CERIF 1.6 Tutorial
Jan Dvořák
jan.dvorak@infoscience.cz

euroCRIS
• CERIF TG Leader since 2013
• CRIS2012 (Prague, June 2012) Organizer

Charles University in Prague
• Faculty of Arts
  – Institute of Information Studies & Librarianship

InfoScience Praha
• Research & Development & Innovation Information System
  (the national CRIS for CZ – www.isvav.cz)

Most slides by Brigitte Jörg
What is Research Information?

Information about:
- Researchers
- Organisations
  - Research performing orgs, Funders, Publishers, Facility Operators
- Scientific Disciplines
- Funding
  - Funding Programmes, Calls
- Projects
  - Proposed, Ongoing, Completed
- Research infrastructures
  - Facilities, Equipment, Services
- Outputs
  - Publications, Patents, Research Data, Research Software, Products
- Outcomes
  - New product on the market, Improved treatment procedure, Regulation update
- Impacts
  - Increased market share, Reduced death rate of a disease
- And their Relationships
Kinds of questions we want to support

- How many articles has author X published in 2012 as a first author?
- How many times have articles by author X been cited by the end of the previous year?
- Did author X publish with institutionally external authors?
- In how many FP7 projects does/did organisation Z participate?
- How many publications have resulted from project Y?
- How many people have been employed in the course of FP7 projects from the 1st call in the New Member States?
- How many PhD students have participated in national research projects in country C? In which countries have they earned their masters degrees?
- How many women have been involved in FP7 projects?
- How often have articles in journal A been requested in 2012?
- How many articles have been published in field B?
The Ultimate Answer: Common European Research Information Format

- Person
  - Skills
  - CV
- Project
- Publication
- Patent
- Product
- Event
- Funding
- Organisation
- Service
- Equipment
- Classification (Semantics)
The CERIF Evolution

Similar Ideas
UN/UNESCO
OECD
CODATA

EU Working Group on Research Databases Workshop

CERIF 91

- Networking of DBs
- Exchange of Records
- EC Recommendation to Member States

Acronym: ERGO
Participant:
Keith Jeffery, Anne Asseron, many more
Organisations:
Rutherford Appleton, University of Bergen, ...

CERIF 2000 Model

- Data Model
- Multilinguality
- Controlled Vocabulary
- Roles / Types
- User-driven
- EC Recommendation to Member States

CERIF 2006 / 2008 Model

- Data Model
- Model Normalization
- Robust/Consistent Structure
- Extensible Structure
- Semantic Layer
- XML Exchange Specification
- Elaboration on Publication
- CERIF Core Semantics (2008 1.2)

CERIF 2000

- Network of DBs
- Exchange of Records
- EC Recommendation to Member States

CERIF 1.6
CERIF 1.5
CERIF 1.4 (XML)
CERIF 1.3

FORMAL SEMANTICS

FEDERATED IDENTIFIERS
08.09.13

EUROCRIS 2000

Person
Organisation
Person
Organisation

2000

CERIF 1.6
CERIF 1.5
CERIF 1.4 (XML)
CERIF 1.3

FORMAL SEMANTICS

FEDERATED IDENTIFIERS
08.09.13

EUROCRIS 2000

Person
Organisation
Person
Organisation

2000
Common European Research Information Format

- CERIF is an EU Recommendation to Member States
  http://cordis.europa.eu/cerif/

- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage
  http://www.eurocris.org/Index.php?page=CERIFreleases&t=1
Model Levels

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

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

- **Physical Level** (Database Scripts)
  Data Definition commands for the database

- **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

---

SQL Script

```
CREATE Table cfPers (...);
CREATE Table cfProj (...);
CREATE Table cfOrgUnit (...);
```
CERIF Base Entities

- Person
- OrganisationUnit
- Project
CERIF Base Entities

- **Person**
  - ID
  - URI
  - Gender
  - FirstNames
  - OtherNames
  - FamilyNames
  - NameVariants
  - ResearchInterest
  - Keywords

- **OrganisationUnit**
  - ID
  - URI
  - Acronym
  - Name
  - HeadCount
  - CurrencyCode
  - Turnover
  - ResearchActivity
  - Keywords

- **Project**
  - ID
  - URI
  - Acronym
  - StartDate
  - EndDate
  - Title
  - Abstract
  - Keywords

**Relationships**
- Person → Project
- Person → OrganisationUnit
- Project → OrganisationUnit
- Project → Person
CERIF Result Entities

ResultPublication

ResultPatent

ResultProduct
CERIF Result Entities

ResultProduct
- ID
- URI

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

ResultPatent
- ID
- URI
- PatentNumber
- Title
- CountryCode
- RegistrationDate
- ApprovalDate
- Description
- Keywords
CERIF Infrastructure Entities

- Facility
- Equipment
- Service
CERIF Infrastructure Entities

cfName

cfDescription

cfKeywords

cfFacility

cfID

cfURI

cfAcronym

cfService

cfID

cfURI

cfAcronym

cfEquipment

cfID

cfURI

cfAcronym
Measuring Impact in CERIF (MICE)

MICE, a JISC-funded Project coordinated by Richard Gartner, Kings College, London, UK
CERIF Measurement & Indicator

Is an Aggregation Entity

cfMeasureIdentifier

cfCountInteger

cfCountIntegerChange

cfValueFloatingPoint

cfCountFloatingPointChange

cfValueJudgementalNumeric

cfValueJudgementalNumericChange

cfValueJudgementalText

cfValueJudgementalTextChange

cfURI
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
Some CERIF Link Entities
Some CERIF Link Entities

- Person
  - Person_OrganisationUnit
    - role=principal investigator
  - Person_Project
    - role=principal investigator
  - Person_ResultPublication
    - role=author
  - Person_Project
    - role=principal investigator

- OrganisationUnit
  - OrganisationUnit_ResultPublication
    - role=author's affiliation
  - OrganisationUnit_OrganisationUnit

- Project
  - Project_OrganisationUnit
    - role=coordinator
  - Project_ResultPublication
    - role=deliverable

- ResultPublication
  - Person_ResultPublication
    - role=author
  - Project_ResultPublication
    - role=deliverable
Result_Publication Instance Diagram
(slide by Keith Jeffery)
Some CERIF Link Entities

### Binary classifications:
- Role

<table>
<thead>
<tr>
<th>Table</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfPers_Pers</td>
<td>cfPersId1 ID NN (PK), cfPersId2 ID NN (PK), cfClassId ID NN (PK), cfClassSchemId ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK)</td>
</tr>
<tr>
<td>cfProj_OrgUnit</td>
<td>cfProjId ID NN (PK), cfOrgUnitId ID NN (PK), cfClassId ID NN (PK), cfClassSchemId ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK)</td>
</tr>
<tr>
<td>cfOrgUnit_Equip</td>
<td>cfOrgUnitId ID NN (PK), cfEquipld ID NN (PK), cfClassId ID NN (PK), cfClassSchemId ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK), cfCurrCode Char(3) (FK), cfAvailability Char(64), cfConditions Char(255), cfPrice Float</td>
</tr>
<tr>
<td>cfOrgUnit_Class</td>
<td>cfOrgUnitId ID NN (PK), cfClassId ID NN (PK), cfClassSchemId ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK)</td>
</tr>
<tr>
<td>cfClass_Class</td>
<td>cfClassId1 ID NN (PK), cfClassId2 ID NN (PK), cfClassSchemId1 ID NN (PK), cfClassSchemId2 ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK)</td>
</tr>
<tr>
<td>cfProj_Pers</td>
<td>cfProjId ID NN (PK), cfPersId ID NN (PK), cfClassId ID NN (PK), cfClassSchemId ID NN (PK), cfFraction Float, cfStartDate Timestamp(6) NN (PK), cfEndDate Timestamp(6) NN (PK)</td>
</tr>
</tbody>
</table>

### Unary classification:
- Type
- Status
- Subject area
Generic Linking Entity Structure

Base object 1 (FK)

- cfStartDate
- cfEndDate

Interval of validity

Fraction (optional)

Base object 2 (FK)

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

Before:

```
P --> Principal Investigator : cfClassification --> X
```

After:

```
P --> Principal Investigator : cfClassification --> Y
P --> Principal Investigator : cfClassification --> X
```

D .. +∞
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
Recursion

is-a
maps-to
is-part-of
is-broader-term

Scheme-Assignment

Time-based
CERIF 1.6

Diagram showing relationships between different CRIS elements such as:
- cfFunding
- cfEquipment
- cfFacility
- cfService
- cfResultPublication
- cfResultProduct
- cfProject
- cfOrganisationUnit
- cfPerson
- cfIndicator
- cfMeasurement
- cfPostalAddress
- cfGeographicBoundingBox
- cfCountry
- cfLanguage
- cfCurrency
- cfFederatedIdentifier
- cfPrize
- cfQualification
- cfCurriculumVitae
- cfCitation
- cfEvent
- cfExpertiseAndSkills
- cfLanguage
- cfCurrency
- cfCountry
- cfGeographicBoundingBox
CERIF Federated Identifiers

- **ResultPublication**
  - DOI
  - WoS Accession Number
  - ISBN
  - ISSN

- **Person**
  - Social Security Number
  - Staff Id in HR system
  - Author identifier
    - ORCID
    - ScopusID

- **Project/Grant**
  - Funder’s reference number

- **Organisation**
  - VAT Identification Number
  - Internal Code
  - FundId

- **Classification**
  - External Code

www.eurocris.org
CERIF Federated Identifiers

- Records the “tag” by which an object is known elsewhere
- For any Base, Result, Infrastructure, or 2\textsuperscript{nd} Level entity
- Federated Identifier Type classification scheme
- (optionally) Connected to a Service representing the issuer of the identifier
  - Usually an information system
- For point-to-point interchange
- XML namespace
- XML Schema
  - Based on the ER model

```
cfCitation
cfEvent
cfLanguage
cfCurrency
cfCountry
cfResultPublication
cfResultPublication
cfResultProduct
cfResultProduct
cfProject

cfEquipment
cfFunding
cfFacility
cfService
cfFedera
dentifier

cfQualification

cfPrize

cfCurriculumVitae

```

---

`www.eurocris.org`

---

**CERIF XML 1.6 Interchange Format**
<CERIF xmlns="urn:xmlns:org:eurocris:cerif-1.6-2">
  <cfProj>
    <cfProjId>internal-project-identifier</cfProjId>
    <cfAcro>ACRO</cfAcro>
    <cfURI>http://www.project-url.ac.uk/acro.html</cfURI>
    <cfTitle cfLangCode="en" cfTrans="o">The title of the project</cfTitle>
    <cfAbstr cfLangCode="en" cfTrans="o">The goals of the project</cfAbstr>
    <cfProj_Class>
      <cfClassId>infrastructure-project-uuid</cfClassId>
      <cfClassSchemeId>-project-types-scheme-uuid</cfClassSchemeId>
    </cfProj_Class>
    <cfFedId>PROJECT NUMBER</cfFedId>
    <cfClassId>project-number-uuid</cfClassId>
    <cfClassSchemeId>-federated-identifier-type-uuid</cfClassSchemeId>
  </cfFedId>
  <cfProj_OrgUnit>
    <cfOrgUnitId>orgunit-1-identifier</cfOrgUnitId>
    <cfClassId>coordinator-uuid</cfClassId>
    <cfClassSchemeId>orgunit-project-roles-scheme-uuid</cfClassSchemeId>
    <cfStartDate>from-datetime</cfStartDate>
    <cfEndDate>to-datetime</cfEndDate>
  </cfProj_OrgUnit>
</cfProj>
</CERIF>
CERIF development

By the CERIF Task Group of euroCRIS

Join euroCRIS
Come to the Task Group meeting
CERIF highlights

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