OpenAIRE Guidelines for CRIS Managers

Supporting Interoperability of Open Research Information through established standards
Outline

• OpenAIRE CRIS Guidelines overview and scope
• Different aspects of interoperability support
• Data exchange based on CERIF-XML
• Semantics
• Harvesting protocol for the transmission of CERIF-XML
• Status and next steps
OpenAIRE CRIS Guidelines overview

- Guidelines for CRIS managers to expose information to OpenAIRE
- Great benefits for CRIS managers: reuse – dissemination of their information by the OpenAIRE infrastructure
- OpenAIRE can be considered a CRIS system – internal OpenAIRE data model is CERIF compliant
- OpenAIRE interoperability with CRIS systems: an example of point-to-point data exchange between CRIS systems
Different aspects of interoperability

• Structure and semantics
  • Structure: data model (CERIF subset)
  • Semantics: vocabularies and terms used for classification and definition of relationships

• System and syntax
  • Access / harvesting protocol (OAI-PMH)
  • Marshalling CERIF XML as OAI-PMH payload
Interoperability in structure

- Define the subset of CERIF relevant for OpenAIRE
- Major part of this work performed during definition of the CERIF-compliant OpenAIRE internal data model
Semantic interoperability

• The meaning and scope of each CERIF entity within OpenAIRE
• Terms and vocabularies used for classifications and definition of relationships – reuse standard CERIF Semantics vocabularies with certain additions
• Ensured consistency with OpenAIRE Guidelines for literature and data repositories (e.g. dataset types, open access types)
Syntactic interoperability

- Reuse CERIF XML

- OpenAIRE CERIF XML Schema with distinct namespace
  - Strict subset of the canonical CERIF XML (data elements and constraints)
  - Distinct namespace (OpenAIRE domain)
  - Nesting only of multi-lingual attributes, federated identifiers and linked entities
  - Referential integrity constraints apply
System interoperability

• Harvesting protocol: OAI-PMH
• Different OAI-PMH sets need to be made available by the data provider
  • One for each entity type (e.g. get all records of cfProject)
  • One for the entire CERIF XML graph
• Relaxed rules about selective harvesting through time stamps and deleted records
Guidelines material

- Main guidelines document
  - Includes definition of allowed vocabularies and terms for each classification and relationship type
- OpenAIRE CERIF XML Schema
- Examples
  - CRIS data expressed in OpenAIRE CERIF XML
  - CERIF XML as OAI-PMH payload
- Material available at the OpenAIRE Guidelines Wiki
  https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS
Process - methodology

• Ensured early involvement of stakeholders and communities
  • OpenAIRE (e.g. technical infrastructure developers, guidelines and community support teams)
  • euroCRIS community – CERIF experts, developers / vendors of CRIS systems
• Early internal review within OpenAIRE
• Review by the CERIF experts and CRIS developers (autumn 2013)
  • Valuable feedback collected and incorporated into guidelines
Status and next steps

• Guidelines went through public review (review period ended 28 March 2014)
  • No major review comments
• Post-review version to be released within May 2014
Thank you!

• Acknowledgements
  • The presented work was partly supported by OpenAIREplus Project (Ref No: 283595) of the European Union FP7-INFRASTRUCTURES Programme.
  • Jan Dvořák’s work on this article was partly supported by the Ministry of Education, Youth and Sports of the Czech Republic through grant no. LG14007.
  • The authors wish to acknowledge the valuable feedback provided by the reviewers of the OpenAIRE guidelines for CRIS managers (the reviewers’ list is available at the OpenAIRE guidelines wiki).