellsten presented technical aspect and architecture of VIRTA platform, as well as data model. Also, implementation of OAI-PMH VIRTA server side was presented. In that moment Dublin Core and VIRTA internal model in XML format were supported OAI-PMH VIRTA server side. The purpose of those meetings was preparation for discussion about extension of OAI-PMH VIRTA API with support of exporting data in accordance with OpenAIRE Guidelines for CRIS Managers (https://openaire-guidelines-for-

cris-managers.readthedocs.io/en/latest/index.html).

- Then, we defined three steps needed to complete previously mentioned extension of OAI-PMH VIRTA API:
    a. Description of mapping VIRTA data model to format prescribed by OpenAIRE Guidelines for CRIS Managers (a profile of CERIF format).
    b. Implementation of previously described mapping using TSQL stored procedure. The result of this stored procedure should be prepared data for exporting in a format prescribed by the OpenAIRE Guidelines.
    c. Extension of OAI-PMH server side implementation:
        i. Support of the **oai_cerif_openaire** metadata prefix.
        ii. Extension of supported list of sets with **openaire_cris_publications**
        iii. Retrieving records - If an OAI-PMH client requests the **oai_cerif_openaire** prefix and **openaire_cris_publications** set, previously prepared data should be returned
- Dragan worked on description of mapping till end of the first week.
- In the second week of the STSM, Lauri and Dragan worked on the implementation of previously described mapping using TSQL stored procedure.
- The last two days of the STSM, the source code of the VIRTA OAI-PMH server side implementation was analyzed by Dragan and VIRTA team members in order to define strategy for its extension.
- At the end of the STSM, Dragan and VIRTA team members discussed further collaboration and research.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

Mapping of VIRTA data model to the format prescribed by the OpenAIRE Guidelines is described: https://docs.google.com/spreadsheets/d/1J8fBScgZ8uVwnro8helPqFXaitlCdXjEzW7joEJl2O4/edit?usp=sharing. Taking into account ENRESSH VIRTA is based on the same data model as the VIRTA platform, the same mapping can be adopted for ENRESH VIRTA with two modifications:

1. Scientific fields should be harmonized and machine and human readable vocabulary for those fields should be defined.
2. Also, publication types should be harmonized and mappings of those types to COAR resource types adopted by the OpenAIRE Guidelines for CRIS managers (https://openaire-guidelines-for-cris-managers.readthedocs.io/en/latest/cerif_xml_publication_entity.html)

Described mapping is implemented using TSQL stored procedure. The implementation was tested using small sample of VIRTA data. Before it is going to be put into operation testing of performance should be conducted with real set of data.

Strategy for extension of the VIRTA OAI-PMH server side implementation has been developed and extension points have been identified. In the next period, the implementation should be performed by VIRTA team members in cooperation with Dragan.

The technical development as regards VIRTA-ENRESSH achieved during this STSM promotes the aims of the WP3 by providing a way of mapping the VIRTA to an international standard, CERIF in particular. It enables the VIRTA publication data to be provided to external users. The work can be exploited in further development of harvesting metadata from various local and national publication databases which use different data models.

The STSM also supplements the current contextual and theoretical work done in ENRESSH by way of providing the needed technical and pragmatic perspective for the development of the European research information

infrastructure in SSH.

FUTURE COLLABORATIONS (if applicable)

Besides completing implementation of the VIRTA OAI-PMH server side extension, further collaboration includes presenting the result of this STSM on next Strategic meeting of euroCRIS which will be held in November 2018 in Warsaw. After collecting feedbacks from that meeting, we are going to write a short paper and to submit to some international conference.

Moreover, the mapping in the opposite direction (from a CERIF profile to VIRTA data model) can be a subject of further collaboration. That mapping will enable ENRESSH VIRTA platform to collect data from CERIF compatible CRIS systems across the Europe.

Further collaboration could be achieved using the contemporary ICT technologies such as email and Skype, or through some funding program for supporting joint research activities.