CERIF 1.5 Tutorial

November 13th, 2013

euroCRIS Membership Meeting
Porto, Portugal
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)

Most slides by Brigitte Jörg
What is Research Information?

Information about:

• Researchers
• Organisations (Research-performing, Funding)
• Funding Programmes, Calls
• Projects (Proposed, Ongoing, Completed)
• Publications, Patents, Data, Products
• Facilities, Equipment, Services
• Addresses, Geographic Bindings, Languages

• And their Relationships
Who needs Research Information?

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

Research Organisations
- integration and interoperability, strategic management

Research Information
Kinds of questions we want to support

- How many articles has author X published in 2011 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 2010?
- How many articles have been published in field B?
The Ultimate Answer: Common European Research Information Format

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

CERIF 91

- Project
- Person
- Organisation
- Results
- Equipment
- Classification

CERIF 2000 Model

- Roles
- Expertise
- OrgUnit
- Person
- Project
- Measurement
- Link
- Language
- Semantics
- Base
- 2ndLevel

CERIF 2006 / 2008 Model

- Data Model
- Multilinguality
- Controlled Vocabulary
- Roles / Types
- User-driven

- EC Recommendation to Member States

CERIF 1.3

- Data Model
- Infrastructure
- Facility, Equipment, Service
- Measurement & Indicator
- Entities and Link Tables
- Geographic Bounding Box
- CERIF 1.3 Vocabulary
- UUIs
- Terms
- Schemes
- CERIF 1.4 new XML format
- CERIF 1.5 Federated Identifiers

CERIF 1.4 (XML)
CERIF 1.5

Linked Data

Similar Ideas
UN/ UNESCO
OECD
CODATA

EU Working Group on Research Databases Workshop

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

- CERIF is an EU Recommendation to Member States

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

**CERIF Entity Types**
- Base Entities
- Result Entities
- Infrastructure Entities
- 2nd Level Entities
- Geographic Bounding Box
- Link Entities

**CERIF Features**
- Multiple Language
- Semantics
- Measures & Indicators
CERIF Base Entities

- **cfPerson**
  - cfID
  - cfURI
  - cfOtherNames
  - cfNameVariants
  - cfFirstName
  - cfBirthdate
  - cfKeywords
  - cfDescription

- **cfOrganisationUnit**
  - cfID
  - cfURI
  - cfAcronym
  - cfHeadCount
  - cfCurrencyCode
  - cfTurnover
  - cfName
  - cfDescription
  - cfKeywords

- **cfProject**
  - cfID
  - cfURI
  - cfAcronym
  - cfStartDate
  - cfEndDate
  - cfTitle
  - cfAbstract
  - cfKeywords
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

**Facility**
- ID
- Acronym
- URI
- Title
- Description
- Keywords

**Equipment**
- ID
- Acronym
- URI
- Title
- Description
- Keywords

**Service**
- ID
- Acronym
- URI
- Title
- Description
- Keywords
CERIF Infrastructure Entities

cfName

Facility
  Equipment
  Service

Service
  cfService
    cfID
    cfURI
    cfAcronym

Equipment
  cfEquipment
    cfID
    cfURI
    cfAcronym

Facility
  cfFacility
    cfID
    cfURI
    cfAcronym


CERIF 1.5

www.eurocris.org
Measuring Impact in CERIF (MICE)

MICE, a JISC-funded Project coordinated by Richard Gartner, Kings College, London, UK
– 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 – Generic Entity Structure

Generic

Identifier
URI
Attributes

Multilingual Entities
Relationships (Links)
Some CERIF Link Entities
Some CERIF Link Entities

- **Person**
  - **Person_ResultPublication**
    - role=author
  - **Person_OrganisationUnit**
    - role=research assistant
  - **Person_Project**
    - role=principal investigator

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

- **Project**
  - **Project_ResultPublication**
    - role=deliverable
  - **Project_OrganisationUnit**

- **ResultPublication**
  - **Person_ResultPublication**
  - **Project_ResultPublication**
  - **OrganisationUnit_ResultPublication**
### Some CERIF Link Entities

#### cfPers_Pers
- **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)

#### cfProj_OrgUnit
- **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)

#### cfOrgUnit_Equip
- **cfOrgUnitId**: ID, NN (PK)
- **cfEquipId**: 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) (PK)
- **cfAvailability**: Char(64)
- **cfConditions**: Char(255)
- **cfPrice**: Float

#### cfClass_Class
- **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)

#### cfProj_Pers
- **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)
Result_Publication Instance Diagram
(slides by Keith Jeffer y)

Person A
- Project leader

OrgUnit M
- member

OrgUnit O
- Part of

OrgUnit N
- member
- Part of

Project P
- member

Publication X
- author
- owns IPR
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 simple Mappings between Schemes
Recursion is-a maps-to is-part-of ls-broader-term
CERIF Federated Identifiers

• Result Publication
  – DOI
  – WoS Accession Number

• 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
• Connected to a Service representing the context:
  – The issuer of the identifier
    • Usually an information system
- For point-to-point interchange
- XML namespace
- XML Schema
  - Based on the ER model
<CERIF xmlns="urn:xmlns:org:eurocris:cerif-1.5-1">
  <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>
    <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 1.5 Release

- CERIF Model Introduction and Specification **coming**
- CERIF XML Data Exchange Format Specification ✓
- CERIF Formal Semantics (Vocabulary) ✓
- CERIF SQL Scripts ✓ *(euroCRIS members only)*
- CERIF XML Schemas ✓
- CERIF XML Examples ✓
- CERIF Semantics (Excel) ✓
Ongoing Activities: CERIF 1.6

- Model Cleaning
- Research Data

**Cross-TG Activities**
- Linked Open Data TG
- Institutional Repositories TG
- Architectures TG
- Indicators TG
- Best Practice TG

**Cooperation with**
- CASRAI
- VIVO
- RDA
- ORCID
What makes CERIF shine

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

Current Research Information System = CRIS

... that means
- of current interest
- not necessarily ongoing

... information about
- Researchers
- Organisations (Research-performing, Funding)
- Funding Programmes, Calls
- Projects
- ...

... driven by
- Concepts
- Model
- Implementation (Information System)

an integrated approach towards managing research information
CRIS and Repositories at an institution (slide by Keith Jeffery)

End-User

CERIF

CRIS
Research Context
[projects, persons, organisational units, funding, products, patents, publications, facilities, equipment, events]

CERIF

OA Repository (hypermedia) Documents

OAI-PMH

Various protocols

e-Research repository
Datasets and Software