CERIF:
Research Metadata

Jan Dvořák

- CERIF Task Group Leader @ euroCRIS
- Researcher @ Charles University in Prague, CZ
- Consultant @ InfoScience Praha, CZ

Research Metadata

- Discovery metadata for information to be found
  - Serve many specific use-cases, scenarios, niches

- Many standards
  - Tens of major ones
  - Hundreds of domain-specific standards
  - …
  - Thousands on experiment-level

The Purpose of Metadata

Enable the **re-use of resources**

- Knowledge stored in publications
- Data in datasets
- Functionality in software
- Events
- Infrastructure
  - Facilities
  - Equipment
  - Services

Common Grounds

- Organisations
  - Universities, Research institutes, Hi-tech companies
  - Research funding agencies
  - Publishers
  - Facility operators
  - Service providers
  - Associations, Societies

- People
  - Researchers
  - People in management positions

One Domain

Research

In Need of Consistency

- Several possible views of the same objects
  - From different points of view

- Inconsistencies look … *sub-optimal*

- Mind the gaps!

Common Metadata Format?

- A *lingua franca* for research metadata

- To be used in the data exchanges

Requirements …

- Complete coverage of research information

- Interlinked – providing the context

- Allow for many perspectives on the research information

- Accommodate multilinguality & support translations

- Accept the world keeps changing: record history

- Declared semantics: definitions rather than terms

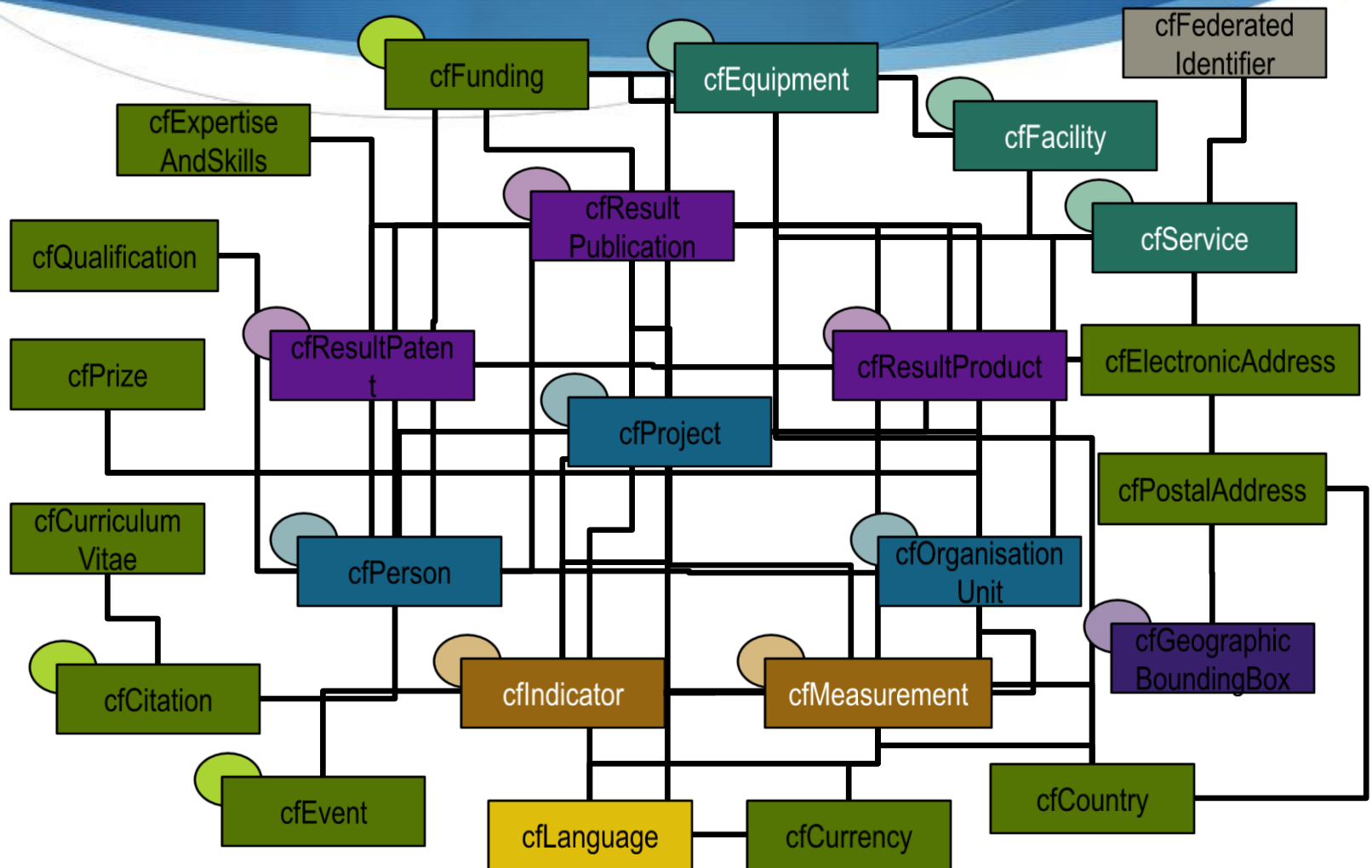- Formal syntax – machine processable & understandable

Satisfied by …

CERIF

Common European Research Information Format

CERIF: a concise history

- CERIF**91** – flat file

- CERIF **2000** – database structured
  - Free to choose your perspective

- CERIF is the EC recommendation to member states (**2002**)
  - http://cordis.europa.eu/cerif/

- CERIF **2006** – semantics moved into Semantic Layer
  - XML exchange format

- CERIF 1.5 (**2012**) – federated identifiers
  - incl. ORCIDs for researchers
  - incl. DOIs for datasets & publications
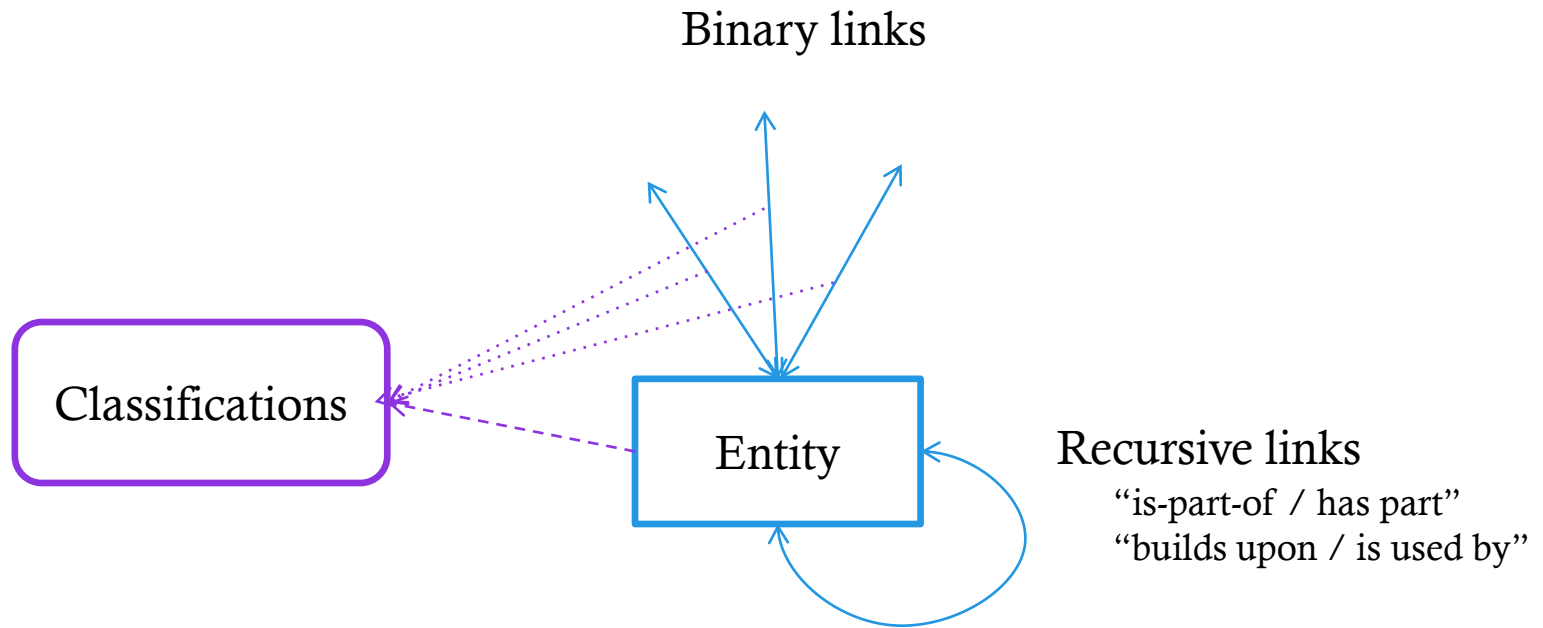
- CERIF 1.6 (**2013**) – support for datasets

CERIF: Complete Coverage

CERIF: Many Perspectives

- Start from any entity:
  - Project – funding, consortium, project team, outputs
  - Publication – authors, publisher, funding
  - Research dataset – creator/contributor, origin project, publications that build upon it
  - Person – outputs, datasets, projects, events, …
  - …

- A mesh, a fully connected graph

CERIF: Interlinking

Binary links

Classifications

Entity

Recursive links
"is-part-of / has part"
"builds upon / is used by"

CERIF: Record History

- Every relationship **records the time interval** (start date – end date) in which it {is|was|will be} true

- Open ends represented by effective $\pm\infty$

- When something changes:
  - the old relationship is NOT removed, just set its end date
  - a new relationship is inserted, start date

- Historic data accumulates naturally

CERIF: Declared Semantics

- Terms can be misleading
  - They only work within a context

- It's the real meaning that matters
  - Definition
  - Description
  - Examples

CERIF: Formal Syntax

- Currently available:
  - Relational database model
  - XML exchange format

- Coming:
  - Linked Open Data

CERIF Adoption

Current Research Information Systems (CRIS)
– CERIF used within CRIS, among them, in/out

- Research-performing organisations
  - Home built
  - Commercial offerings

- Funding agencies

- National research portals
  - UK, BE/Flanders, NL, NO, SK, SI, LV, ET, SE, DE

- FP7 projects (ENGAGE, OpenAIREplus, EPOS)

- …

euroCRIS

- Mission:
  Advance Interoperability in the Research Community through CERIF

- Not-for-profit association of research information management professionals
  - ~120 institutional members, ~350 people

- Est. 2002

- Appointed the stewardship of CERIF by the EC
  - http://cordis.europa.eu/cerif/

- Work done in Task Groups

euroCRIS Partnerships

International Council for Science;
Commission on Data Access

European Association of Research
Managers and Administrators

All European Academies

euroCRIS Partnerships

CERIF 2.0: The Outlook

- Input from other euroCRIS Task Groups

  - Linked Open Data TG

  - Institutional Repositories TG

  - Architectures TG

  - Indicators TG

  - Best Practice TG

- Cooperation with CASRAI, VIVO, RDA, ORCID