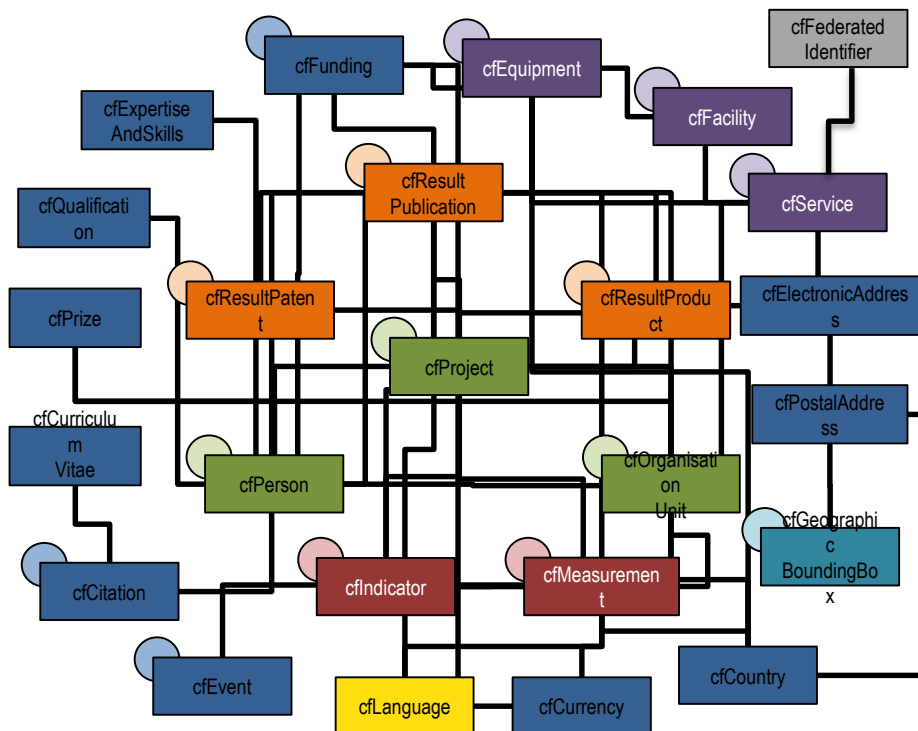


# CERIF Tutorial: Part 2



**Jan Dvořák**

October 2<sup>nd</sup>, 2018

I encuentro de gestores  
de información en CTI - Perú  
ESAN University, Lima, Perú

# Interoperability

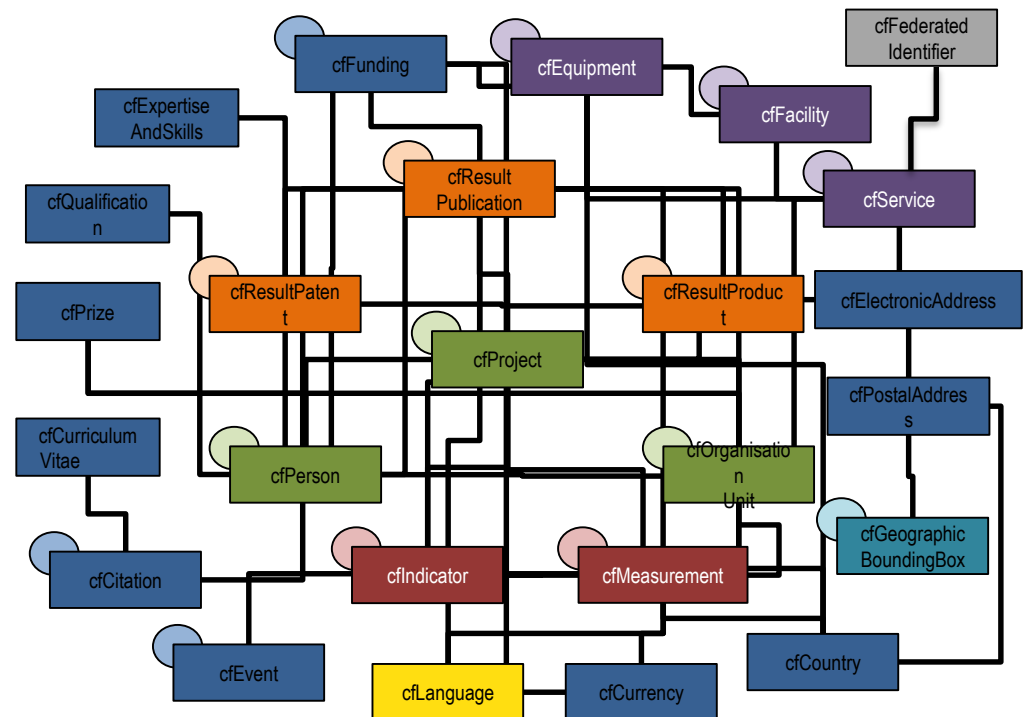
1. Agreed Semantics
2. Agreed Format
3. Agreed Protocol

# CERIF XML Interchange Format

XML Schema

Based on the ER model

Undergone a big update



# OpenAIRE Guidelines for CRIS Managers

## 1.1

- Dvořák, Jan, Bollini, Andrea, Rémy, Laurent, & Schirrwagen, Jochen. (2018). OpenAIRE Guidelines for CRIS Managers 1.1 (Version 1.1.0). Zenodo. DOI [10.5281/zenodo.1298650](https://doi.org/10.5281/zenodo.1298650)
  
- Document on the web <https://openaire-guidelines-for-cris-managers.readthedocs.io/en/latest/index.html>
  - Introduction
  - CRIS Information elements relevant for OpenAIRE
    - CERIF, CERIF XML
  - Technical implementation guidelines
    - CERIF XML as payload of OAI-PMH 2.0

# Version 1.1: Aligned vocabularies

- COAR Resource Types
- COAR Access Rights
- ISSN Media List

OpenAIRE Guidelines for CRIS Managers  
latest

Search docs

Introduction

- CRIS information elements relevant for OpenAIRE
  - Publication
    - Internal Identifier
    - Type
    - Language
    - Title
    - Subtitle
    - PublishedIn
    - PartOf
    - PublicationDate
    - Number
    - Volume
    - Issue
    - Edition

Read the Docs v: latest

## Type

**Description:** The type of the publication

**Use:** mandatory (1)

**Representation:** XML element `Type` from namespace [https://www.openaire.eu/cerif-profile/vocab/COAR\\_Publication\\_Types](https://www.openaire.eu/cerif-profile/vocab/COAR_Publication_Types)

**CERIF:** the ResultPublication\_Classification ([https://w3id.org/cerif/model#ResultPublication\\_Classification](https://w3id.org/cerif/model#ResultPublication_Classification))

**Vocabulary:** Publication types extracted from the COAR Resource Types concept scheme: Types of publications as extracted from the COAR Resource Types concept scheme ([http://vocabularies.coar-repositories.org/documentation/resource\\_types/](http://vocabularies.coar-repositories.org/documentation/resource_types/), the term 'text' and its descendants in the hierarchy except 'patent').

- **text** ([http://purl.org/coar/resource\\_type/c\\_18cf](http://purl.org/coar/resource_type/c_18cf)): A resource consisting primarily of words for reading. Examples include books, letters, dissertations, poems, newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre Text.
  - **annotation** ([http://purl.org/coar/resource\\_type/c\\_1162](http://purl.org/coar/resource_type/c_1162)): An annotation in the sense of a legal note is a legally explanatory comment on a decision handed down by a court or arbitral tribunal.
  - **bibliography** ([http://purl.org/coar/resource\\_type/c\\_86bc](http://purl.org/coar/resource_type/c_86bc)): A systematic list or enumeration of written works by a specific author or on a given subject.
  - **book** ([http://purl.org/coar/resource\\_type/c\\_2f33](http://purl.org/coar/resource_type/c_2f33)): A non-serial publication that is complete in one volume or a designated finite number of volumes. (adapted from CITO; EPrint Type vocabulary)

# Information scope of the OpenAIRE Guidelines for CRIS Managers

## 1.1

### The Core:

- The CRIS as a Service
- Organisations
- Researchers

### Activities:

- Projects
- Funding
- Events

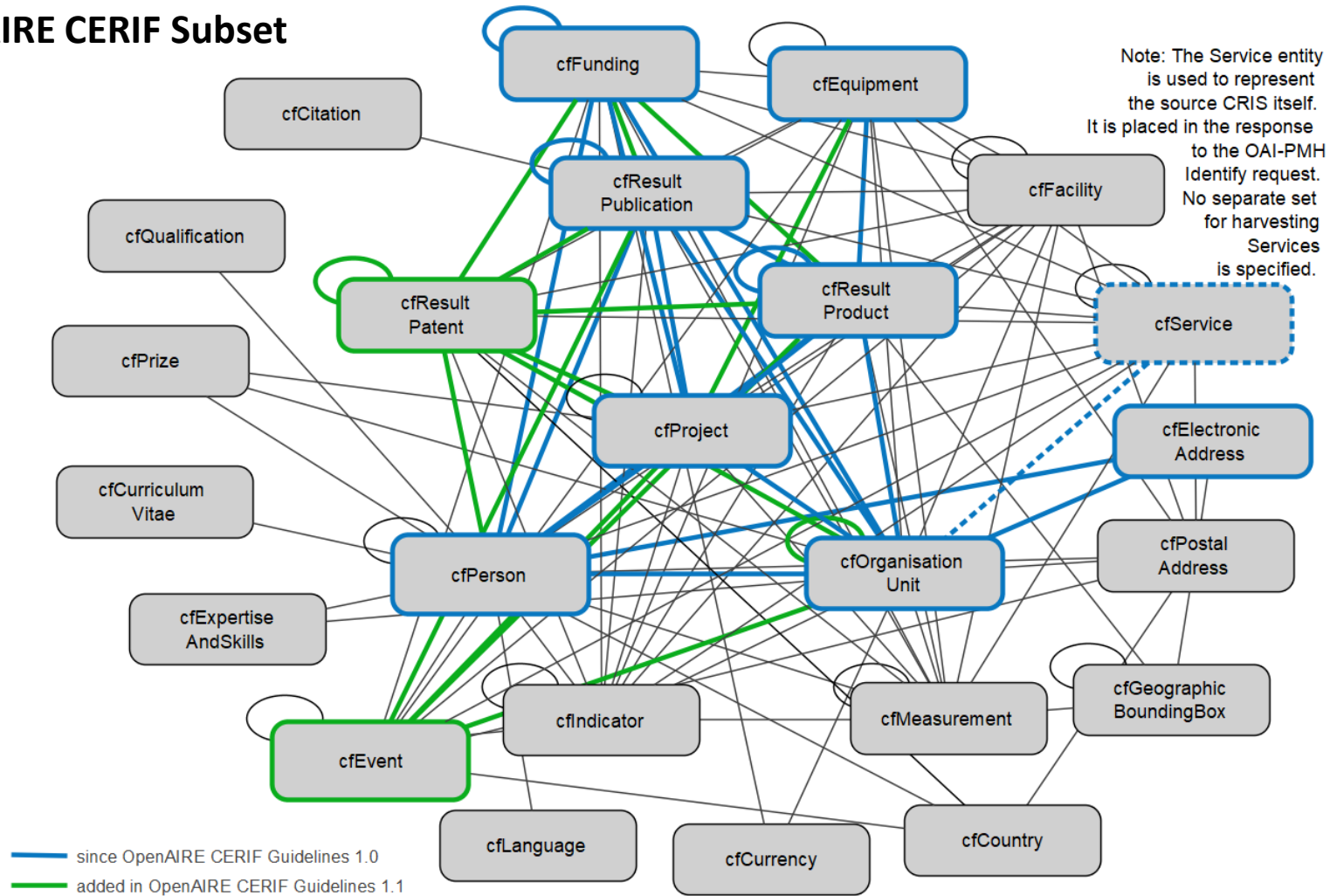
### Outputs:

- Publications
- Datasets, Software & other Products
- Patents

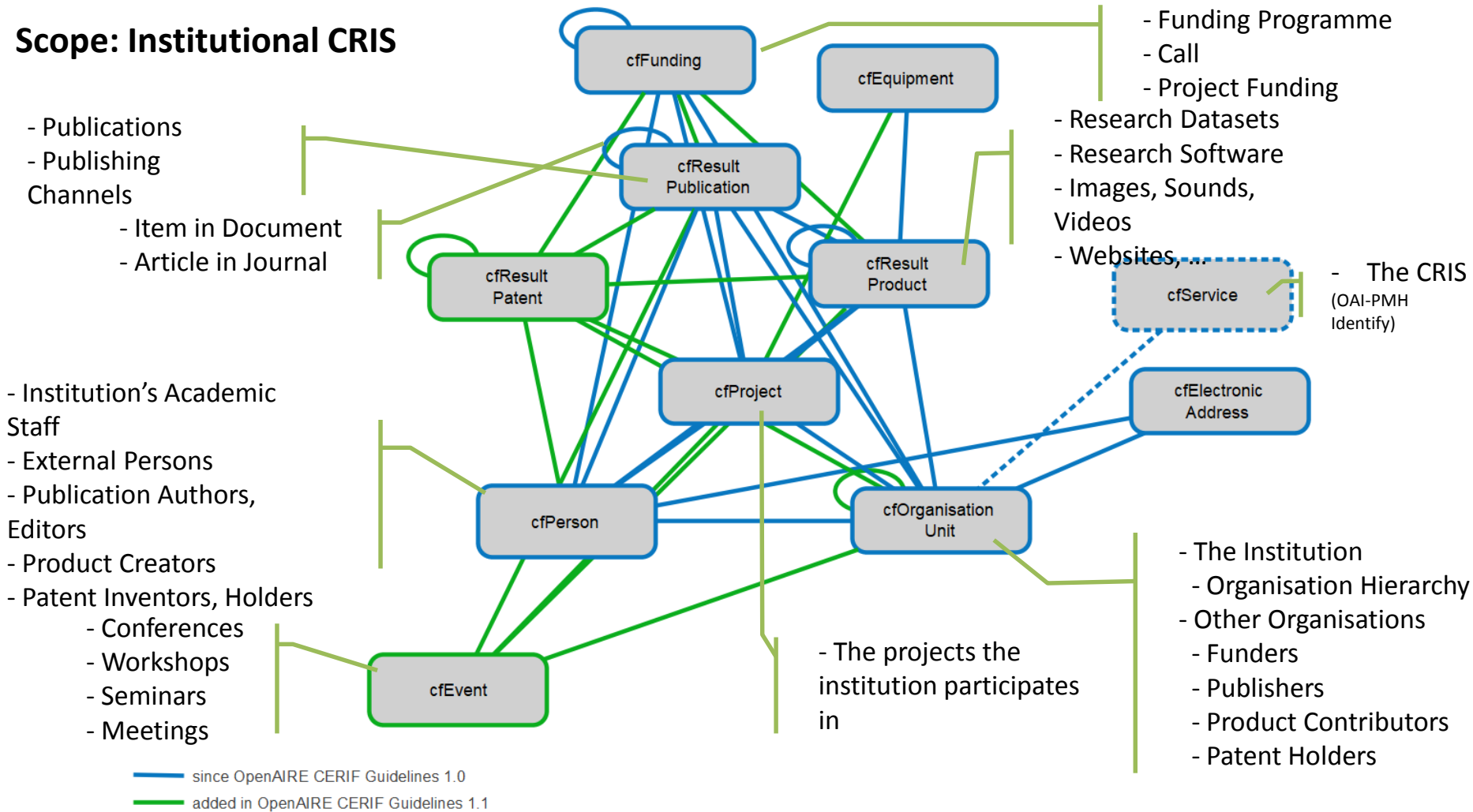
### Infrastructure:

- Equipment

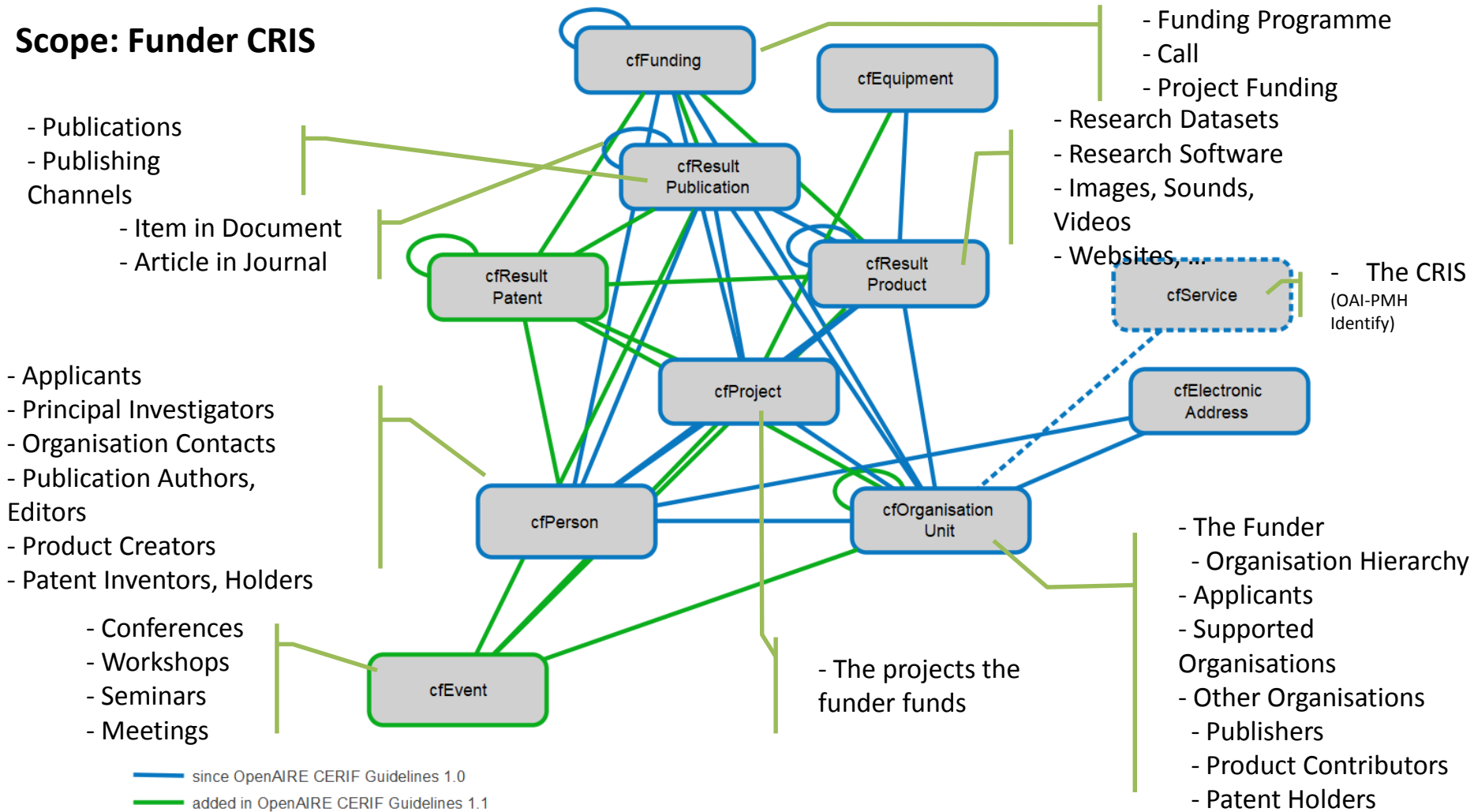
## OpenAIRE CERIF Subset



## Scope: Institutional CRIS



## Scope: Funder CRIS



# Identifiers

- Publication:
  - DOI
  - Handle
  - PMCID
  - ISI Number (WoS Accession Number)
  - SCP Number
  - ISSN
  - ISBN
  - URL
  - URN
- Persons:
  - ORCID
  - ResearcherID
  - Scopus Author ID
  - ISNI
- Products:
  - ARK
  - DOI
  - Handle
  - URL
  - URN

Elsewhere generic

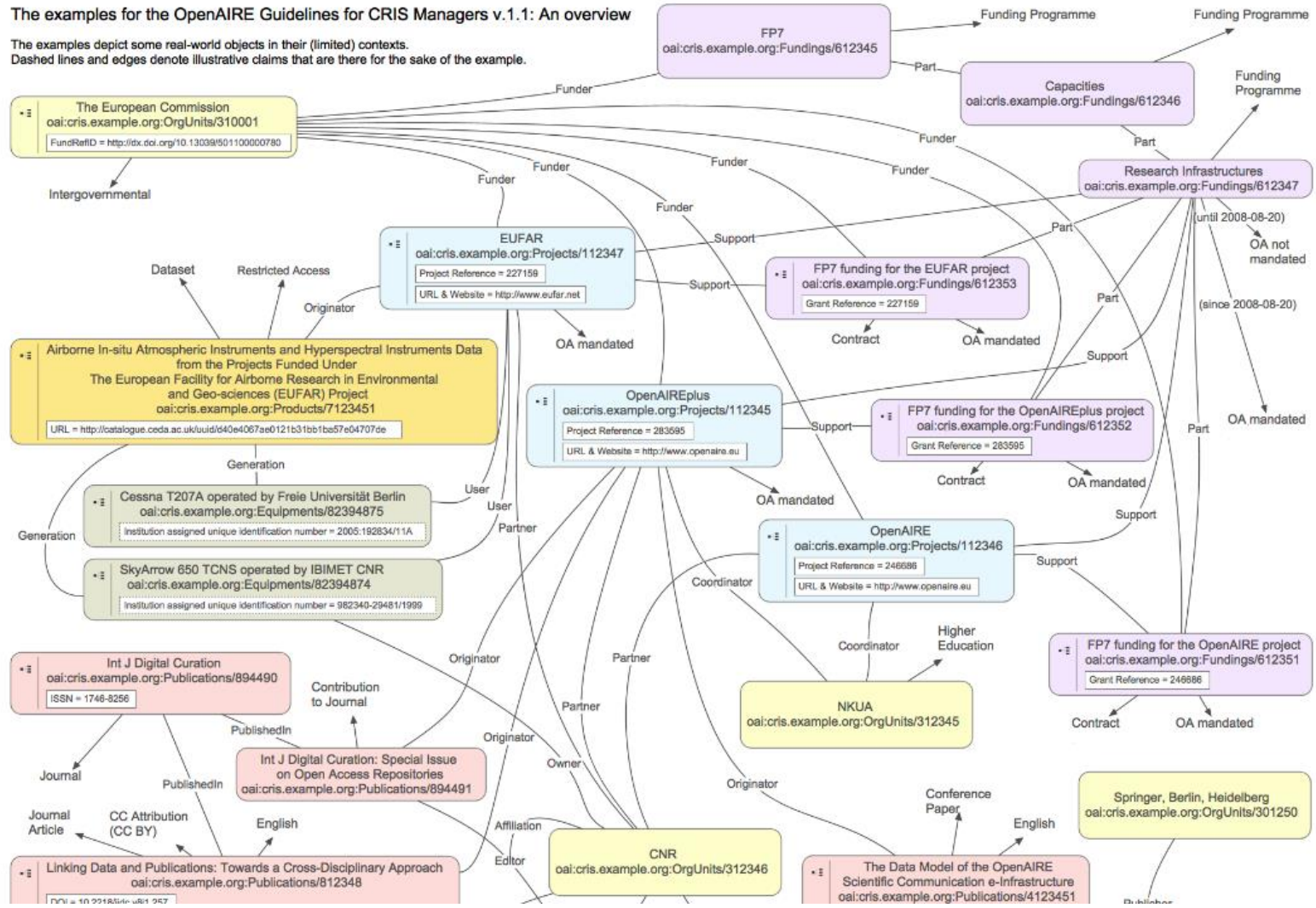
## Supporting artifacts

- XML Schema
  - Namespace <https://www.openaire.eu/cerif-profile/1.1/>
  - A few Schematron integrity rules embedded
- Comprehensive set of examples
- Prototype validator

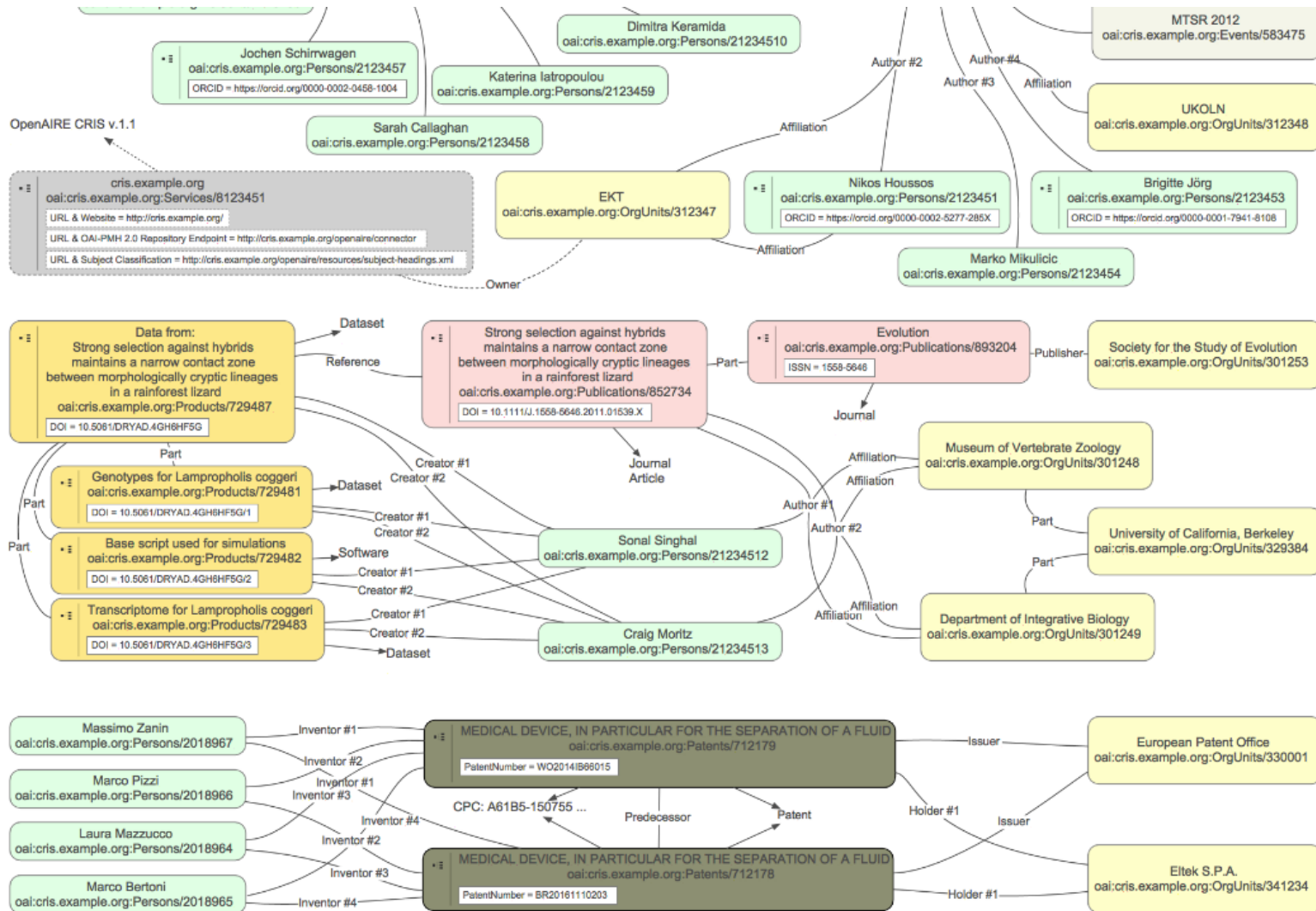


## The examples for the OpenAIRE Guidelines for CRIS Managers v.1.1: An overview

The examples depict some real-world objects in their (limited) contexts.  
Dashed lines and edges denote illustrative claims that are there for the sake of the example.







```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.1/" id="852734">
  <Type xmlns="https://www.openaire.eu/cerif-
profile/vocab/COAR_Publication_Types">http://purl.org/coar/resource_type/c_6501<!-- journal article --></Type>
  <Title xml:lang="en">Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic
lineages in a rainforest lizard</Title>
  <PublishedIn>
    <Publication id="893204">
      <Type scheme="https://www.openaire.eu/cerif-
profile/vocab/COAR_Publication_Types">http://purl.org/coar/resource_type/c_0640<!-- journal --></Type>
      <Title xml:lang="en">Evolution</Title>
      <ISSN medium="http://issn.org/vocabularies/Medium#Online">1558-5646</ISSN>
      <Publishers>
        <!-- [ ... ] -->
      </Publishers>
    </Publication>
  </PublishedIn>
  <Volume>66</Volume>
  <Issue>5</Issue>
  <StartPage>1474</StartPage>
  <EndPage>1489</EndPage>
  <DOI>10.1111/J.1558-5646.2011.01539.X</DOI>
  <!-- [ ... ] -->
</Publication>
```

```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.1/" id="852734">
  <!-- [ ... ] -->
  <Authors>
    <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> <!-- [ ... ] -->
  </Authors>
  <!-- [ ... ] -->
</Publication>
```

```
<Publication xmlns="https://www.openaire.eu/cerif-profile/1.1/" id="852734">
  <!-- [ ... ] -->
  <Keyword xml:lang="en">hybrid zones</Keyword>
  <Keyword xml:lang="en">phylogeography</Keyword>
  <Abstract xml:lang="en">Phenotypically cryptic lineages comprise an important yet understudied part of biodiversity;
  <!-- [...] --></Abstract>
  <References>
    <Product id="729487">
      <Type xmlns="https://www.openaire.eu/cerif-
profile/vocab/COAR_Product_Types">http://purl.org/coar/resource_type/c_ddb1<!-- dataset --></Type>
      <Name xml:lang="en">Data from: Strong selection against hybrids maintains a narrow contact zone between
morphologically cryptic lineages in a rainforest lizard</Name>
      <VersionInfo xml:lang="en">1</VersionInfo>
      <DOI>10.5061/DRYAD.4GH6HF5G</DOI>
      <!-- [...] -->
    </Product>
  </References>
</Publication>
```

# METIS2OpenAIRE Project

## METIS

- An in-house built CRIS at Radboud University [NL]
- Now at 2 Dutch Universities

METIS harvested  
through  
the OpenAIRE  
CRIS endpoint

| OAI identifier prefix         | # objects |
|-------------------------------|-----------|
| oai:metis.ru.nl:OrgUnits/     | 541       |
| oai:metis.ru.nl:Persons/      | 178405    |
| oai:metis.ru.nl:Products/     | 17160     |
| oai:metis.ru.nl:Projects/     | 5080      |
| oai:metis.ru.nl:Publications/ | 109429    |

# Prototype Validator Implementation

<https://github.com/jdvorak001/openaire-cris-validator#readme>

(Java, Apache License 2.0)

Deliverable in the METIS2OpenAIRE project

- A suite of 13 JUnit tests
- Runnable from the command line or IDE

# Prototype Validator Rules

## Endpoint:

- `?verb=Identify` - gives the basic description as per the spec
- `?verb=ListMetadataFormats` - lists `oai_cerif_openaire`
- `?verb=ListSets` - lists all 9 sets defined by the spec

## Each record:

- Validity w.r. to the XML Schema
- OAI identifiers follow the spec

## All records:

- OAI identifiers unique
- Referential integrity
- Functional dependency

## Another profile: NL P-O-P-F

- Scope:
  - Person – Organisation – Project – Funding
- Notification:
  - Funder → Institution
  - When project funding has been approved



# CERIF Profiles in general

Useful subsets of CERIF  
for specific research information  
interchange scenarios

- Entities & attributes:
  - Profile  $\subset$  CERIF
- Semantic vocabularies:
  - Profile – specific choices
  - Sources: CERIF & beyond
- Integrity constraints:
  - Profile  $\supset$  CERIF

Profile data is  
CERIF

# CERIF Profiles in general

Producers know  
what to include



Consumers know  
what to expect

# CRIS Interoperability Protocols

- OAI-PMH
  - CERIF XML profile as the payload format
  - For mirroring metadata between systems
- REST API
  - CERIF XML profile as the response format
  - Querying (incl. mirroring)
- Linked Open Data
  - RDF data in CERIF Ontology
  - Good for the public part of research information
- Submissions  
(emails, file attachments, submission services)

## Mirrored records use-cases

- Replace our records
  - If authoritative!
- Basis for our new records
- Compare with our records
  - Have users resolve differences
- Aggregation

## Authorization in CRISs

- Typically based on:
  - User's roles in publications/projects/...
  - User's roles within the organisation
  - Organisational structure
- Can be very complex!

Gracias por su  
atención!