Implementation of OpenAIRE Guidelines for CRIS managers to Finnish VIRTA Publication Information Service

Ivanović Dragan[1] and Puuska Hanna-Mari[2]   Nikkanen Joonas[2]   Hellsten Lauri[2]   Eskola Olli[2]

[1]University of Novi Sad

[2]Research Information Management and Interoperability, CSC – IT Center for Science, Espoo, Finland

ENRESSH - Podgorica 2019

Outline

1. **Introduction**
   - STSM
   - VIRTA
   - OpenAIRE

2. **Methodology**

3. **Implementation**

4. **Conclusion**

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Outline

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Participants and time frame

- Dragan Ivanovic, University of Novi Sad, Serbia
- Host: Hanna-Mari Puuska, CSC – IT Center for Science Ltd, Espoo
- Also involved Joonas Nikkanen, Lauri Hellsten and Olli Eskola
- First two weeks of September 2018

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Aim and activities

- Aim: to investigate the potential use of international metadata standards, CERIF in particular in harvesting and exporting publication data in the frame of the VIRTA system
- Activities:
    - Matching of VIRTA data model entities' properties to CERIF entities' properties
    - Implementation of OpenAIRE guidelines for CRIS managers

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Result

- T-SQL procedure for mappings CERIF data to VIRTA data model
- Extension of VIRTA OAI-PMH server side implementation in accordance with OpenAIRE guidelines

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Outline

Introduction
Methodology
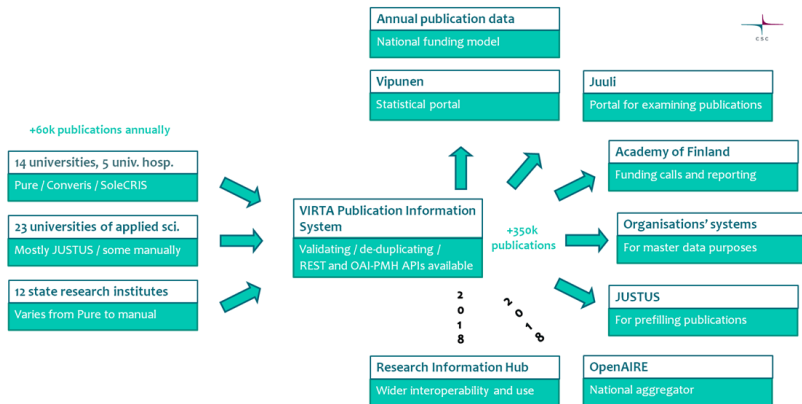Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Collection

- Collects bibliographic information on scientific publications from 54 Finnish research organizations
- About 60,000 publications per year are harvested, and all scientific fields are covered
- Publication metadata is based on a national data model

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Usage

- For the Finnish universities, 13% of core funding (200 mill. euros) is allocated via publication points calculated based on publication types and their level
- The bibliographic data are published at JUULI (www.juuli.fi) and statistics at Vipunen (www.vipunen.fi/en-gb/)
- The queries are available in a machine readable format via authenticated REST API (XML, JSON) and OAI-PMH API (Dublin Core, XML)

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

VIRTA interfaces

Annual publication data
National funding model

Vipunen
Statistical portal

Juuli
Portal for examining publications

+60k publications annually

14 universities, 5 univ. hosp.
Pure / Converis / SoleCRIS

23 universities of applied sci.
Mostly JUSTUS / some manually

12 state research institutes
Varies from Pure to manual

VIRTA Publication Information System
Validating / de-duplicating / REST and OAI-PMH APIs available

+350k publications

Academy of Finland
Funding calls and reporting

Organisations' systems
For master data purposes

JUSTUS
For prefilling publications

Research Information Hub
Wider interoperability and use

OpenAIRE
National aggregator

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

ENRESSH-VIRTA vs VIRTA

- Same software architecture
- Same data model
- SSH Publications' metadata collected from repositories all over the Europe vs publications' metadata collected from Finnish repositories

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Outline

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

Mission

- The mission of the H2020 project OpenAIRE is to shift scholarly communication towards openness and transparency and facilitate innovative ways to communicate and monitor research
- OpenAIRE project is a platform (https://www.openaire.eu/) representing a hub of scientific results collected from research infrastructure nodes across the Europe

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

OpenAIRE and Finnish publications

- By 2018, the OpenAIRE platform harvests Finnish publications' metadata from 9 repositories
- 5 universities' repositories as well as most of the research institutions in Finland are missing in OpenAIRE
- This is due to cumbersome implementation needed on repository-level to be compliant with OpenAIRE specifications
- In addition, repositories rarely include any metadata on scientific publications that are not self-archived
- For these reasons a major share of Finnish publications' metadata is not transferred to OpenAIRE.

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

OpenAIRE Guidelines for CRIS managers

- Version 1.1 was released in June 2018
- The Guidelines provide orientation for CRIS managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure
- Protocol for collecting data according to this guidelines is the OAI-PMH protocol, while the format for collecting data is a CERIF profile

Introduction
Methodology
Implementation
Conclusion

STSM
VIRTA
OpenAIRE

OpenAIRE Guidelines for CRIS managers - implementations

- METIS2OpenAIRE is an OpenAIRE-funded project led by euroCRIS
- Omega-PSIR, PURE, VIRTA, NARCIS, and CRIS UNS also working on the implementation
- DSpace-CRIS planning to implement

Outline

Mapping VIRTA data model to OpenAIRE (CERIF) data model

- There are many similarities between VIRTA and CERIF data models
- However, there are some key differences which should be addressed to properly implement the mapping
  - different classification of publication types,
  - presentation of IDs,
  - open access classifications and
  - Finnish national classifications of scientific fields

Making procedure for converting data from VIRTA to CERIF-XML

- The new T-SQL procedures which implements mappings from the previous point should be implemented
- Those procedures should be run each night to update CERIF-XML data table

Extension of the VIRTA OAI-PMH endpoint in accordance with the Guidelines

- OAI-PMH was already implemented in VIRTA for providing metadata in both Dublin Core and VIRTA-XML formats
- This implementation is used as basis for implementing OpenAIRE specifications
- However, this implementation should be extended with support for
  - new metadata prefix - oai_cerif_openaire,
  - the OpenAIRE supported sets: openaire_cris_publications, openaire_cris_persons, openaire_cris_events, openaire_cris_orgUnits
  - the new format - CERIF XML

VIRTA OAI-PMH
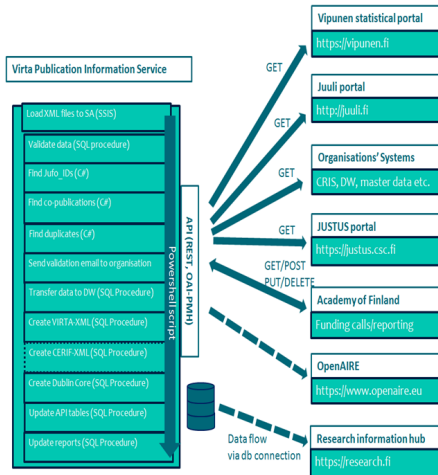
Making agreement with organizations on what information they want to be made available for harvesting

- VIRTA only stores a copy of publication metadata and research organizations act as owners of the metadata
- As such, permissions are needed for metadata to be allowed to external services / use
- Organizations need to decide if
    - OpenAIRE can harvest their data
    - Which publication years should be included in the harvest
    - The other limitations for the harvest (e.g. publication types)

Outline

Mapping VIRTA data model to CERIF data model

- Mapping has been defined - https://wiki.eduuni.fi/pages/viewpage.action?pageId=80941717
- Classification of publications types prescribed by Ministry of Education and Culture, Finland are mapped to classification prescribed by OpenAIRE
- Local identifiers and generated identifiers based on ISSN, ISBN, ORCID, etc.
- Finnish Thesaurus and Ontology Service has been used for definition of vocabulary of Finnish national scientific fields

Making procedure for converting data from VIRTA to CERIF-XML

Script name	Number of lines	Purpose of script
JulkaisutCERIF_Luonti	970	Transfer publications from VIRTA to CERIF-XML publication entity
JulkaisutCERIF_ISSN_Luonti	870	Transfer parent publications with ISSN ID from VIRTA to CERIF-XML publication entity
JulkaisutCERIF_ISBN_Luonti	960	Transfer parent publications with ISBN ID from VIRTA to CERIF-XML publication entity
OrganisaatiotCERIF_Luonti	610	Transfer organisations from VIRTA to CERIF-XML organisation entity
TapahtumatCERIF_Luonti	260	Transfer events from VIRTA to CERIF-XML event entity
TekijatCERIF_Luonti	460	Transfer persons from VIRTA to CERIF-XML person entity

Extension of the VIRTA OAI-PMH endpoint in accordance with the Guidelines

- The OAI-PMH endpoint has been extended
- The testing and validation has been performed using OpenAIRE CRIS-validator - https://github.com/jdvorak001/openaire-cris-validator
- Test endpoint is available at https://dwitjutife1.csc.fi/api/cerif

Making agreement with organizations on what information they want to be made available for harvesting

- Written permissions from organizations on data harvest are being collected
- The T-SQL script will convert data to CERIF-XML as agreements come and thus make available publications via OAI-PMH endpoint

Outline

Summary

- The vision of OpenAIRE project is to allow citizens, educators, funders, civil servants and industry find ways to make science useful for themselves, their working environments, and the society

- The presented implementation of OpenAIRE Guidelines for CRIS managers can be adjusted for the needs of exporting data from CRIS systems all over the Europe to OpenAIRE platform

Summary

- The vision of OpenAIRE project is to allow citizens, educators, funders, civil servants and industry find ways to make science useful for themselves, their working environments, and the society
- The presented implementation of OpenAIRE Guidelines for CRIS managers can be adjusted for the needs of exporting data from CRIS systems all over the Europe to OpenAIRE platform

Future work

- We are in contact with OpenAIRE software developers and working on integration of VIRTA and OpenAIRE
- OpenAIRE Guidelines for CRIS managers will be slightly modified (version 1.1.1)
- A conference paper describing results of the STSM accepted for presentation (ICIST 2019), the full paper should be written and published in proceedings
- STSM in Novi Sad

Application of similar approach to ENRESSH-VIRTA

- ENRESSH-VIRTA is instance of VIRTA system with different data
- Implementation of VIRTA mapping and extension of VIRTA OAI-PMH server side can be adopted
- Possible issues
    - publication types - should be mapped to OpenAIRE publication types
    - scientific fields - machine readable vocabulary should be created (rdf)
- The European centralized solutions (ENRESSH-VIRTA) can provide information on research in a consistent format so that each organization does not need to build its own reporting solution to external systems (OpenAIRE)

Questions

- Thank you for your attention!!!
- If you have any questions, please do not hesitate to
  - ask me now
  - ask me during the conference breaks
  - contact me via email address dragan.ivanovic@uns.ac.rs