

Role of Ontologies in Semantic Digital Libraries

Sebastian Ryszard Kruk,

Bernhard Haslhofer, Piotr Piotrowski, Adam Westerski, Tomasz Woroniecki

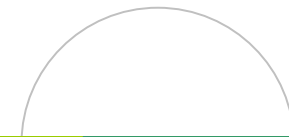
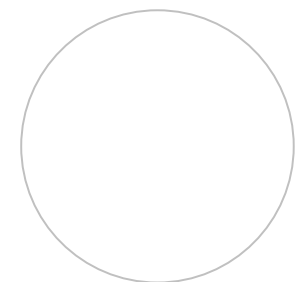
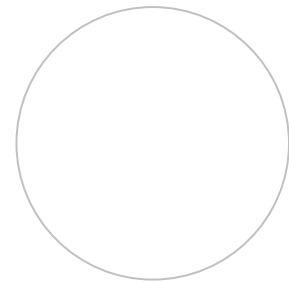
Sebastian Ryszard Kruk
sebastian.kruk@deri.org

NKOS Workshop, 2006
21/09/2006
Alicante, Spain

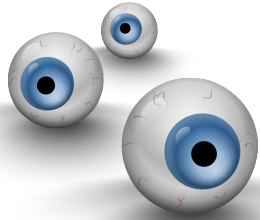
Takeaway home message



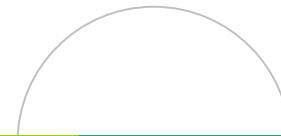
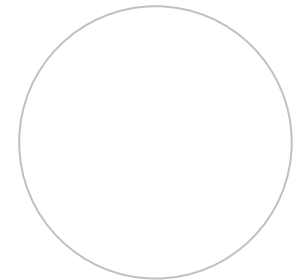
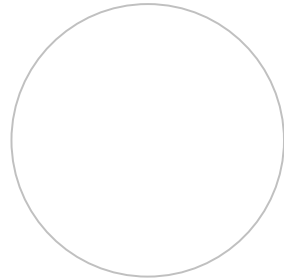
- Ontologies play the key role in the semantic digital libraries
- We need ontologies supporting the main aspects of contemporary knowledge repositories:
 - Bibliographic descriptions
 - Extensible structure of resources
 - Community-aware features



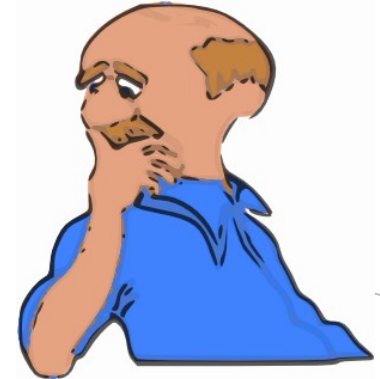
Presentation overview



- Motivations
- What is a semantic digital library?
- Bibliographic Ontology
- Structure Description Ontology
- Community-aware Ontology
- Ontologies in JeromeDL
- Future – Mash-up Digital Libraries
- Conclusions



- Networks of digital libraries linger for more expressive interoperability solutions than existing ones
- Community-aware solutions change the face of the Internet as we knew it
digital libraries should be a part of these changes
- Multimedia play bigger and bigger role on the Internet, while there is a need for accessible and adaptive access solutions

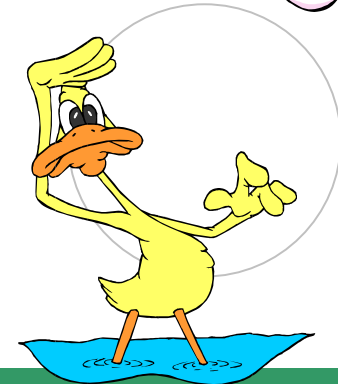
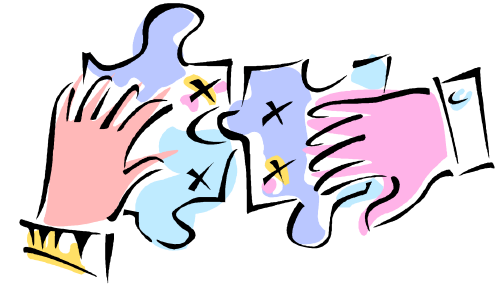


What is a Semantic Digital Library?



Semantic digital libraries

- **integrate** information based on different metadata, e.g.: resources, user profiles, bookmarks, taxonomies
- provide **interoperability** with other systems (not only digital libraries) on either metadata or communication level or both
- delivering more robust, **user friendly and adaptable search and browsing** interfaces empowered by semantics



Metadata is the key concept

- the Web **does not have** metadata
 - the idea of a Semantic Web is nice but difficult to implement
- many digital libraries **do have** metadata in place

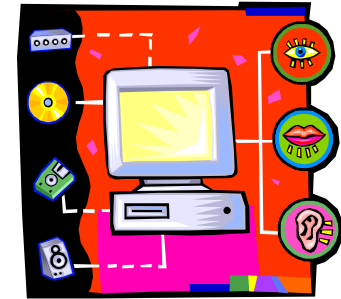


RDF:

- is a framework to model any kind of metadata
- it delivers certain level of technical interoperability

- **Thesauri & Controlled Vocabularies**

- qualified DublinCore
- DMOZ, DDC-based taxonomies
- SKOS, WordNet and other thesauri



- **Schema Mappings / Crosswalks**

- MarcOnt Ontology – aims to cover concepts from MARC21, BibTeX and DublinCore
- MarcOnt Mediation Services – an open mediation framework between common legacy metadata standards

- **Metadata Integration**

- RDF as a common data model for integrating metadata from various autonomous and heterogeneous data sources
- OWL for modeling the data source's semantics
- SPARQL as a common query language

Semantic DL as Evolving Knowledge Space



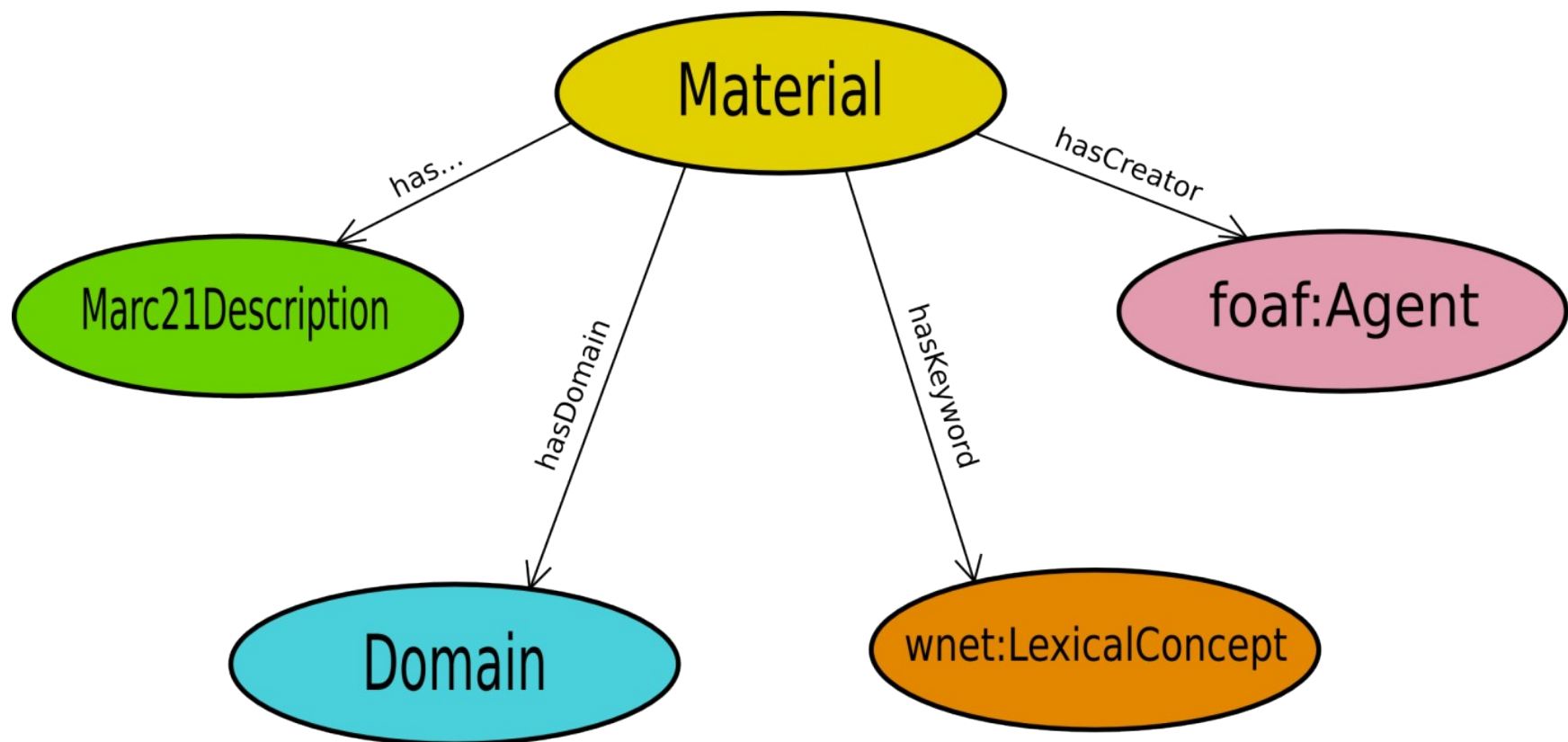
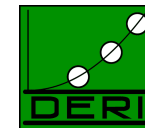
- In state-of-the-art digital libraries users are **consumers**:
Retrieve contents based on available bibliographic records
- Recent trends: user communities
 - Connetea, del.icio.us
 - Flickr
- In Semantic digital libraries users are **contributors** as well
 - Tagging (Web 2.0), Annotations
 - Social Semantic Collaborative Filtering
- Semantic Digital libraries enforce the transition from a static information to a **dynamic (collaborative) knowledge space**

Bibliographic Ontologies



- Build to capture the semantics of the legacy metadata
- Examples of bibliographic ontologies:
 - MarcOnt ontology aiming at capturing concepts from MARC21 and BibTeX
 - RDF Schema for FRBR

MarcOnt Ontology – Main Concepts



Generating various bibliographic descriptions



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

The screenshot shows a web page titled "MarcOnt" with a search bar and an RSS icon. The main content area displays RDF data for a person named Tomas Vitvar. The data includes information about his affiliation (DERI INTERNATIONAL), his name, and a book titled "D04.01 WSMO, WSML and WSMX Working Groups Participation". The book is linked to a library page at deri.ie. At the bottom of the page, there is a navigation menu with buttons for "BibTeX", "BibTeXML", "BibTeX-RDF", "Dublin Core", "MARC21", "MARC XML", "MARC-RDF", and "MarcOnt". The "MARC21" button is highlighted with a blue border. Three arrows point from the text labels "DublinCore", "BibTeX", and "MARC21" to their respective buttons in the navigation menu.

```
<rdf:RDF >
<rdf:Description rdf:about='mailto:tomas.vitvar@deri.org' >
<rdf:type rdf:resource='http://xmlns.com/foaf/0.1/Person' />
<foaf:surname rdf:datatype='http://www.w3.org/2001/XMLSchema#string' >Tomas Vitvar
</foaf:surname >
</rdf:Description >
<rdf:Description rdf:about='http://www.marcont.org/ontology#8778119' >
<rdf:type rdf:resource='http://www.marcont.org/ontology#TitleStatement' />
<marcont:titleValue rdf:datatype='http://www.w3.org/2001/XMLSchema#string' >D04.01 WSMO, WSML
and WSMX Working Groups Participation </marcont:titleValue >
</rdf:Description >
<rdf:Description rdf:about='http://library.deri.ie//pages/show.jsp?id=c8af806b' >
<rdf:type rdf:resource='http://www.marcont.org/ontology#Book' />
<marcont:hasCreator rdf:resource='mailto:tomas.vitvar@deri.org' />
<marcont:hasTitles rdf:resource='http://www.marcont.org/ontology#8778119' />
</rdf:Description >
<rdf:Description rdf:nodeID='node10oqnbtcax1' >
<rdf:subject rdf:resource='http://library.deri.ie//pages/show.jsp?id=c8af806b' />
<rdf:predicate rdf:resource='http://www.marcont.org/ontology/marcont.owl#creator' />
<rdf:object rdf:resource='mailto:tomas.vitvar@deri.org' />
<rdf:type rdf:resource='http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement' />
<marcont:order rdf:datatype='http://www.w3.org/2001/XMLSchema#integer' >1 </marcont:order >
</rdf:Description >
</rdf:RDF >
```

Mediation service used during searching



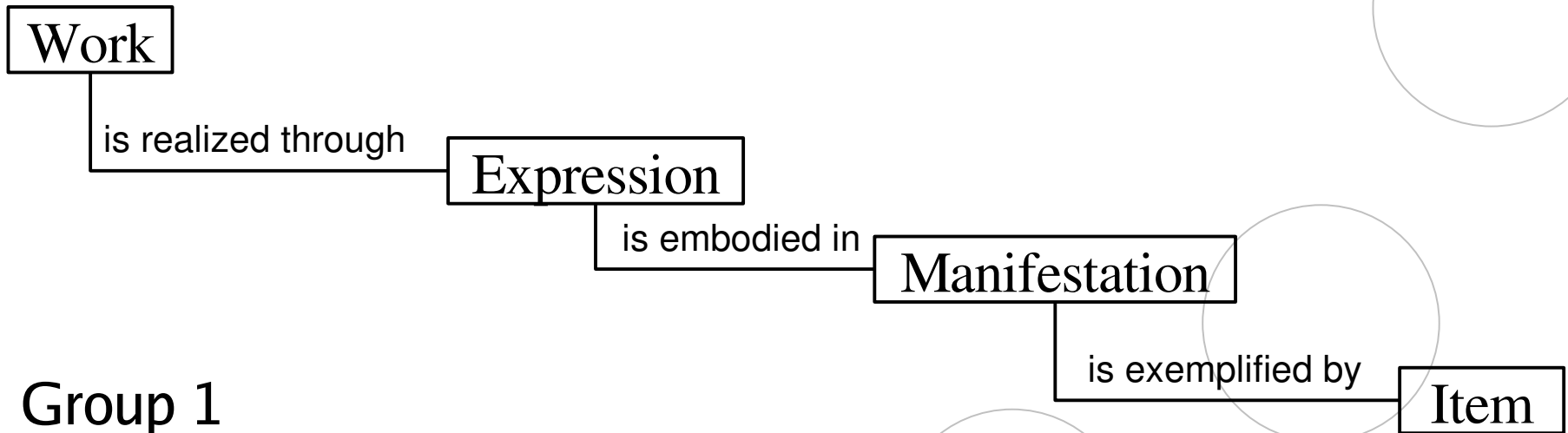
- User can select from wide range of description **properties**, defined in **different metadata**, during query building

A screenshot of a search interface. At the top, there are two search input fields labeled 'Search for' and 'in content'. Below them is a dropdown menu showing a list of properties. The list is divided into three sections: 'DublinCore' (with 'abstract' and 'type' highlighted in blue), 'BibTeX' (with a list of properties including 'author', 'title', 'chapter', 'series', 'year', 'address', 'number', 'month', 'editor', 'volume', 'type', 'pages', 'booktitle', 'note', 'edition', and 'publisher'), and 'MARC21' (with 'source aud' highlighted in blue). To the right of the dropdown menu, there are checkboxes for 'All queries' and 'Perform', a 'More properties' button, and a 'results' label. At the bottom right, there is a 'form' button. Arrows point from the labels 'DublinCore', 'BibTeX', and 'MARC21' on the left to their respective sections in the dropdown menu.

FRBR and Bibliographic Ontology



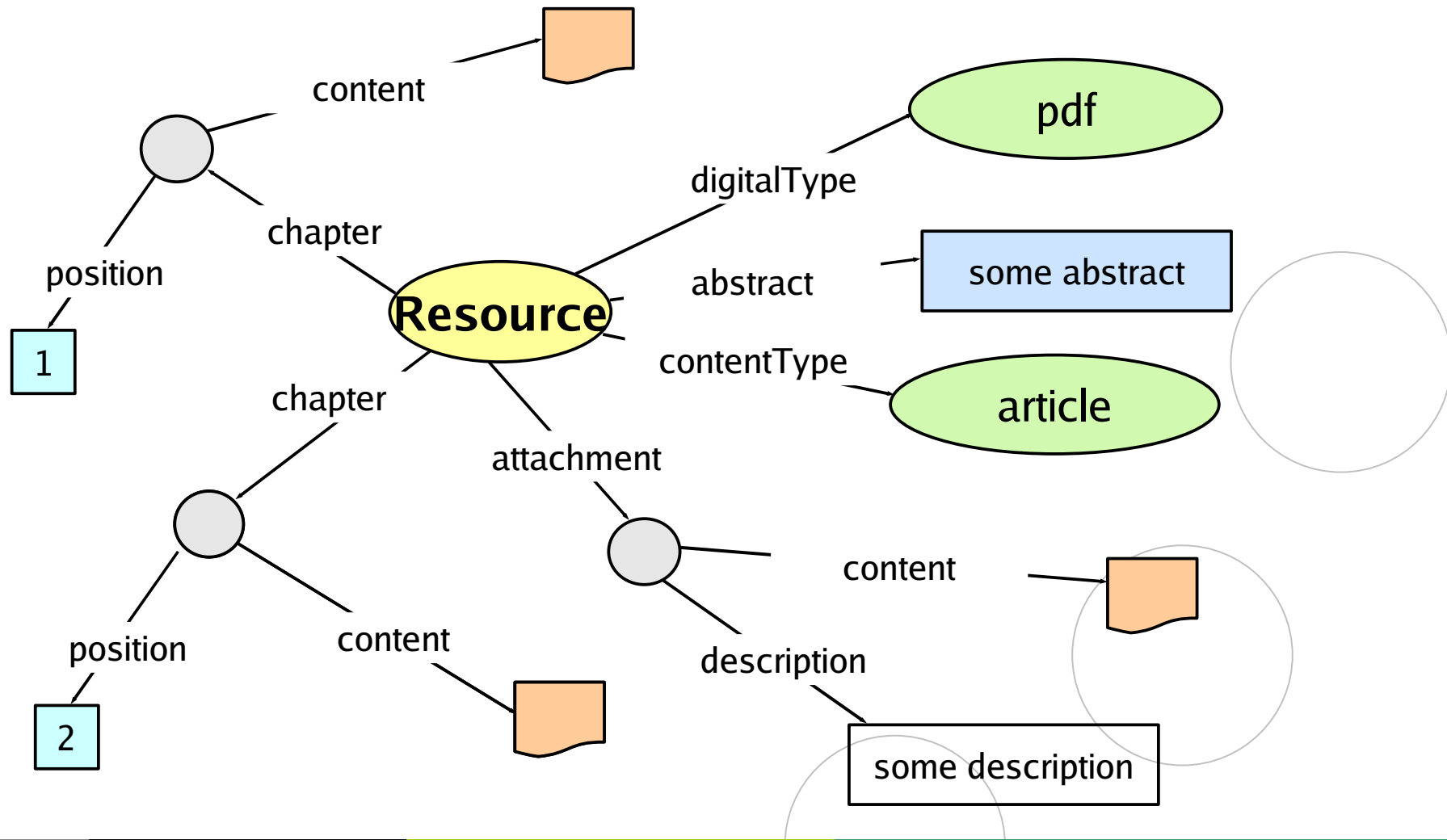
- FRBR was published in 1998 -> **does not address virtual electronic resources** except as downloadable copies of documents
- conceptual model based on **entities-attributed-relationships**



- Semantic Web is based on “**entity**” metadata for resources (documents, people, concepts, etc.)
- FRBR and RDF
 - RDF Schema: <http://vocab.org/frbr/core>
 - 13 distinctive classes
 - 48 properties (most of them with coupled with their inverse counterparts)

- Describing structure of resources with RDF allows flexible content model
- Library resource can be decomposed into sub-resources
- Each part of the resource can be later additionally annotated to deliver:
 - Accessibility features (using e.g. WAI ontology)
 - Adaptive hypermedia (with an appropriate user client)
 - Rendering to different platforms
- Library resource can be:
 - Easily extended with new type of content
 - Versioned and internationalized
 - Decomposed to deliver fine-grained access control

Example of the structure description



Community-aware Ontology



- Bibliographic descriptions were always too complex for average user
- Seems that Semantic Web is suffering similar problems with understanding by larger group of users
- The notion of community-based computing (so called Web 2.0) gains larger and large group of users

Community Tagging



- One of the key aspects of Web 2.0 apart from collaboration is the simplicity of descriptions (so called tagging)
- What do people tag:
 - Resources (URLs): del.icio.us, connotea.org
 - Photos: flickr.com
 - Events: upcoming.org
- How do people tag:
 - Free tagging – any keyword goes
 - Controlled/suggested vocabulary - based on established folksonomy
 - Geo-tagging – drag&drop (Flickr Maps), GPS info (Google Maps)

- A tag itself has no meaning
- A tag within a context of other tags or some actions/states has meaning
- We can come up with an ontology based on:
 - Groups of tags users use
 - Keywords in users' queries
 - Explicitly defined groups of tags (e.g. bundles in del.icio.us)
- Folksonomies are reverse-engineered ontologies of users' tagging/querying actions

What is Social Semantic Collaborative Filtering?



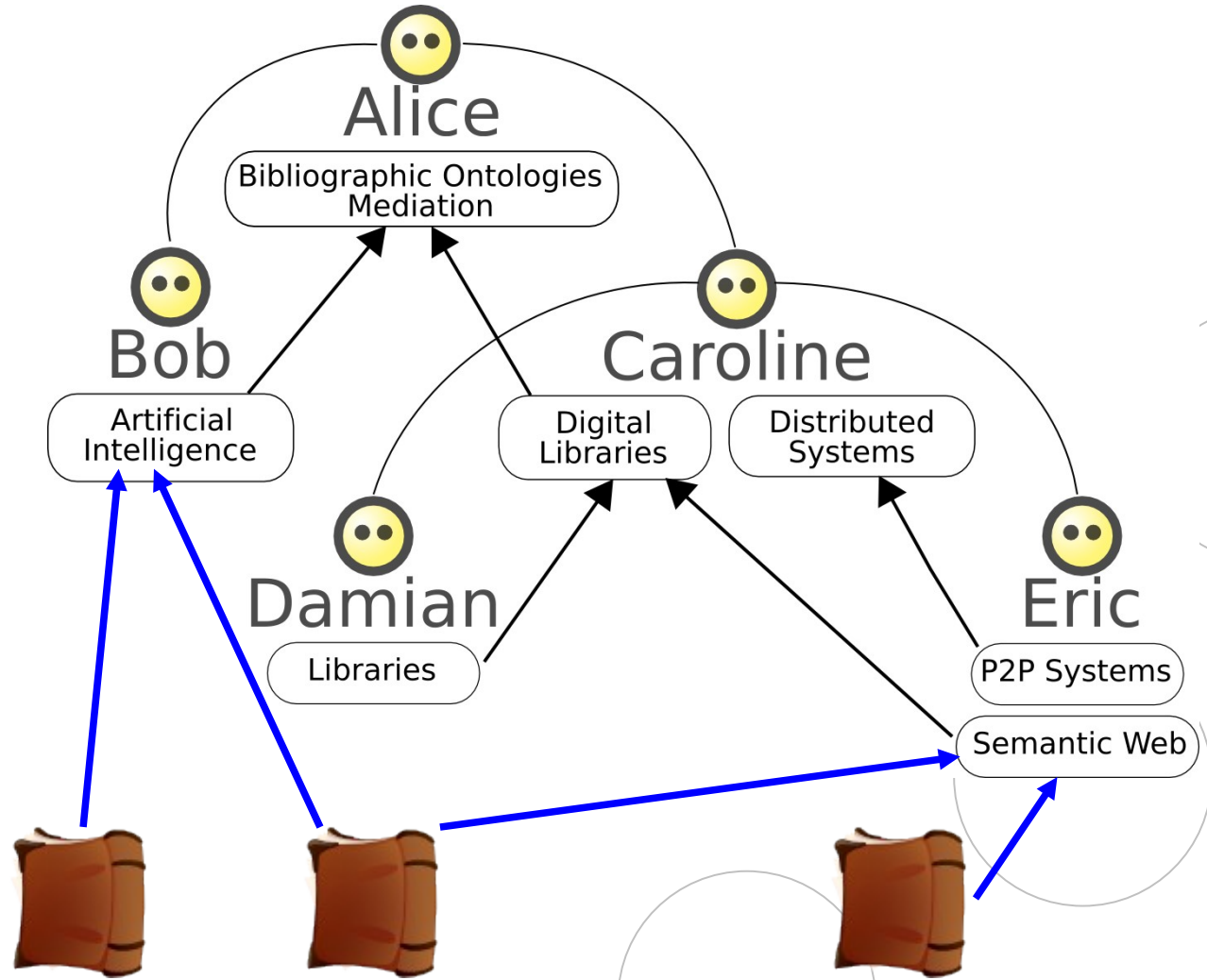
- Goal: *to enhance individual bookmarks with shared knowledge within a community*
- Users annotate catalogues of bookmarks with semantic information taken from taxonomies or thesauri
- Catalogs can include (*transclusion*) friend's catalogues
- Access to catalogues can be restricted with social networking-based policies
- SSCF delivers:
 - Community-oriented, semantically-rich taxonomies
 - Information about a user's interest
 - Flows of expertise from the domain expert

Example of Social Semantic Collaborative Filtering

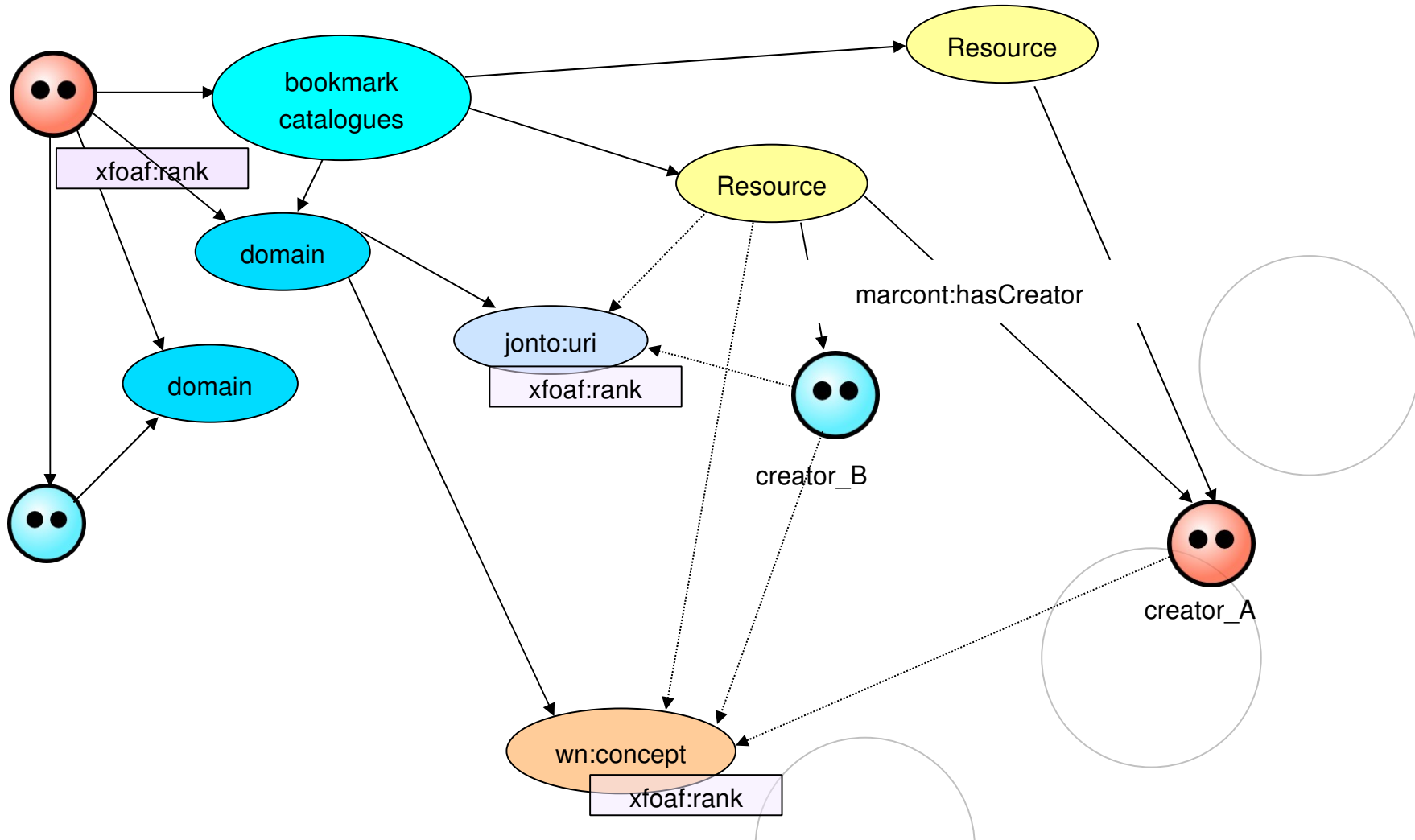


foaf:knows

xfoaf:include
----->
xfoaf:bookmark
----->



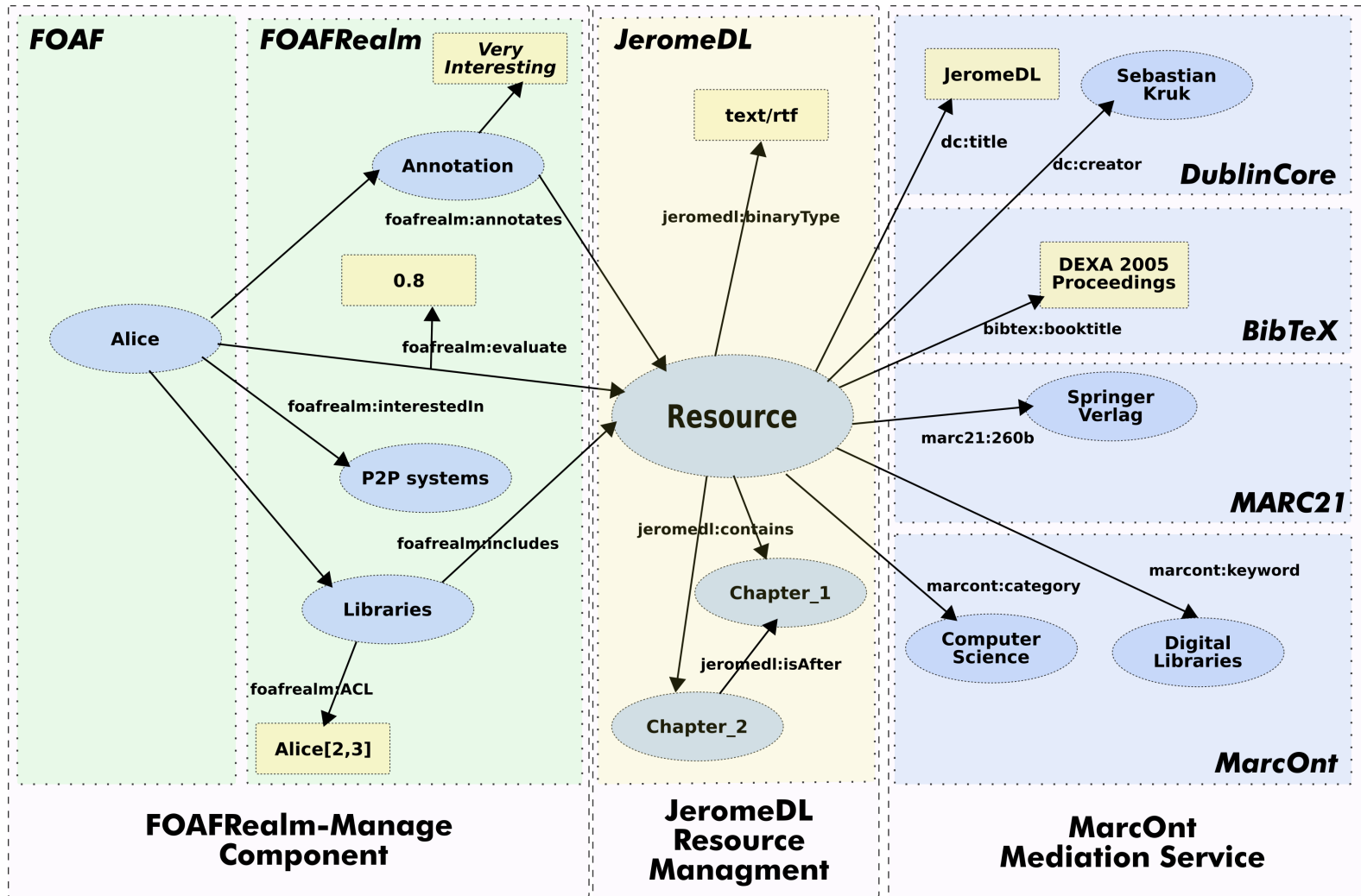
Social Semantic Collaborative Filtering



- Digital Library build with semantics and communities in mind
- Build to reflect requirements of:
 - Librarians
 - Researchers
 - Average users
- Ultimate goal – accessibility achieved through
 - Interface design
 - Search and browsing technologies
 - In-depth internationalization effort

- Structure (system administrators):
 - JeromeDL structure ontology
- Bibliographic and legacy descriptions (domain experts and expert users):
 - MarcOnt bibliographic ontology
 - Extensible MarcOnt suggestions
- Communities (normal users, expert users with restricted vocabulary):
 - FOAF and FOAFRealm identity management ontology
 - Social semantic collaborative filtering (SSCF) ontology
 - Semantical Interlinked Online Communities (SIOC) ontology

Ontologies in JeromeDL



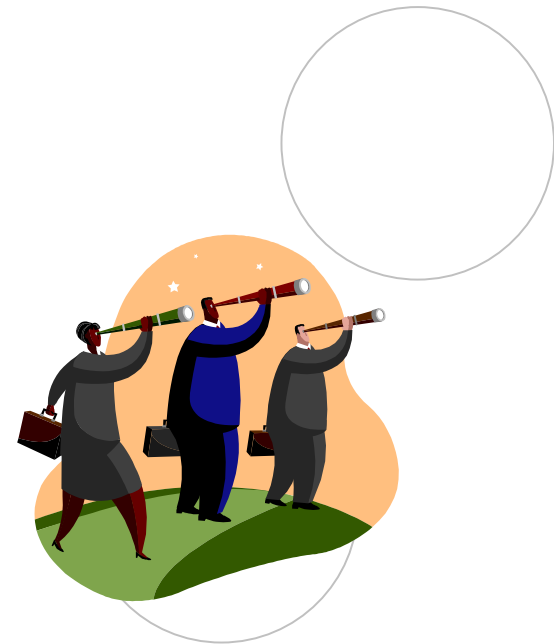
JeromeDL – Delivering Semantic Content



- Providing semantic annotations during uploading process:
 - open module for handling any taxonomies
 - keywords based on WordNet and free tagging
 - defining structure of resources in the JeromeDL ontology
- Lifting legacy metadata to MarcOnt ontology
- Community maintained annotations
 - social semantic collaborative filtering
 - semantic descriptions based on the FOAF metadata



- Keyword-based search with semantic query expansion
- Semantic search:
 - Direct RDF querying
 - Natural language templates
- Social Semantic Collaborative Filtering
- Heterogeneous communication:
 - Bibster
 - A9
 - OAI



Mash-Up Digital Libraries



- Business world aims towards SOA – to easily integrate, choreograph and orchestrate existing services
- Users tend to mash-up various Web 2.0 services to deliver solutions answering their needs
- Community-aware semantic digital libraries can easily become one of the mashed-up services
- But the real challenge is to build mash-up features directly into the digital library to provide users with completely new experience of browsing beyond the resources of DLs.

- Ontologies play a key role in the evolution of digital libraries
- Current streams of Semantic Web and Web 2.0 should and can be combined into the development of digital libraries
- We have identified 3 applications areas for ontologies: legacy, structure and community-aware descriptions
- However the future of DLs seems to lay beyond integration of information, reaching towards integration with other services

Gracias

[thank you]

Questions & Answers

Sebastian Ryszard Kruk
sebastian.kruk@deri.org
<http://www.corrib.org/>
DERI, NUI Galway, Ireland