CERIF Tutorial

Jan Dvořák
June 13th, 2018
CRIS 2018 conference
Umeå, Sweden
Jan Dvořák
jan.dvorak@ff.cuni.cz, https://orcid.org/0000-0001-8985-152X

euroCRIS
• CERIF TG Leader since 2013
• CRIS 2012 (Prague, June 2012), Organization Committee Chair
• Membership meeting Autumn 2010, Organizer

Charles University in Prague, Faculty of Arts, Institute of Information Studies & Librarianship
• Researcher & Lecturer

Czech Technical University, Computing and Information Centre
• IS Analyst: EZOP+V3S – institutional CRIS

InfoScience Praha
• Research, Development & Innovation Information System

---

This deck of slides is based on the CERIF Tutorial by Brigitte Jörg
CERIF TG Leader 2004-2012
What is Research Information?

The process of research
- Research projects
- Funding
- Research infrastructures

The research actors
- Researchers
- Institutions
- Funders
- Publishers
- Facility operators
- Associations

Research results
- Outputs (Publications, Research Datasets, Patents, …)
- Outcomes, Impacts
Who needs Research Information?

- **Researchers**: visibility, finding collaborations, competitors, CV generation
- **Decision Makers**: performance, strategic decisions, priorities, comparisons
- **Project Managers**: overview of ongoing activities
- **Libraries**: acquisition, dissemination
- **Publishers**: finding reviewers, editors
- **Educators**: integration of relevant findings into lectures and training
- **General Public**: information and education, interest
- **Enterprises**: finding information for participation in projects, partnerships, usage of results
- **Intermediaries / Brokers**: finding research results of potential market or innovative value
- **Research Organisations**: integration and interoperability, strategic management
- **Funding Organisations**: distribution of programs, evaluation of results, finding reviewers
Common European Research Information Format

- CERIF is an EU Recommendation to Member States

- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage

http://cordis.europa.eu/cerif/
Model Levels

- **Conceptual Level** (Specification)
  Concepts relevant for the research domain and their relationships

- **Logical Level** (ER Model)
  Entities and their relationships

- **Semantic Layer** (Declared Semantics)
  A formalized controlled vocabulary describing a general contextual semantics of the research domain inline with the conceptual, logical and machine description
Common European Research Information Format
CERIF Base Entities

- Project
- Person
- OrganisationUnit

Relationships:
- Person -> Project
- OrganisationUnit -> Project
- Project -> Person
- Project -> OrganisationUnit
CERIF Base Entities

Person
ID
URI
Gender
FirstNames
OtherNames
FamilyNames
NameVariants
ResearchInterest
Keywords

Project
ID
URI
Acronym
StartDate
EndDate
Title
Abstract
Keywords

OrganisationUnit
ID
URI
Acronym
Name
HeadCount
CurrencyCode
Turnover
ResearchActivity
Keywords
CERIF Result Entities

- ResultPublication
- ResultPatent
- ResultProduct
CERIF Result Entities

- ResultPublication
  - ID
  - URI
  - Title
  - Subtitle
  - Abstract
  - Bibl. Note
  - PublicationDate
  - TotalPages
  - StartPage
  - EndPage
  - Keywords

- ResultPatent
  - ID
  - URI
  - PatentNumber
  - Title
  - CountryCode
  - RegistrationDate
  - ApprovalDate
  - Description
  - Keywords

- ResultProduct
  - ID
  - URI
CERIF Result Entities

**cfResultPublication**
- cfID
- cfURI
- cfNumber
- cfPublicationDate
- cfStartPage
- cfEndPage
- cfTotalPages
- cfEdition
- cfSeries
- cfIssue
- cfVolume
- cfISBN
- cfISSN

**cfResultPatent**
- cfID
- cfURI
- cfPatentNumber
- cfCountryCode
- cfRegistrationDate
- cfApprovalDate

**cfResultProduct**
- cfID
- cfURI

**cfBibliographic**
- Note

**cfVersionInfo**

**cfAbbreviation**

**cfDescription**

**cfKeywords**
CERIF Infrastructure Entities

- Facility
- Equipment
- Service
A typical CERIF entity has:

- **Identifier (internal)**
- **Attributes**
  - the basic ones
  - the multi-lingual ones
- **External Identifiers**
- **Classifications**
  - Type
  - Status
  - Subject area
- **Links**
  - to other entities
  - recursive
CERIF 1.6
Some CERIF Link Entities

- Person_ResultPublication
  - role=author
- Person_OrganisationUnit
  - role=research assistant
- Person_Project
  - role=principal investigator
- Project_ResultPublication
  - role=deliverable
- OrganisationUnit_ResultPublication
  - role=author's affiliation
- Project_OrganisationUnit
  - role=coordinator
- Person_OrganisationUnit

CERIF Link Entities:
- Person
- OrganisationUnit
- Project
- ResultPublication
- ResultPatent
- Person_OrganisationUnit
- Project_OrganisationUnit
- Person_Project
- OrganisationUnit_ResultPublication
- Project_ResultPublication
- Project_OrganisationUnit
- Person_OrganisationUnit
- Person_ResultPublication
- Person_Project
Result_Publication Instance Diagram
(slide by Keith Jeffery)

- **Person A**: project leader
- **OrgUnit M**: member
- **OrgUnit O**: employee
- **OrgUnit N**: member
- **OrgUnit N**: partner
- **Project P**: deliverable
- **Publication X**: author
- **Publication X**: owns IPR
- **Measurement Z**: has associated

The diagram illustrates the relationships and roles within an organization.
Generic Linking Entity Structure

- **Base object 1** (FK)
- **Base object 2** (FK)

  - **cfStartDate**
  - **cfEndDate**
  - **Time range of validity**

  - **cfFraction**
    - **Fraction (optional)**

  - **role : cfClassification** (FK)
Example: The Principal Investigator of project $P$ changes: $X$ is replaced by $Y$ effective date $D$.

<table>
<thead>
<tr>
<th>Before:</th>
<th>Validity range</th>
<th>Role</th>
<th>After:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>$(-\infty .. +\infty)$</td>
<td>Principal Investigator: cfClassification</td>
<td>$X$</td>
</tr>
<tr>
<td>$P$</td>
<td>$(-\infty .. D)$</td>
<td>Principal Investigator: cfClassification</td>
<td>$X$</td>
</tr>
<tr>
<td></td>
<td>$D .. +\infty$</td>
<td>Principal Investigator: cfClassification</td>
<td>$Y$</td>
</tr>
</tbody>
</table>
CERIF 1.6
Measuring Impact in CERIF (MICE)

MICE, a JISC-funded Project coordinated by Richard Gartner, Kings College, London, UK
Measurement & Indicator (some examples)

– economic and commercial

  • economic
    – impact on business
      » improving performance of existing businesses
        • increased turnover by 1.2M€ in 2012
        • time savings of 14.56%
        • reduced costs by 42%
    » new products/processes
      • creating numbers of new products/services
      • commercialising / other success measures

Extract from the MICE List of Indicators
CERIF 1.6
CERIF Semantic Layer

Allows to capture any Schema or Structure
- Flat Lists
- Thesauri
- Classification Systems (e.g. SKOS, ...)
- Taxonomies
- Ontologies

Open / Extensible in all directions
- New Schemas
- New Concepts / Terms
- New Relationships

Enables to manage
- Roles / Types Semantics
- Subject Headings
- Archiving (Time component)

Allows for Mappings between Schemes
CERIF Semantic Layer (Declared Semantics)

Recursion

Is-a
Maps-to
Is-part-of
Is-broader-term

Scheme-Assignment
Time-based
CERIF Federated Identifiers

- **ResultPublication**
  - ISBN
  - ISSN
  - DOI
  - WoS Accession Number
  - Scopus EID
  - PubMed Central ID
- **Person**
  - Social Security Number
  - Staff Id in HR system
  - Author identifier
    - ORCID, IdRef, DAI, ResearcherID, ScopusID
- **Project/Grant**
  - Funder’s reference number
  - Organisation’s reference number
- **Organisation**
  - VAT Identification Number
  - FundRefID
  - GridID
- **Classification**
  - External Code
CERIF Federated Identifiers

- Records the “tag” by which an object is known elsewhere
- For any Base, Result, Infrastructure, or 2\textsuperscript{nd} Level entity
- “Identifier Types” classification scheme
- (optionally) Connected to a Service representing the issuer of the identifier
  - Usually an information system
CERIF XML Interchange Format

- XML Schema
- Based on the ER model
- Undergone a big update
CERIF Profiles

- Entities & attributes:
  - Profile $\subseteq$ CERIF

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

- Integrity constraints:
  - Profile $\supset$ CERIF

Profile data is CERIF

Producers know what to include

Consumers know what to expect
Subset of CERIF entities
Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard
<Author>
    <Person>
        <PersonName>
            <FamilyNames>Singhal</FamilyNames>
            <FirstNames>Sonal</FirstNames>
        </PersonName>
    </Person>
    <Affiliation>
        <OrgUnit id="301248">
            <Name xml:lang="en">Museum of Vertebrate Zoology</Name>
            <PartOf>
                <OrgUnit id="329384">
                    <Name xml:lang="en">University of California, Berkeley</Name>
                </OrgUnit>
            </PartOf>
        </OrgUnit>
    </Affiliation>
</Author>
Phenotypically cryptic lineages comprise an important yet understudied part of biodiversity;
CERIF development

By the CERIF Task Group of euroCRIS

Adopting open-source software projects tools & best practices:

→ https://github.com/EuroCRIS/CERIF-DataModel
→ CC BY license

Two branches:
- master: latest official release (1.6.1)
- develop: on-going development
Basic Information

<table>
<thead>
<tr>
<th>Project</th>
<th>CERIF: the Common European Research Information Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>CERIF Data Model</td>
</tr>
<tr>
<td>Version</td>
<td>CERIF current development DM</td>
</tr>
<tr>
<td>Company</td>
<td>euroCRIS</td>
</tr>
<tr>
<td>Author</td>
<td>CERIF Task Group</td>
</tr>
<tr>
<td>Created</td>
<td>30.6.2014</td>
</tr>
<tr>
<td>Last Modified</td>
<td>22.12.2017</td>
</tr>
</tbody>
</table>

Statistic Information

<table>
<thead>
<tr>
<th>Entities</th>
<th>295</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>1829</td>
</tr>
<tr>
<td>Domains</td>
<td>1</td>
</tr>
<tr>
<td>User Data Types</td>
<td>0</td>
</tr>
<tr>
<td>Relationships</td>
<td>671</td>
</tr>
<tr>
<td>Indexes</td>
<td>0</td>
</tr>
<tr>
<td>Keys</td>
<td>298</td>
</tr>
<tr>
<td>Views</td>
<td>0</td>
</tr>
<tr>
<td>View Relationships</td>
<td>0</td>
</tr>
<tr>
<td>Procedures</td>
<td>0</td>
</tr>
<tr>
<td>Defaults</td>
<td>0</td>
</tr>
<tr>
<td>Schemas</td>
<td>0</td>
</tr>
<tr>
<td>Users</td>
<td>0</td>
</tr>
<tr>
<td>User Groups</td>
<td>0</td>
</tr>
<tr>
<td>Notes</td>
<td>0</td>
</tr>
</tbody>
</table>
CERIF highlights

• Right level of abstraction
• Normalized model
  – Record information only once
  – Reference rather than copy
• Versatile Semantic Layer
• Time-based relationships
• Clean design, regular structure
Metadata Layers

Discovery metadata
DC, VIVO, MODS, METS, eGMS, DCAT, ...

Contextual metadata
CERIF

Detailed metadata
Domain-specific standards

Generate
Reference