CERIF XML update

Jan Dvořák
Andrea Bollini
Thomas Vestdam
Outline

• The Goal
• Motivation
• Solution
• Tools and processes
• When?
CERIF XML update

• A novel approach to constructing the CERIF XML exchange format

• Goals:
  – Easier to understand & use
  – More in line with general expectations
Motivation

Current CERIF XML perceived as:

– Not easy to embed in documents
– Fragmented
– Unsuitable for processing of longer streams
– Not attractive for developers

Vestdam T., Plauborg B., Van Campe L.:
CERIF XML in large scale exchange of research information. CRIS 2016

http://dspacecris.eurocris.org/handle/11366/495
Real world CERIF usage

• Anyone needs the full CERIF at once (in one CRIS)?
  – Very unlikely
• Everybody needs to support specific information interchange scenario(s)
  – Specific subsets of CERIF
  – Profiles
A CERIF Profile

• Entities & attributes:
  – Profile $\subseteq$ CERIF

• Semantic vocabularies:
  – Profile – more specific
  – Sources: CERIF & beyond

• Integrity constraints:
  – Profile $\supseteq$ CERIF

Profile data is CERIF

Producers know what to include
Consumers know what to expect
Example of a CERIF Profile:
The OpenAIRE Guidelines for CRIS Managers

https://doi.org/10.5281/zenodo.17065
CERIF Profile Requirements

• Easy to construct and share
• Effective to use
• Remain CERIF
  – through the links to the CERIF Data Model
CERIF Profile Artefacts

• Description
• Examples
• XML Schema
  • with CERIF annotations
Example
<FundingAwardNotice xmlns="urn:xmlns.nl:pop-project:certif-profile-1.0-EXPERIMENTAL" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Project id="13045b7f-ceb0-4a3f-a82d-82c19e1994e7">
    <Title xml:lang="en">Televising Locality: Industrial, historical and socio-cultural investigations of local television</Title>
    <Identifier type="http://openaire.eu/certif/vocab/IdentifierTypes#ReferenceNumber">OND1336115</Identifier>
    <StartDate>2010-01-01</StartDate>
    <EndDate>2015-11-30</EndDate>
    <Funders>
      <Funder>
        <OrgUnit id="87a6333e-0943-4669-a3a4-151826c45c97">
          <Acronym>NWO</Acronym>
          <Name xml:lang="en">Netherlands Organisation for Scientific Research</Name>
          <Name xml:lang="nl">Nederlandse Organisatie voor Wetenschappelijk Onderzoek</Name>
          <FundRefId>http://dx.doi.org/10.13039/501100003246</FundRefId>
        </OrgUnit>
      </Funder>
    </Funders>
    <Consortium>
      <Secretariat>
        <OrgUnit id="e77efdc0-b5f7-4851-8273-ce0fbebfc4938">
            <Acronym>ICON</Acronym>
            <Name xml:lang="en">Institute for Cultural Inquiry</Name>
            <Name xml:lang="nl">Instituut voor Cultuurwetenschappelijk Onderzoek</Name>
            <PartOf>NA</PartOf>
        </OrgUnit>
      </Secretariat>
      <Collaboration>
        <OrgUnit id="7276968b-cb6e-4e41-bb86-61bf5054b00f">
            <Acronym>HUIZINGA</Acronym>
            <Name xml:lang="en">Huizinga Institute - The Netherlands Graduate School for Cultural History</Name>
            <Name xml:lang="nl">Huizinga Instituut Interuniversitaire Onderzoekschool voor Cultuurgeschiedenis</Name>
            <PartOf>NA</PartOf>
        </OrgUnit>
      </Collaboration>
    </Consortium>
    <Team>
      <Supervisor>
        <Person id="5bdbdd3e-369a-41ba-8ab4-e35f55a58d3c">
          <PersonName>
            <FamilyNames>Arps</FamilyNames>
            <FirstName>Bern</FirstName>
            <LastName>Arps</LastName>
            <Email>aparps@uu.nl</Email>
            <Phone>+31-318-483524</Phone>
            <Fax>0-318-483524</Fax>
            <Address>Utrecht University, Netherlands</Address>
          </PersonName>
        </Person>
      </Supervisor>
    </Team>
  </Project>
</FundingAwardNotice>
CERIF Profile Workflow

CERIF Semantic Vocabularies

CERIF Profile Definition
specifies:
- which CERIF entities and attributes are used
- which semantic vocabularies are used

CERIF Profile Processor
- filters out the entities/attributes that are not used
- plugs in the semantic vocabularies that are used
- expands certain syntactic constructs

CERIF Template XML Schema
- written by hand
- has the syntactic structures
- no namespace (to be always included)
- mostly agnostic of the semantic vocabularies
- just some constructs represent specific roles or types

CERIF Profile XML Schema
- generated
- in the specific XML namespace of the profile
- additional restrictions:
- enumerations for the semantic vocabulary terms

CERIF Data Model

Input

References

Full correspondence required and checked

Output
Benefits

• For users:
  – Readable
  – Understandable
  – Closer to the domain

• For developers:
  – Tool-friendly
  – Streamable
CERIF development update

More like an open source software project
With feature requests, releases, ...

Using proper infrastructure:
  GitHub, Wiki, Issue tracker

Using structured governance and roles:
  stable members, contributors, etc.
Next Steps

Applications:
• Pilot in NL
• Snowball
• OpenAIRE

CERIF TG:
• Work out the tooling
• Implement the new working mode

Architecture TG:
• Review CERIF API to support profiles
When?

Help by joining the TG!

Embrace the updated CERIF XML for new projects
The Future

Several CERIF profiles

All of them derive from the same model

Used
  by many parties
  around the world
  in many research information interchange scenarios