MarcOnt – Integration Ontology for Bibliographic Description Formats

Sebastian R. Kruk, Marcin Synak, Kerstin Zimmermann

sebastian.kruk@deri.org
www.marcont.org

DC2005, Madrid
13.09.2005

Outline of the presentation

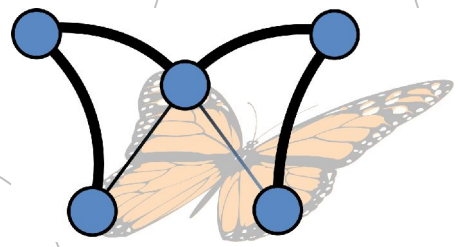
- Motivations
- MarcOnt Initiative Overview and Related Work
- Different types of Libraries
- Bibliographic Description Formats
- MarcOnt Ontology
- MarcOnt Mediation Services
- JeromeDL – proof of concept
- Future Work

Motivations

- Support for different kinds of bibliographic metadata, like: DublinCore, BibTeX and MARC21 at the same time.
- Utilize existing legacy bibliographic descriptions (like BibTeX or MARC21)
- Interlinking heterogeneous digital library networks.

MarcOnt Initiative Overview and Related Work

- Goal: utilize the existing, legacy metadata in semantically enabled libraries
- Defines ontology that would cover concepts from MARC21, BibTeX and DublinCore
- Provides interoperability services to communicate with other entities using legacy metadata
- Similar solutions:
  - ABC ontology
  - GINF mediation framework



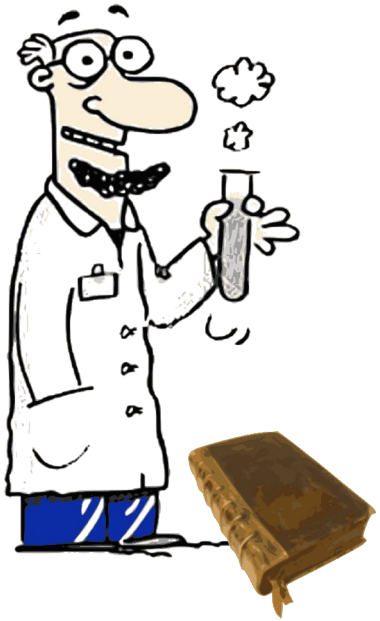
Different Kind of Libraries (Evolution of Libraries)

Different Kind of Libraries (Evolution of Libraries)

- Classic Library

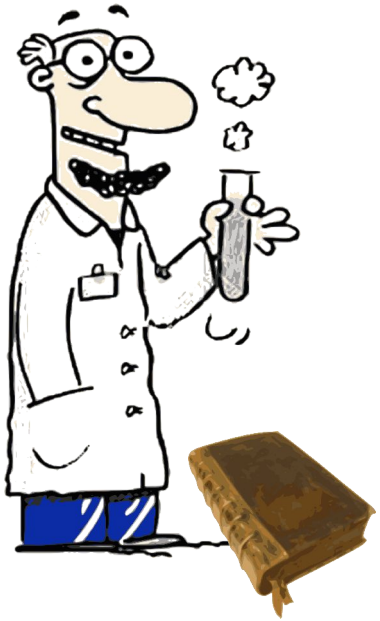
Different Kind of Libraries (Evolution of Libraries)

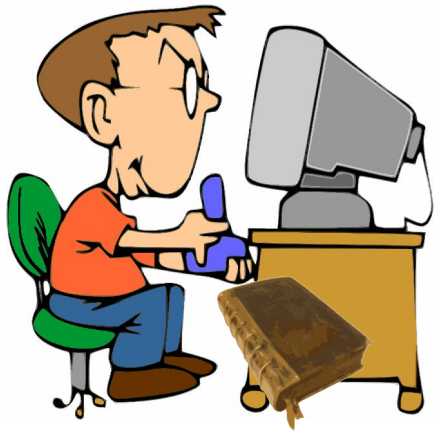
- Classic Library
- Publishers' Library



Different Kind of Libraries (Evolution of Libraries)

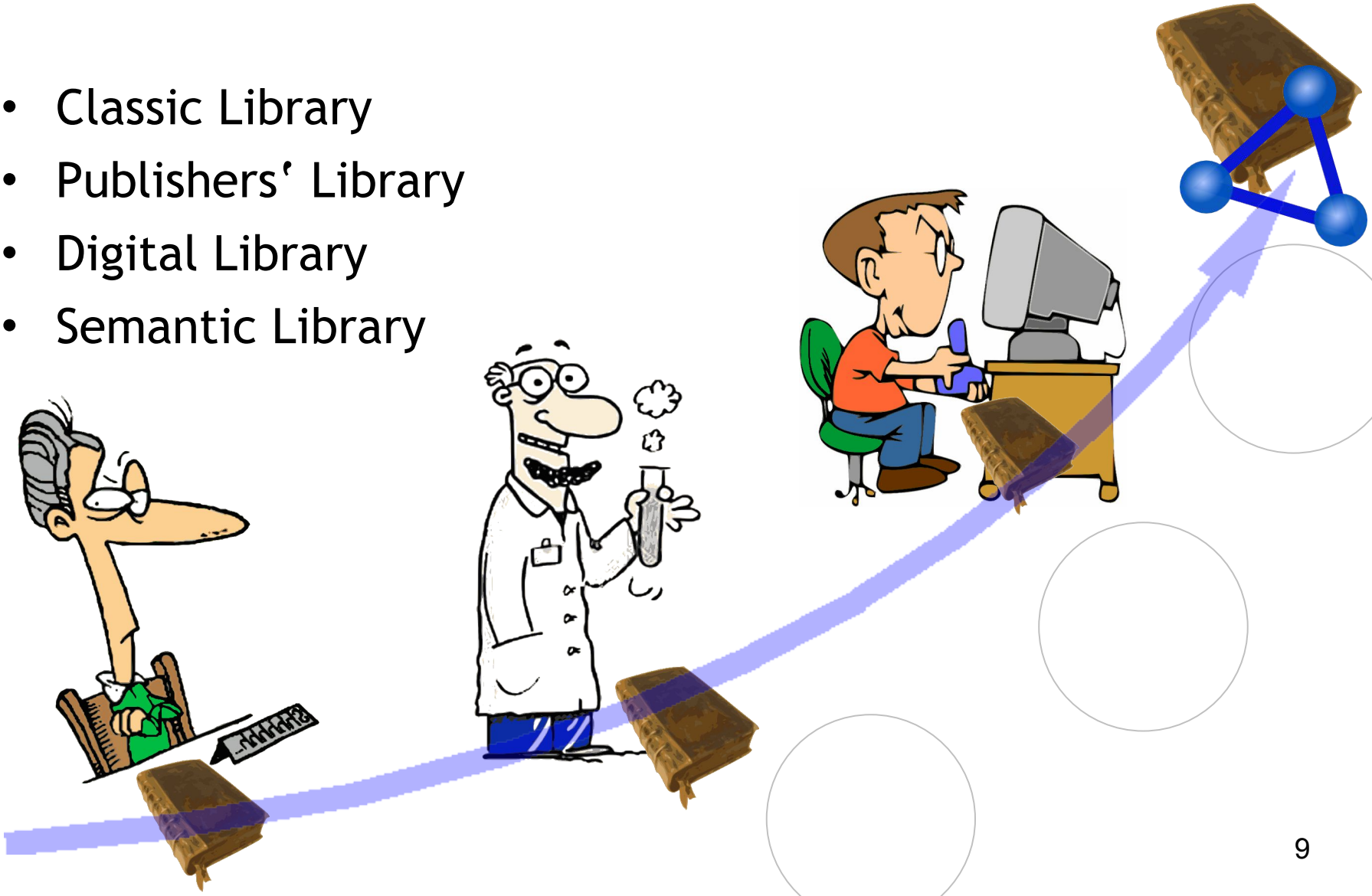
- Classic Library
- Publishers' Library
- Digital Library





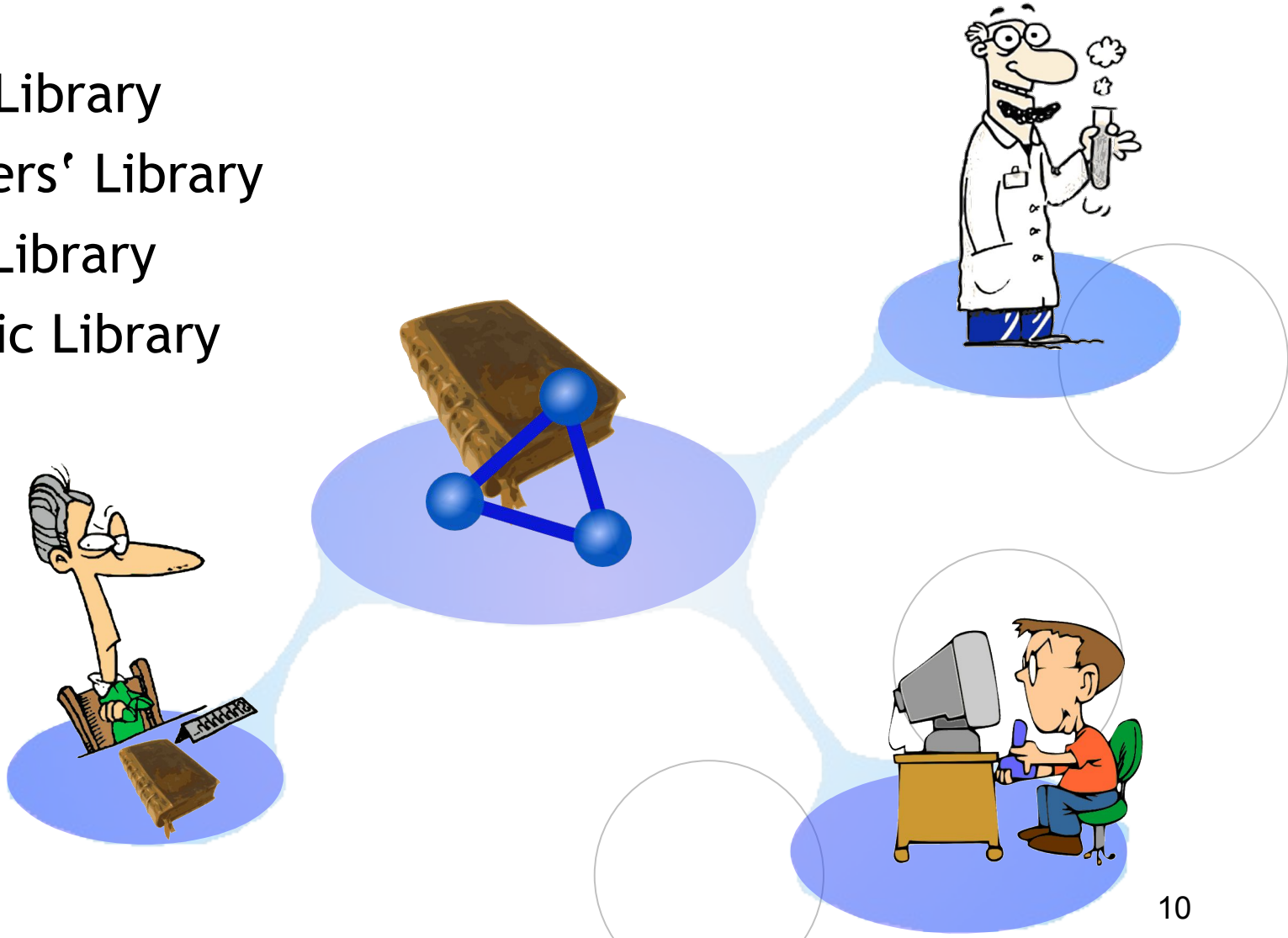
Different Kind of Libraries (Evolution of Libraries)

- Classic Library
- Publishers' Library
- Digital Library
- Semantic Library



Different Kind of Libraries (Evolution of Libraries)

- Classic Library
- Publishers' Library
- Digital Library
- Semantic Library



Description Formats

- MARC21:
  - numerical fields 001-887
  - 8 major types of material
  - 7 major types of records
  - number of various classifications
- BibTeX:
  - variable elements
  - 12 types of publications
- Dublin Core:
  - 15 meta elements
  - 12 types of categories

Different Types

Books
Continuing resources
Computer files
Maps
Music
Sound recordings
(non music)
Visual materials
Mixed materials

Language (textual)
material
Manuscript (textual)
language material
Computer file
Cartographic material
Manuscript cartographic
material
Notated music
Manuscript music

MARC21

String
Book
InBook
InCollection
Proceedings
InProceedings
Article
MasterThesis
PhDThesis
TechReport
Manual
Misc

BibTeX

Collection
Dataset
Event
Image
MovingImage
StillImage
Sound
Text
Interactive resource
Physical object
Service
Software

DublinCore

Different Encoding

01450cas 922004331i
450000100...019c19329999gw qr|p|
||||0 |0ger | a0044-2992
9a200412140219bVLOADc20040407152
5dvkulc200310071018dvbjc200303101
205dkopumky200209211341zVLOAD
aGD U/MPcGD U/MPdGD U/MFdGD
U/KKsdWR O/EJ0 ager1 aZ.
Kunstgesch. 0aZeitschrift für
Kunstgeschichte00aZeitschrift für
Kunstgeschichte.18aZfK aMünchen
;aBerlin :bDeutscher
Kunstverlag,c1932-. c26-29 cm.
aKwart.0 a1 Bd. (Juni 1932)-. aOpis
na podst.: LCC. aW 1932
założycielami czasopisma byli
Wilhelm Waetzoldt i Ernst Gall....

```
@InProceedings { jeromedexa2005,
  author = "Sebastian Ryszard Kruk and  ... ",
  title = "{JeromeDL - Adding Semantic  ...}",
  booktitle = "{In Proceedings to DEXA 2005}",
  year = 2005}
```

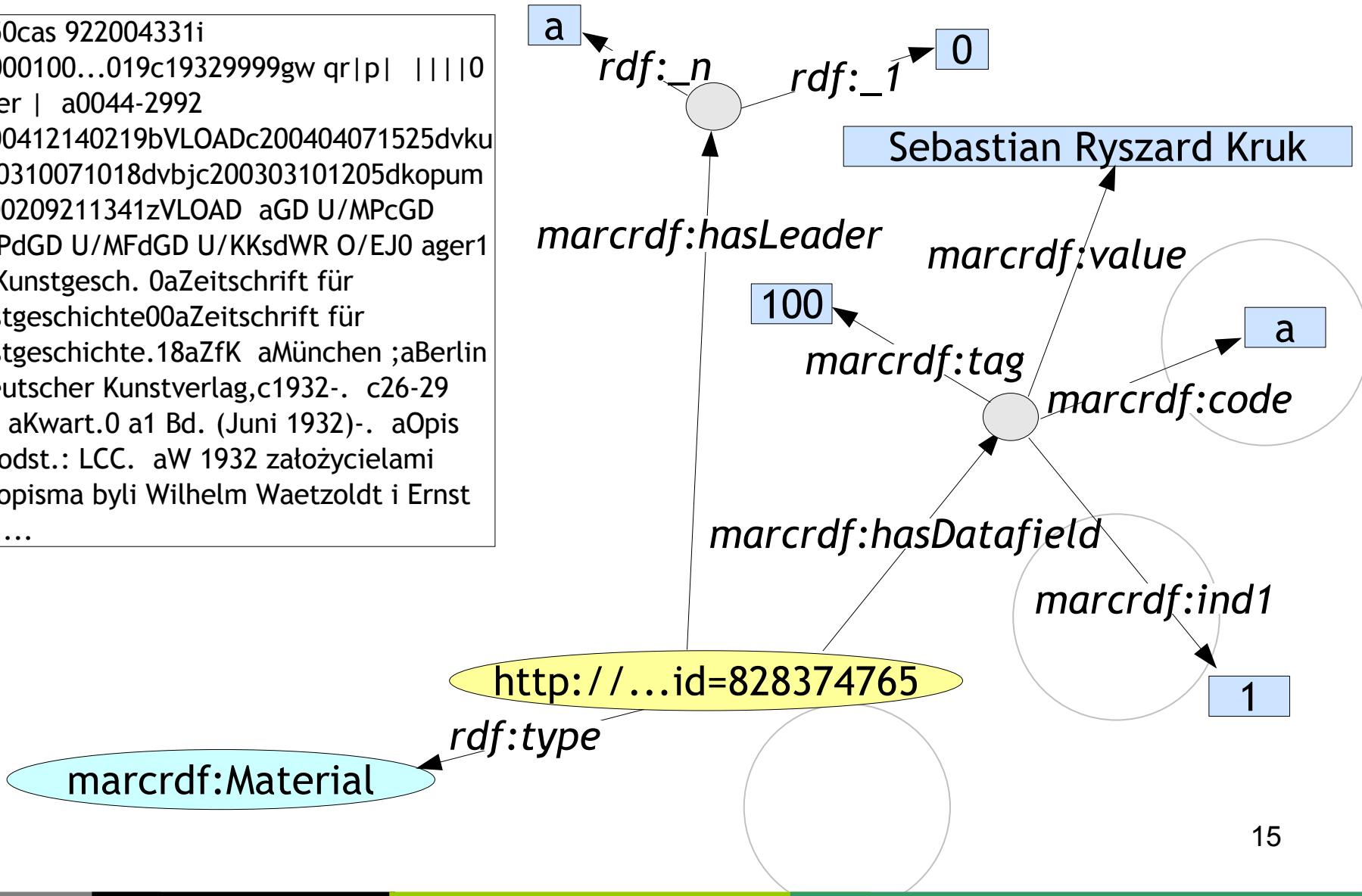
```
<?xml version="1.0" encoding="UTF-8" ?>
<Description>
    <dc:title>JeromeDL - Adding Semantic
    Web Technologies to Digital
    Libraries</dc:title>
    <dc:creator>Sebastian Ryszard
    Kruk</dc:creator>
    <dc:type> ... </dc:type>
</Description>
```

Interoperability challenges

- providing access in a joined query language
- unified data representation – RDF
- unified vocabulary – ontology

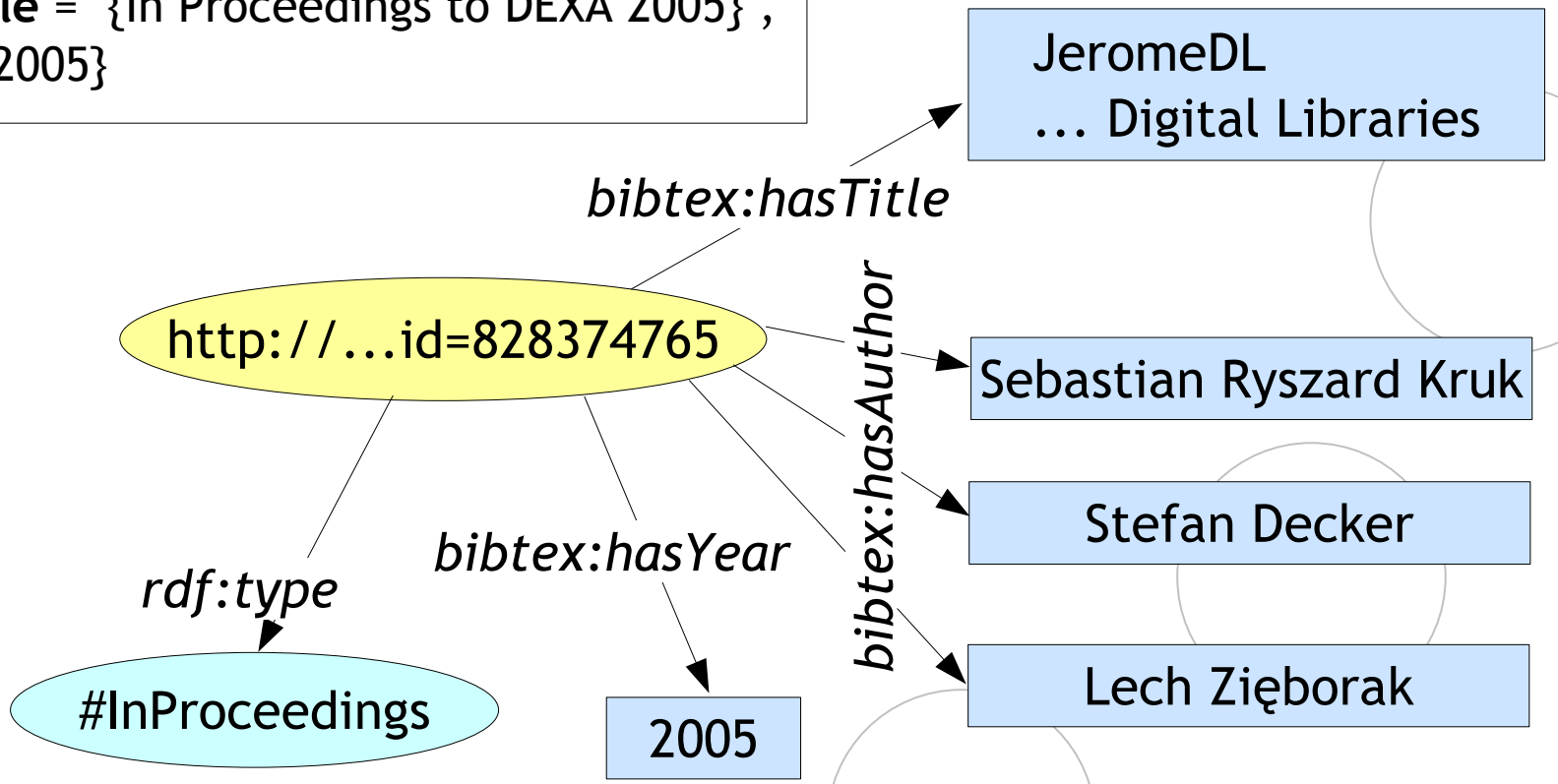
Bibliographic Description - MARC21

01450cas 922004331i
450000100...019c19329999gw qr|p| ||||0
|0ger | a0044-2992
9a200412140219bVLOADc200404071525dvku
lc200310071018dvbjc200303101205dkopum
ky200209211341zVLOAD aGD U/MPcGD
U/MPdGD U/MFdGD U/KKsdWR O/EJ0 ager1
aZ. Kunstgesch. 0aZeitschrift für
Kunstgeschichte00aZeitschrift für
Kunstgeschichte.18aZfK aMünchen ;aBerlin
:bDeutscher Kunstverlag,c1932-. c26-29
cm. aKwart.0 a1 Bd. (Juni 1932)-. aOpis
na podst.: LCC. aW 1932 założycielami
czasopisma byli Wilhelm Waetzoldt i Ernst
Gall....

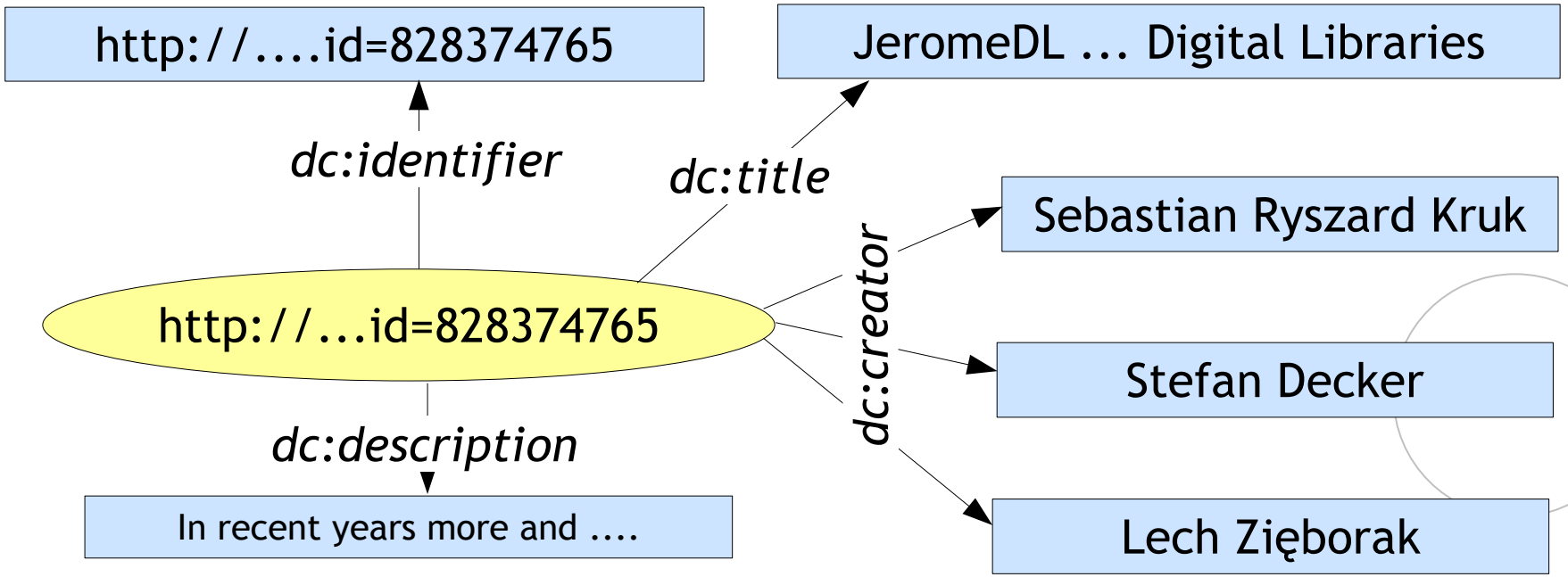


Bibliographic Description - BibTeX


```
@InProceedings { jeromedexa2005,
  author = "Sebastian Ryszard Kruk and  ... ",
  title = "{JeromeDL - Adding Semantic  ...}",
  booktitle = "{In Proceedings to DEXA 2005}",
  year = 2005}
```

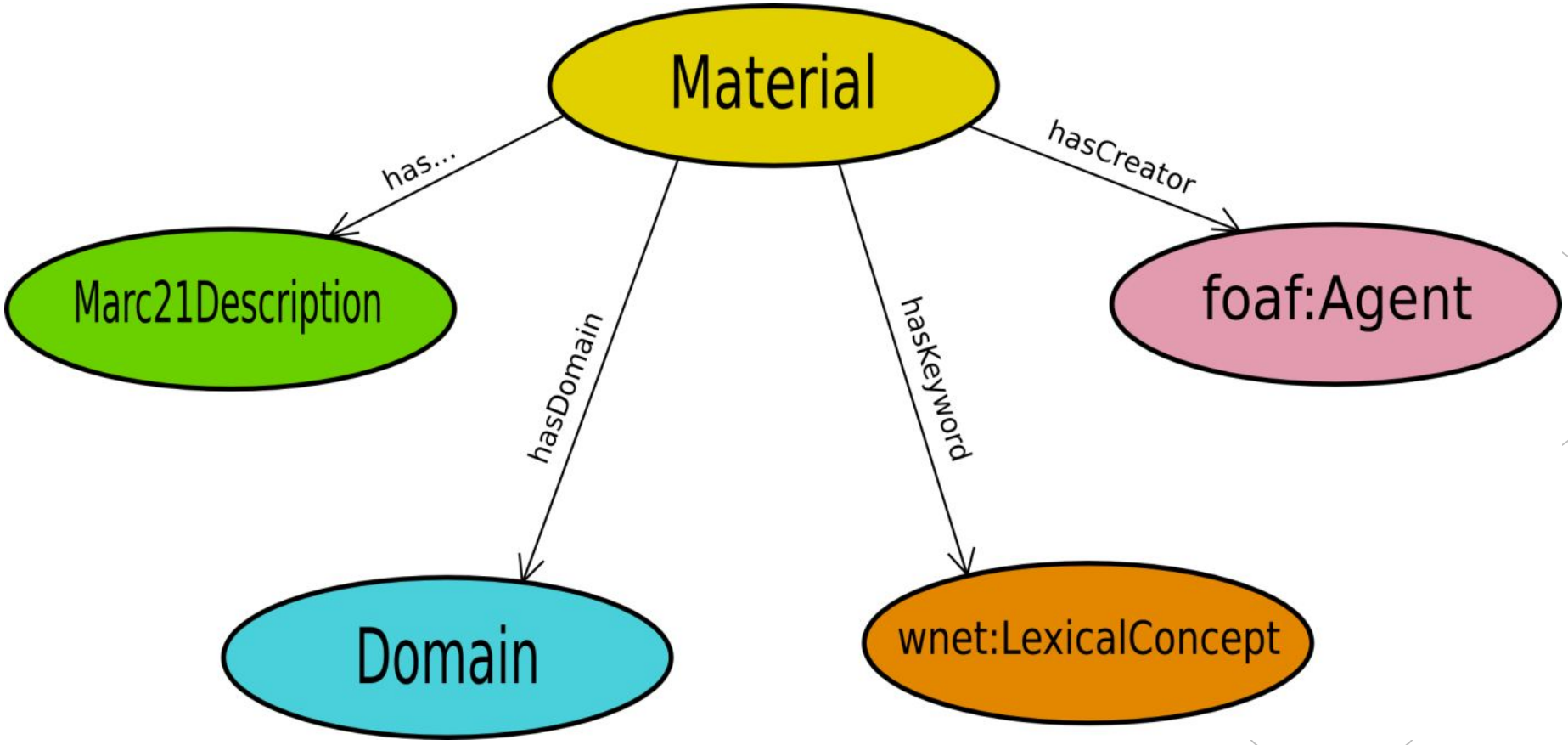


Bibliographic Descriptions in RDF – Dublin Core




```
<?xml version="1.0" encoding="UTF-8" ?>
<Description>
    <dc:title>JeromeDL - Adding Semantic Web Technologies to Digital
    Libraries</dc:title>
    <dc:creator>Sebastian Ryszard Kruk</dc:creator>
    <dc:type> ... </dc:type>
</Description>
```

Common vocabulary - MarcOnt Ontology



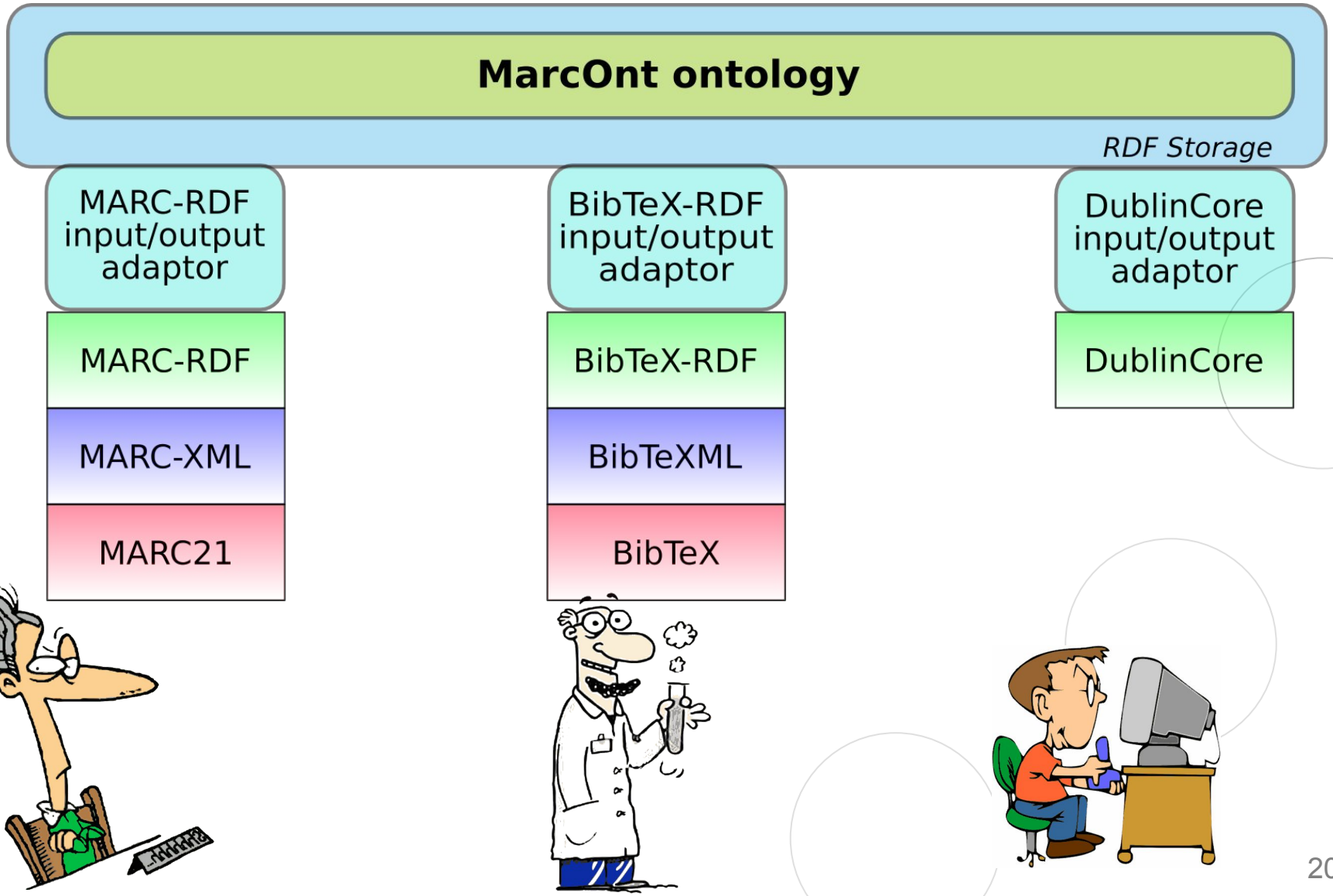
MarcOnt Mediation Services

01450cas92200433li450000100..
.019c19329999gw qr|p| ||||0 |
0gera004429929a200412140219
bVLOADc20040407

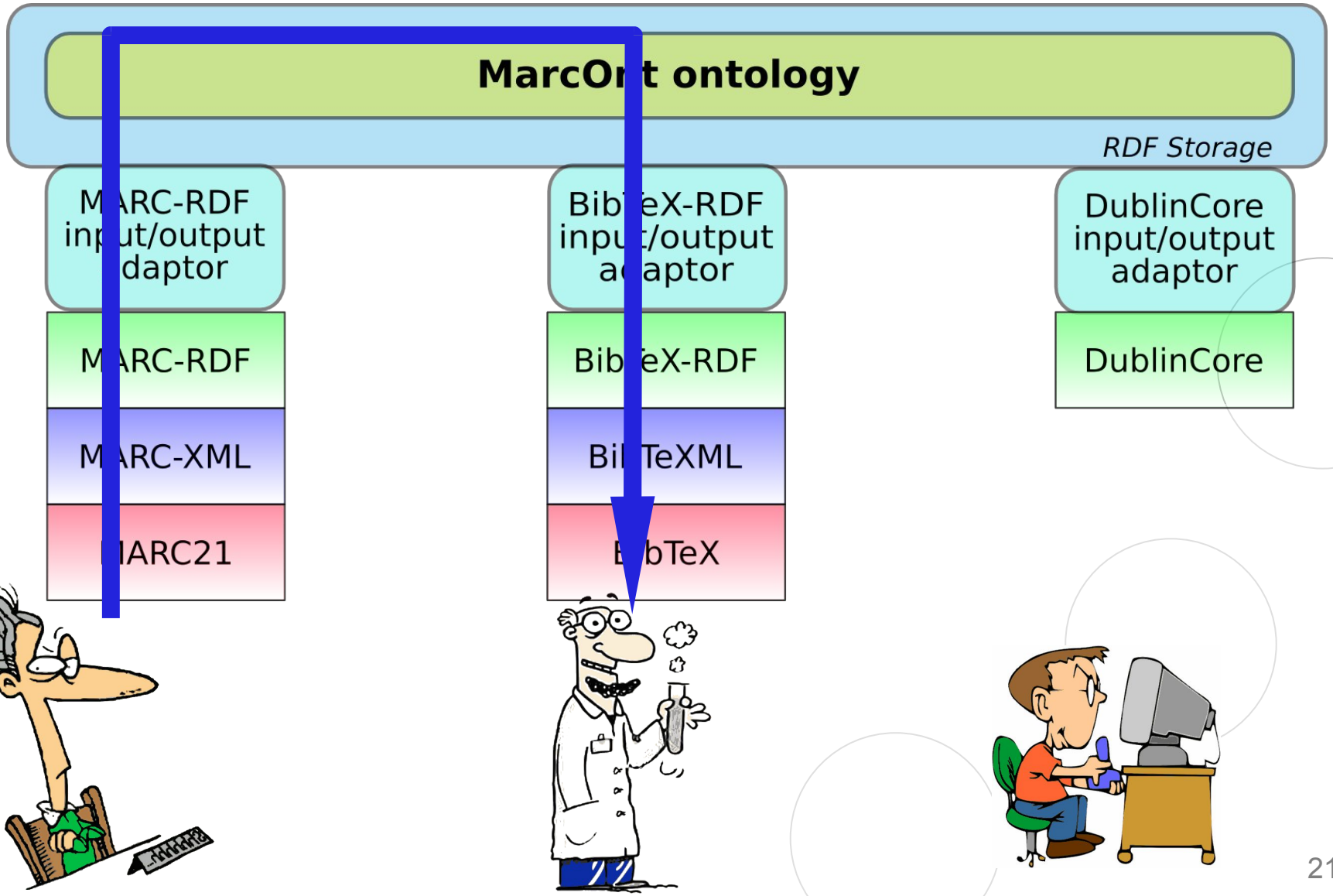
@InProceedings {jeromedexa2005,
author = "Sebastian Kruk and ... ",
title = "{JeromeDL - ...}",
year = 2005}

<Description>
<dc:title>MarcOnt ...</dc:title>
<dc:creator> Kruk</dc:creator>
</Description>

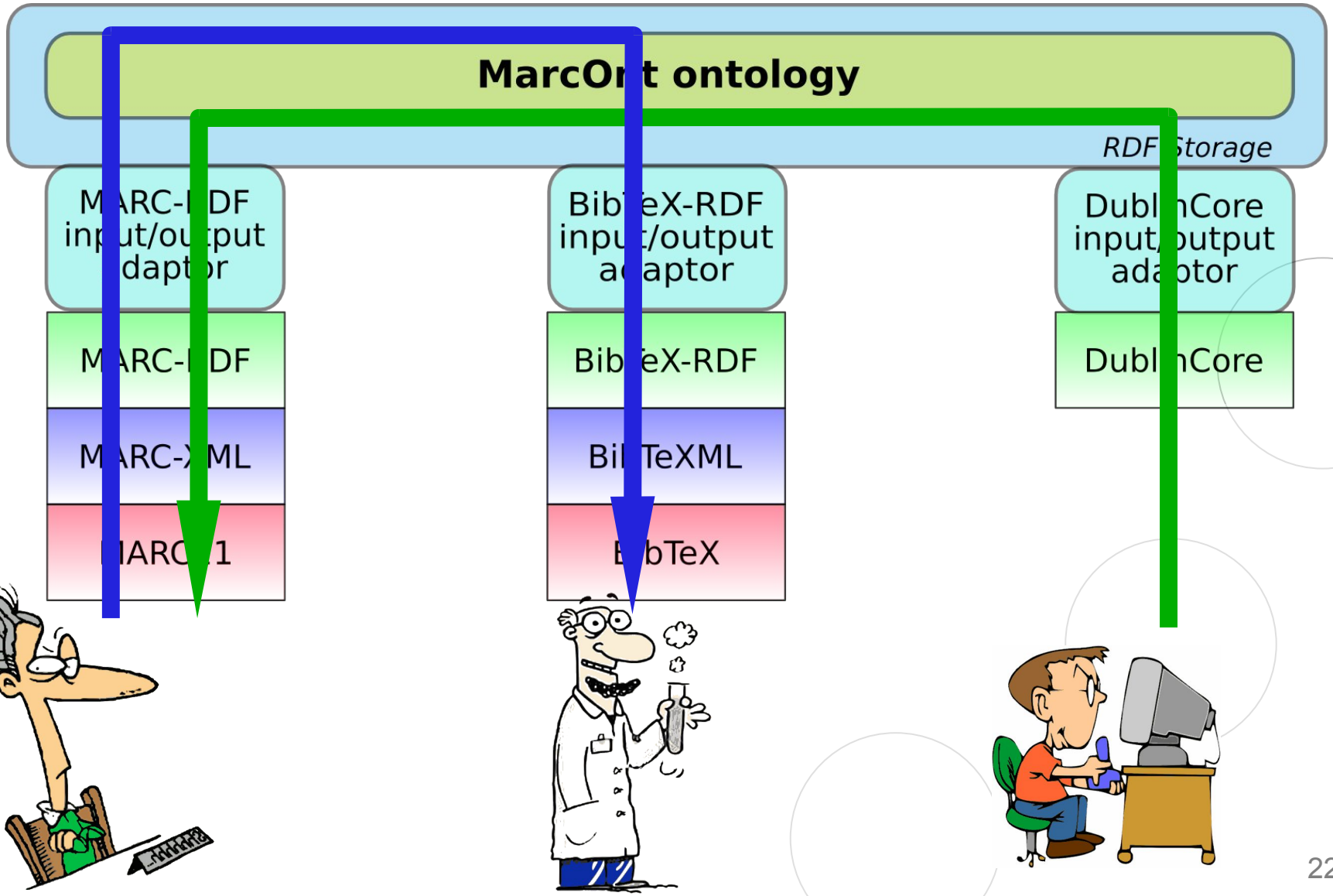
MarcOnt Mediation Services



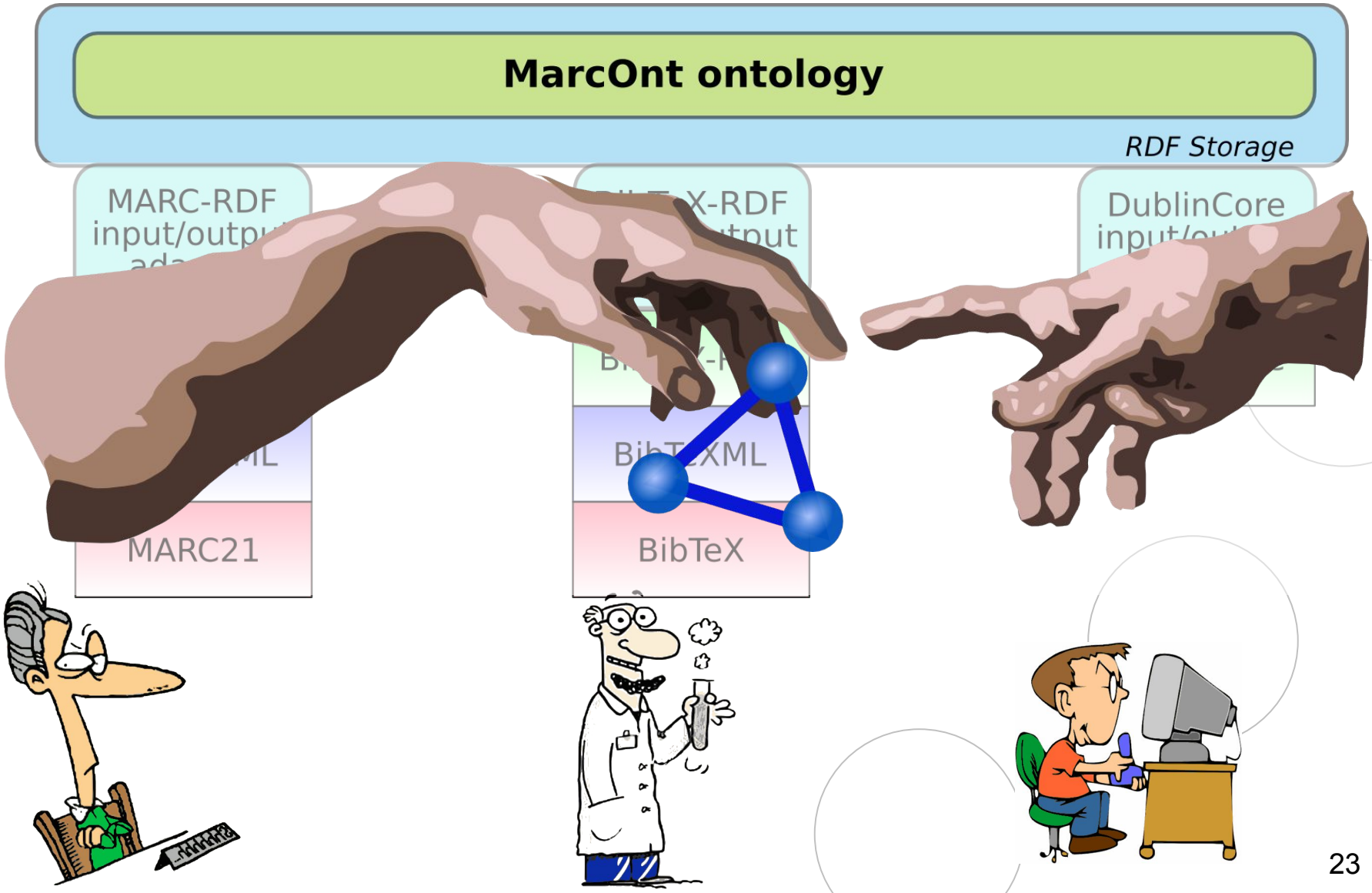
MarcOnt Mediation Services



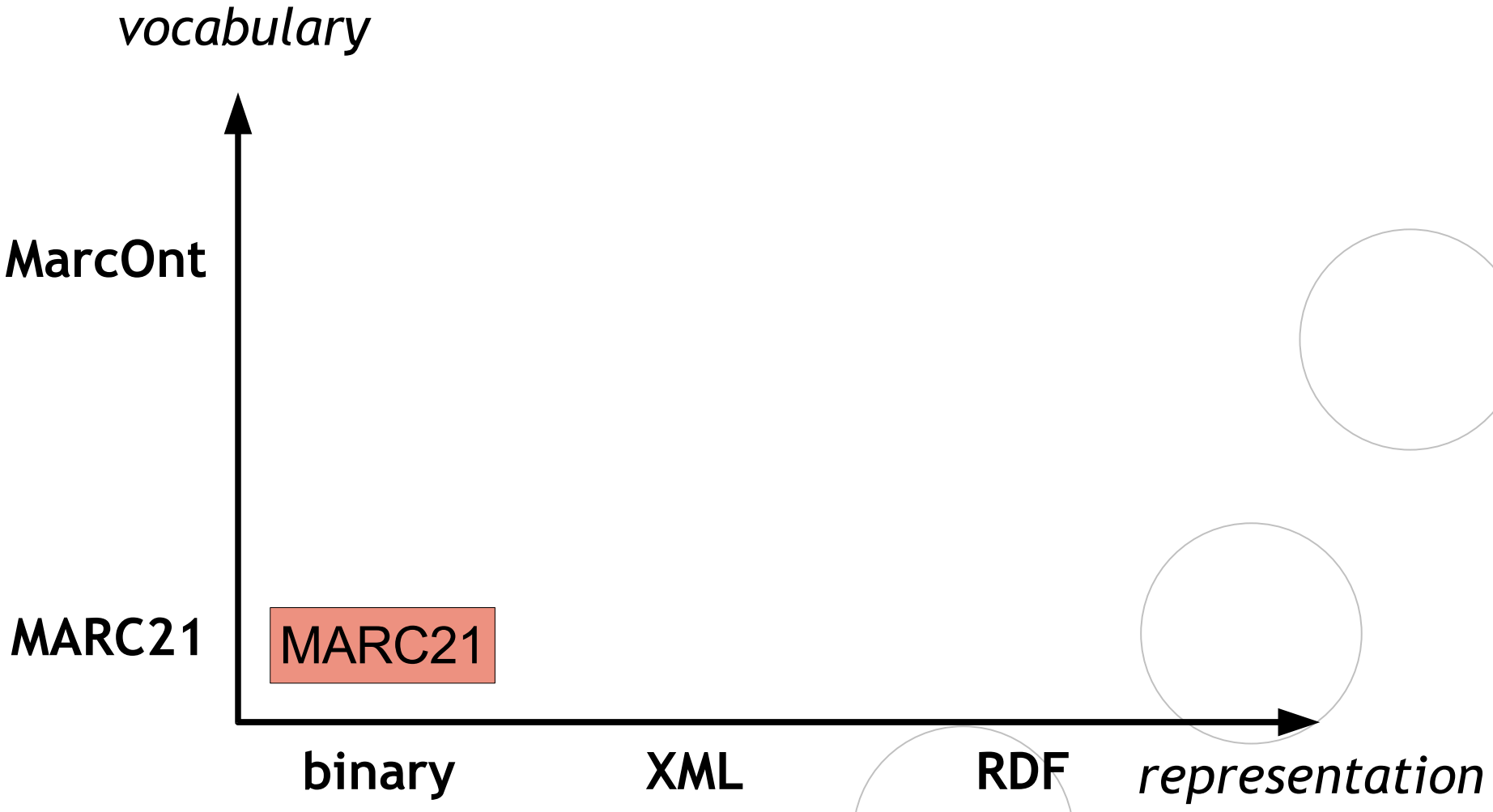
MarcOnt Mediation Services



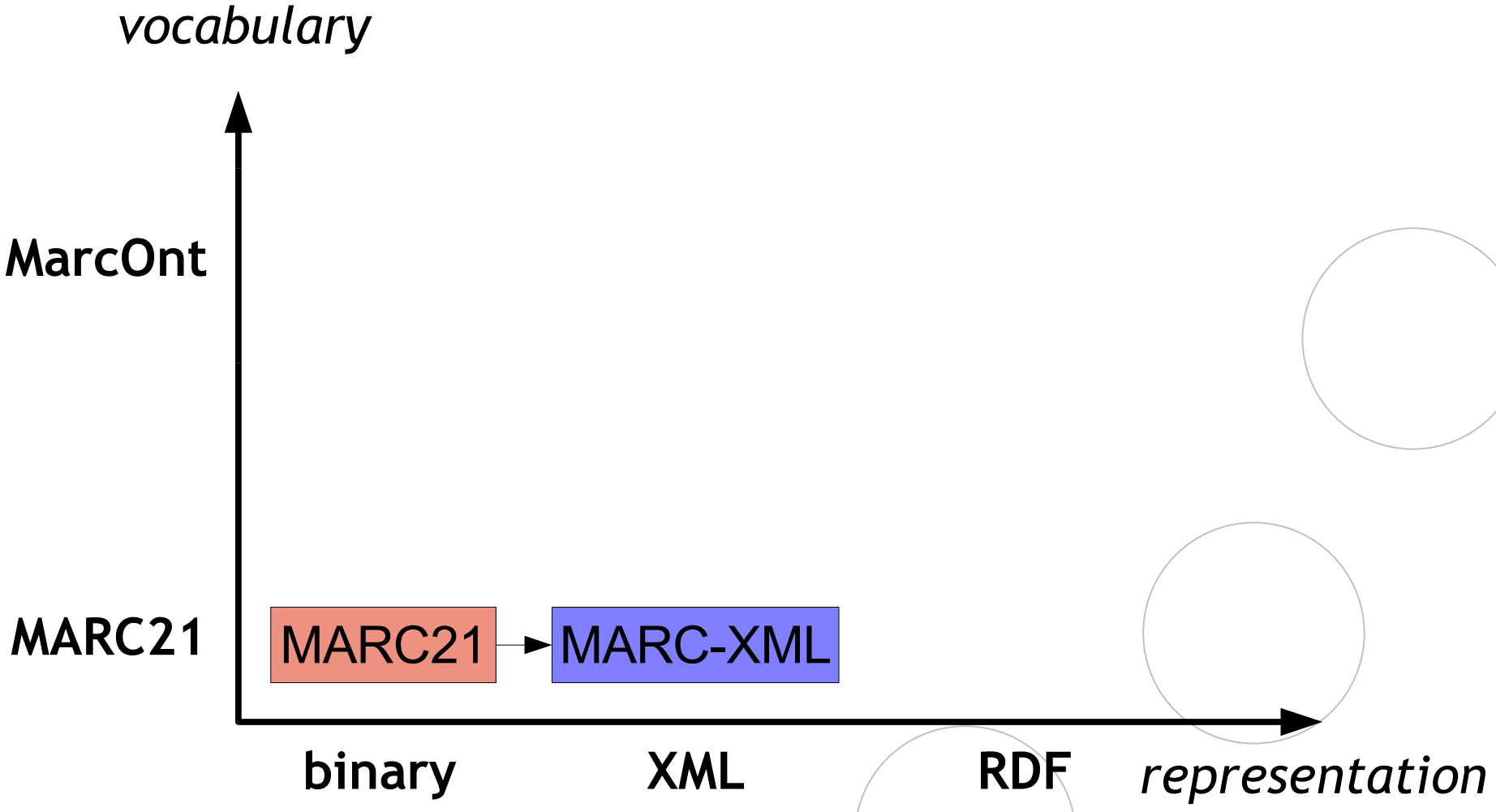
MarcOnt Mediation Services



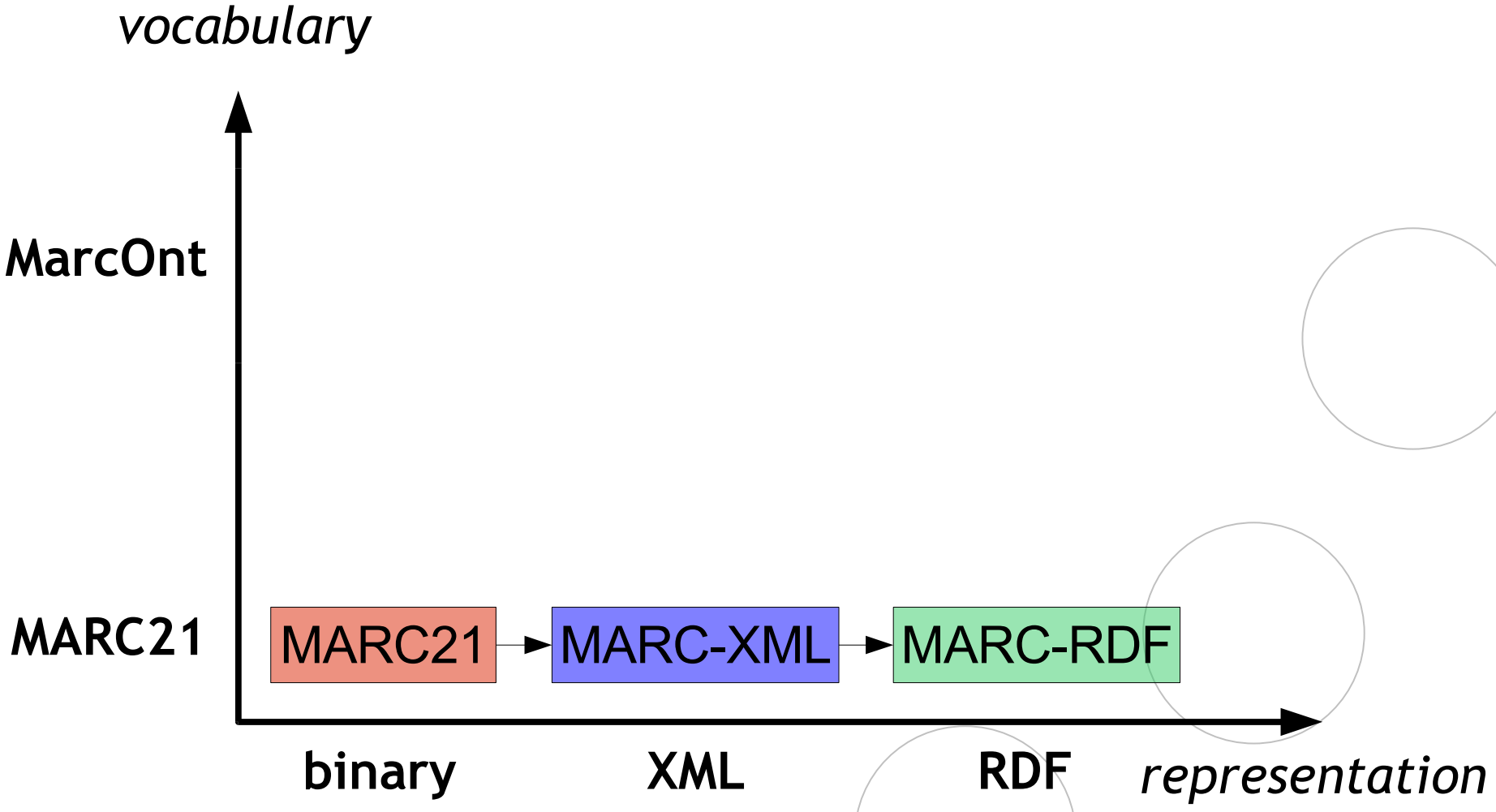
From Legacy Description to MarcOnt Ontology



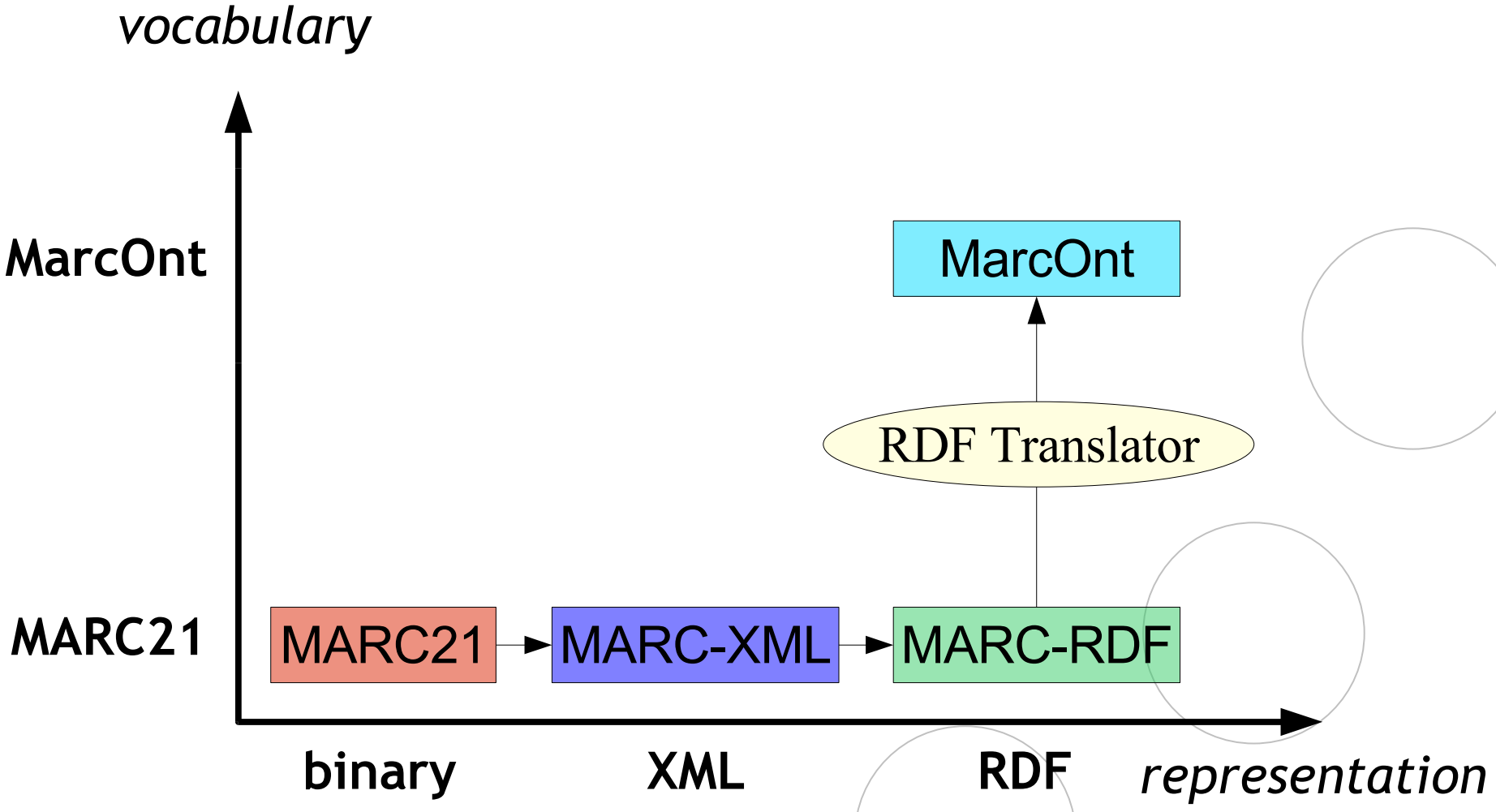
From Legacy Description to MarcOnt Ontology



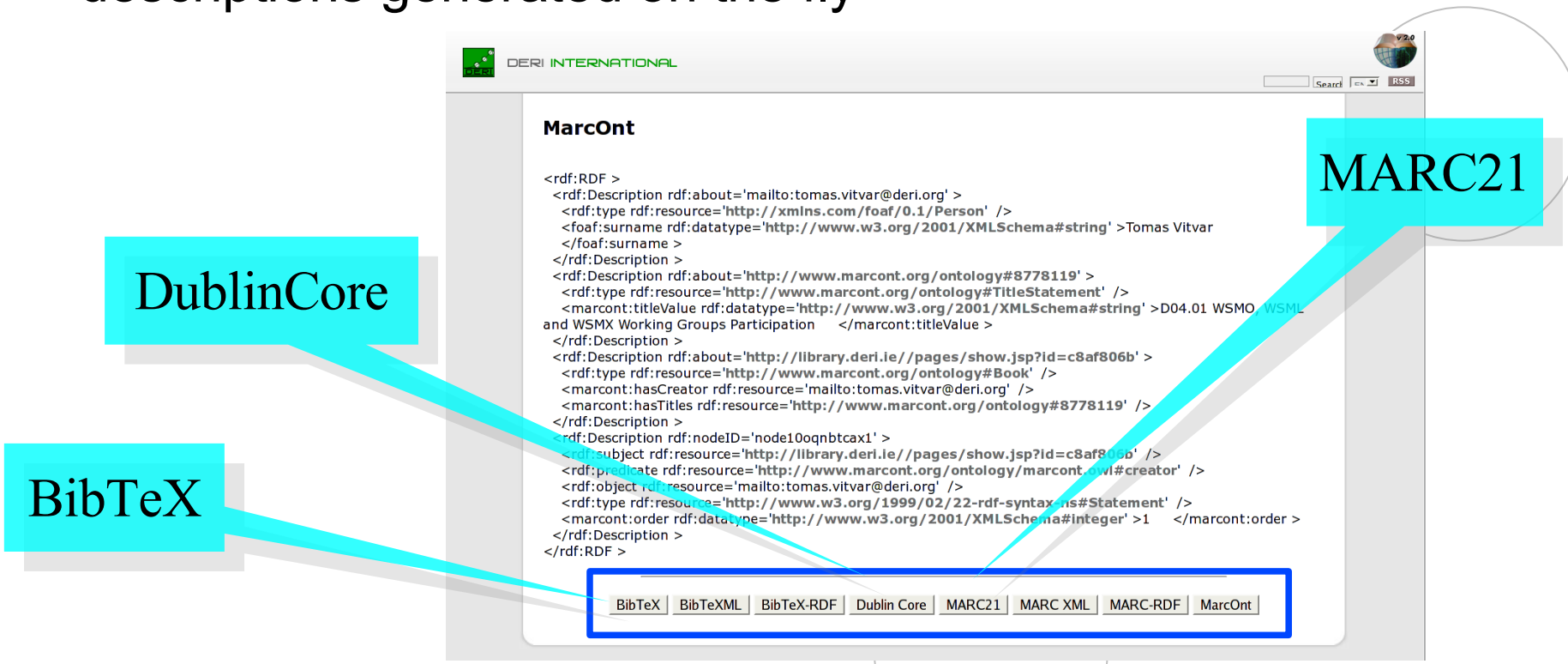
From Legacy Description to MarcOnt Ontology



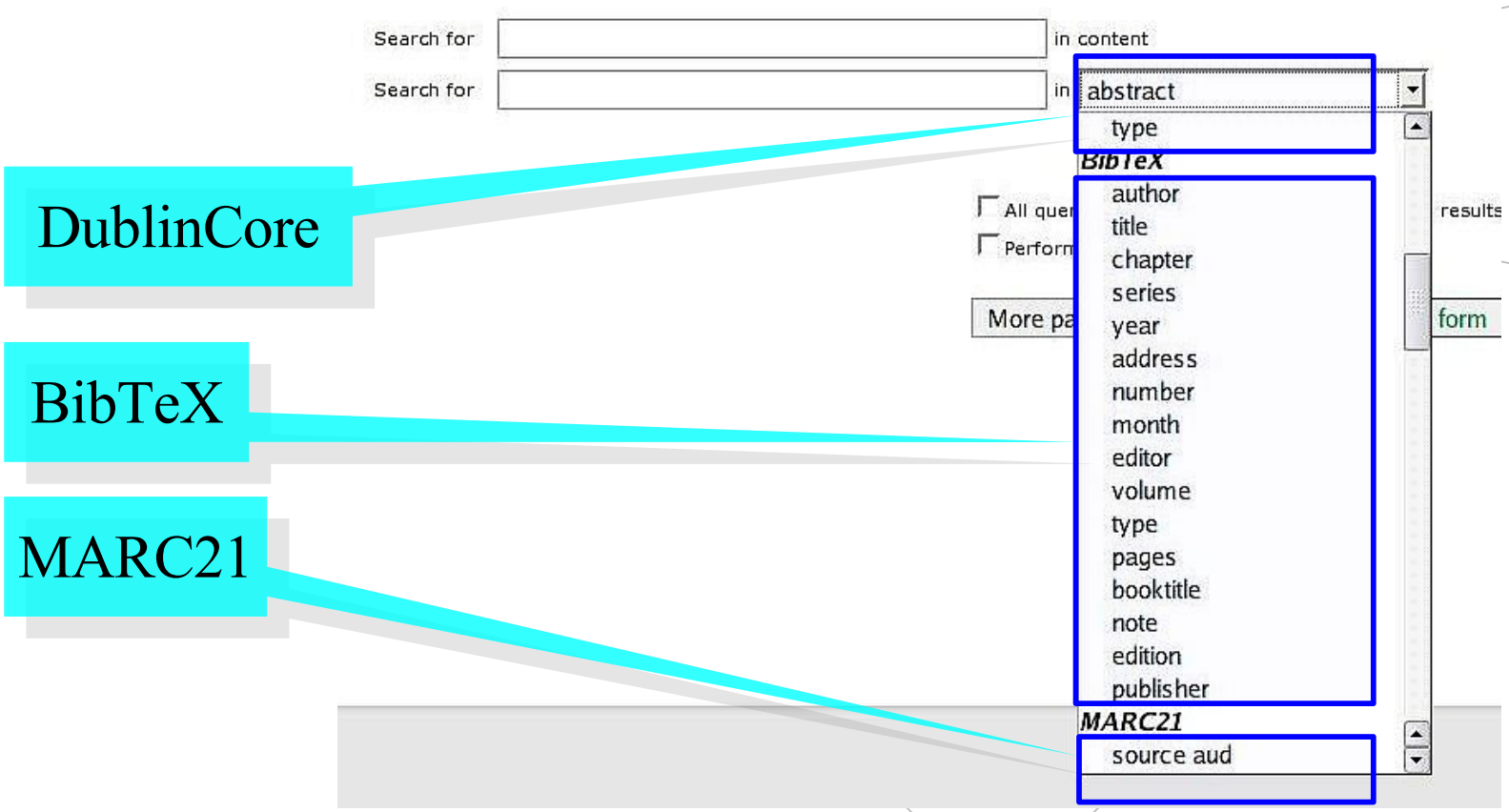
From Legacy Description to MarcOnt Ontology



- All resources are described in **MarcOnt Ontology**, but user can access **MARC21**, **BibTeX** and **DublinCore** descriptions generated on the fly



- User can select from wide range of description **properties**, defined in **different metadata**, during query building



- Performing advanced **RDF queries**

Your query:

```
select * from {a} b {c}
```

Query Language SeRQL Res

Result:

http://www.jeromedl.org/structure#published	"true"^^http://www.w3.org/2001/XMLSchema#boolean
http://www.jeromedl.org/structure#uploadDate	"20050702"^^http://www.w3.org/2001/XMLSchema#integer
http://www.jeromedl.org/structure#abstract	"The core technology area \workÂ°ow management" or \process manageme process meta-model deÂ¯nes its own set of process deÂ¯nition concepts wi process execution semantics. Current practice is to implement a separate int consequently, as many interpreters exist as process meta-models. Various a their process-related function- ality. As a consequence, the IT staÂ® as well required to learn, to use and to main- tain several process management proc procedures. From their viewpoint, it would be tremendously advantageous to domain speciÂ¯c products. The state-of-the-art in process management doe execution models of all process meta-models."^^http://www.w3.org/2001/XML
http://www.marcont.org/bibtexrdf#hasAuthor	"Simeon Petkov"^^http://www.w3.org/2001/XMLSchema#string
http://www.marcont.org/bibtexrdf#hasTitle	"m3pe light - Multi-Meta-Model Process Execution Environment"^^http://www.w
http://www.marcont.org/bibtexrdf#hasYear	"2005"^^http://www.w3.org/2001/XMLSchema#string
http://www.marcont.org/bibtexrdf#hasSchool	"Information Systems Institute Vienna University of Technology, Digital Enterpr Galway"^^http://www.w3.org/2001/XMLSchema#string
http://www.jeromedl.org/structure#href	""^^http://www.w3.org/2001/XMLSchema#string
http://www.jeromedl.org/structure#usageCount	"27"^^http://www.w3.org/2001/XMLSchema#integer
http://www.jeromedl.org/structure#preprint	"true"^^http://www.w3.org/2001/XMLSchema#boolean
http://www.jeromedl.org/structure#published	"true"^^http://www.w3.org/2001/XMLSchema#boolean
http://www.jeromedl.org/structure#uploadDate	"20050803"^^http://www.w3.org/2001/XMLSchema#integer
http://www.jeromedl.org/structure#abstract	"The contemporary Internet offers various services ranging from electronic ne have to register to on-line services. However, most of the authentication and u the registration process must be repeated each time from the beginning, req constraints. Very often, user management systems do not allow user to view actual information gathered about them after registration. To overcome the pr Microsoft Passport have been proposed. In this article we elaborate on poten access to profile data. We present how required security levels in a user ma service. We define how the potential user can benefit from this user manage we present the D-FOAF, a distributed user management system based on F solutions for secure distributed user management system."^^http://www.w3.org

Conclusions

- Different groups of users require different descriptions of resources
- To provide interoperability in heterogeneous networks we need to be able to translate between different metadata without loss of information
- MarcOnt ontology aim to deliver semantic version of existing description metadata
- MarcOnt Mediation Services can translate between various bibliographic description formats
- MMS has been successfully deployed in JeromeDL

Future work

- Build portal for collaborative MarcOnt ontology management and development
- Provide full support for MARC21, BibTeX and DublinCore in MarcOnt ontology and mediation rules
- Deploy MarcOnt ontology in BRICKS project
- Build network of heterogeneous networks based on the MarcOnt ontology

Thank you for your attention

Sebastian Ryszard Kruk
http://www.MarcOnt.org/