Semantic Technologies and Tools

G. Philip Rogers, PMP

Senior Business Analyst, School of Public Health,
Instructional and Information Systems, UNC Chapel Hill

g.philip.rogers@unc.edu

Doctoral Student, Information Science, UNC

http://www.ils.unc.edu/~gerogers/

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Metadata Research Center <MRC>

DCMI Tools Working Group

UNC SCHOOL OF PUBLIC HEALTH

- **Professional experience**. Over the past 20 years, worked in Business Analyst, Project/Program Management, Technical Communications, and Management roles (before joining UNC, worked for Cisco Systems, Web startup, Intel Corporation, USAF).

- **Academic interests**. Doctoral student in UNC's School of Information and Library Science – academic interests include metadata interoperability, Semantic Web, Business Intelligence, and the role of IT as an enabler for addressing major research challenges.

- Connotea.
- http://www.connotea.org/how
- Go to a web page > click Add to Connotea > Add keywords.
- http://www.connotea.org/wiki/WebAPI
- Provides API for building services (libraries currently available for Perl and Ruby). Currently supported services:
  - Retrieving a list of bookmarks (bookmarks are shared)
  - Retrieving a list of posts (bookmarks are unique to individual users)
  - Retrieving a list of tags
  - Creating a new post
  - Editing an existing post
  - Removing an existing post

- Web API queries are based on a set of filtering criteria. Many of the filtering criteria are analogous to DC elements:

  – user. Specifies Connotea user to filter on

  – tag. Specifies tag name

  – date. Specifies the date

  – uri. Specifies the URL (or an MD5 hash of a URL)

Retrieving Data from Connotea

- Query results are returned in the body of the HTTP response, which can be parsed as RDF or XML.
- All API operations are standard HTTP requests (GET the relevant URL).
- The base URL for API operations is http://www.connotea.org/data
- This is followed by an indicator of the type of data you want, followed by the user, tag and uri filters in a defined order, after which come the search, number and start parameters, which can appear in any order.

So the URL for retrieving data is constructed as follows:

http://www.connotea.org/data
    /bookmarks or /tags or '' (empty string, which means 'posts')
    /user/ [username]
    /tag/ [tagname]
    /date/ [date of form YYYY-MM-DD]
    /uri/ [uri or hash]
    ?
    q= [free text search string]
    &
    num= [number of results per]
    &
    start= [result number to start at]

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:prism="http://prismstandard.org/namespaces/1.2/basic/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns="http://www.connotea.org/2005/01/schema#" >


<dcterms:URI rdf:about=" [ The link out to the original web page. ]">
    <link> [ The link is here too. ] </link>


<dc:title> [ The Bookmark title -- i.e. the title found in the HTML. ]
</dc:title> <!-- list of tags assigned to this URL by all users who've
    bookmarked it -->
```

<citation>
<rdf:Description> <prism:title>[Title of the reference. This may be different from dc:title above.] </prism:title>

<!-- list of authors -->
 <foaf:maker> <foaf:Person> <foaf:name> [Name of author 1] </foaf:name>

<dc:date> [Publication date for the reference.] </dc:date>

<dc:identifier>[The DOI in the form doi:xx.xxxx/xxxxxx] </dc:identifier>
<pmidResolver rdf:resource=" [If the article has a Pubmed ID, a link to its entry in the Pubmed database will appear here.] " />
<dc:identifier> [The Pubmed ID in the form PMID: xxxxxxxx] </dc:identifier> </rdf:Description> </citation>

- Large enterprises are starting to take notice and invest in Web 2.0 functionality. For example, the IBM Dogear project:

http://domino.watson.ibm.com/cambridge/research.nsf/9975

"Despite widespread usage, there have been few, if any, social bookmarking systems appropriate for large organizations and businesses. The goal for the dogear project was to design and build a world-class social bookmarking service aimed at the enterprise, which explored these ideas. Dogear exploits the enterprise by allowing people to bookmark pages within their Intranet. In addition it uses enterprise directories to authenticate the user's identity. This allows people to find experts on specific topics within the company."

- IBM is very active in this area. For example:

- IBM Web Ontology Manager (repository for managing ontologies expressed in OWL, web interface for ontology upload, interface for generating Java APIs from uploaded ontology files)

  http://www.alphaworks.ibm.com/tech/wom

- IBM Integrated Ontology Development Toolkit. Principal components include:
  - **EMF Ontology Definition Metamodel (EODM)** – derived from OMG Ontology Definition Metamodel (ODM), implemented in Eclipse Modeling Framework (EMF); includes RDFS/OWL parsing and serialization, reasoning, and transformation between RDFS/OWL and other data-modeling languages.
  - **EODM workbench** – Eclipse-based environment for building, managing, and visualizing ontologies.
  - **OWL Ontology Repository (Minerva)** -- RDBMS-based storage, interference, and query system.
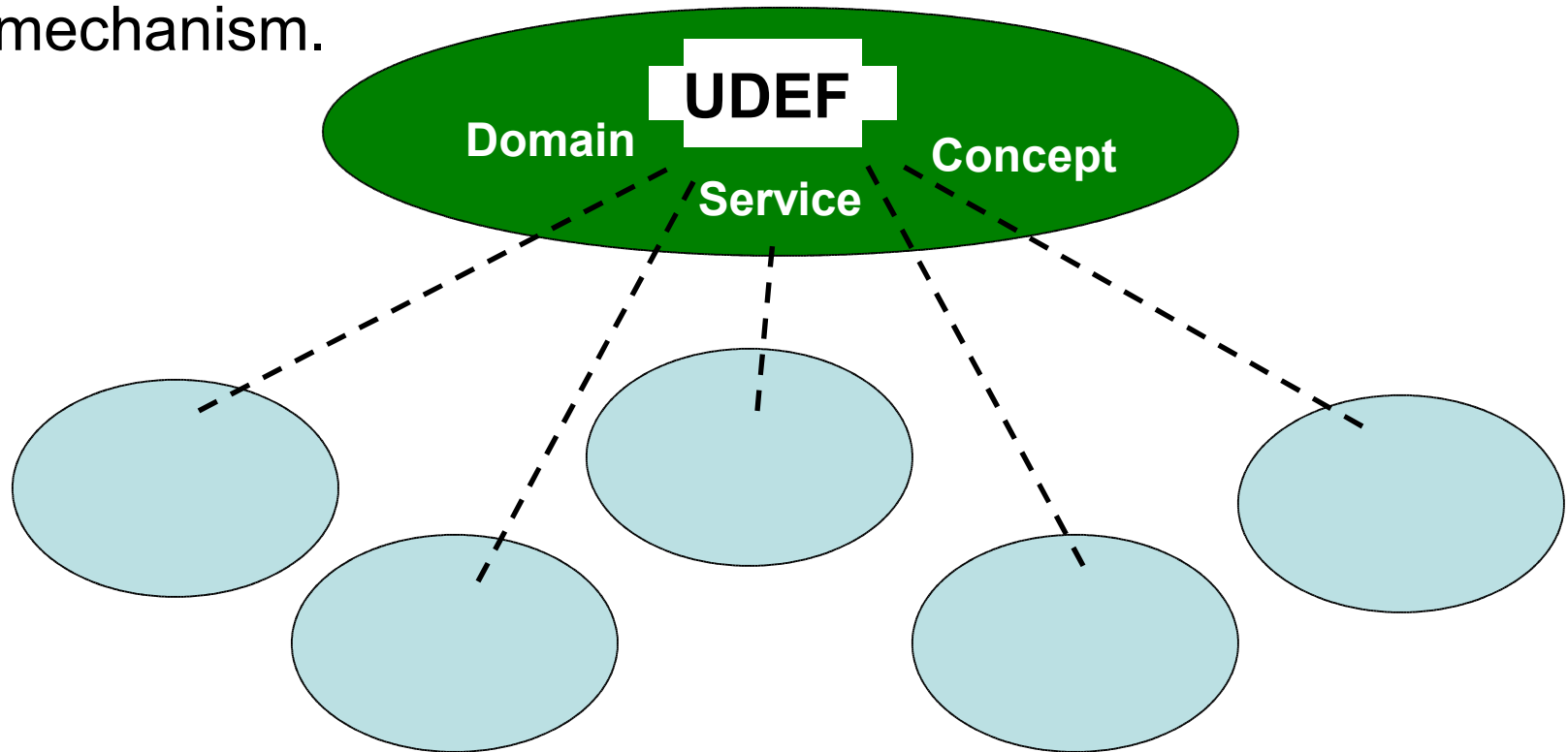
- OS's supported: Windows, Linux

http://www.alphaworks.ibm.com/tech/semanticstk

SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

UNC
SCHOOL OF
PUBLIC HEALTH

Backup Slides

Metadata Research Center <MRC>

Semantic DNS

- The Open Group proposes an interoperability solution based on "semantic DNS" using UDEF as indexing mechanism.

Domain Ontologies

- Train people in different communities to classify their data in a consistent way
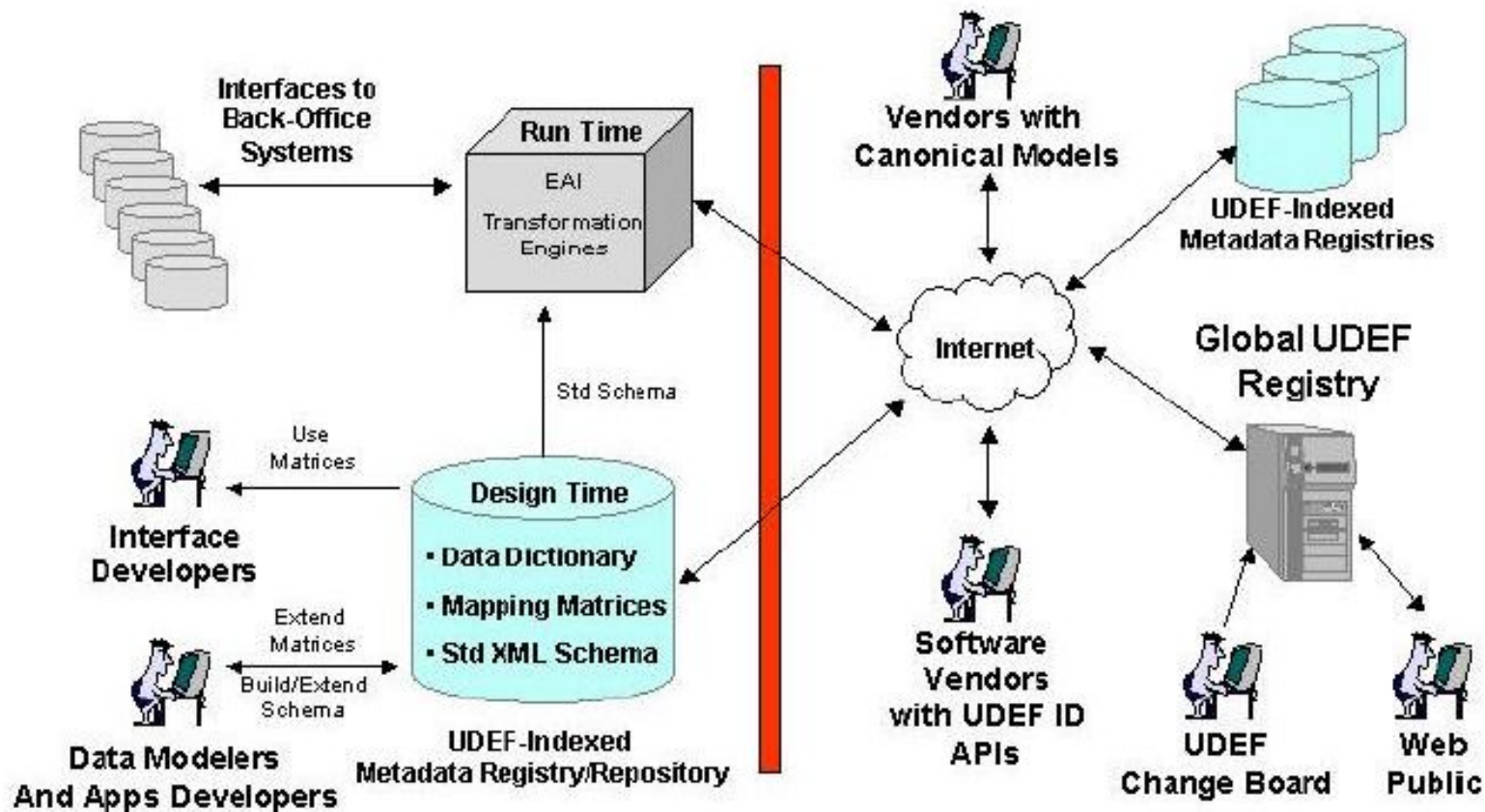- Leverage the resulting ontologies for machine processing.
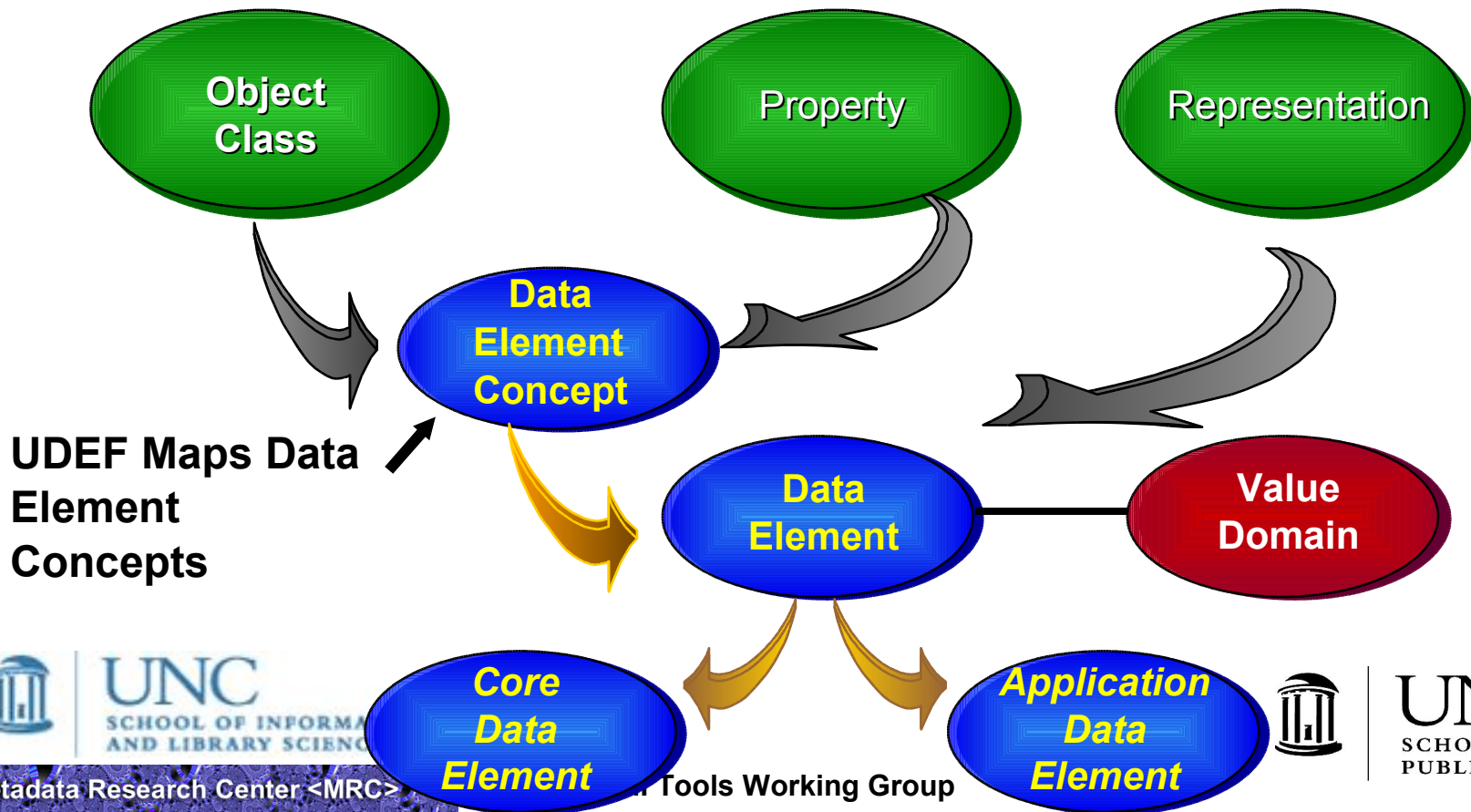
1. Identify applicable UDEF property word that characterizes the dominant attribute (property) of the data element concept (e.g., Name, Identifier, Date),

2. Identify dominant UDEF object word that dominant property (selected in step 1) describes (e.g., Person_Name, Product_Identifier, Document_Date).

3. Reviewing UDEF tree for selected property (identified in step 1), identify applicable qualifiers that can unambiguously describe property word term (e.g., Last Name).

4. Reviewing UDEF tree for selected object (identified in step 2), identify applicable qualifiers that can unambiguously describe object word term (e.g, Customer Person).

5. Concatenate object term and property term to create UDEF naming convention compliant name (e.g., Customer Person_Last Name).

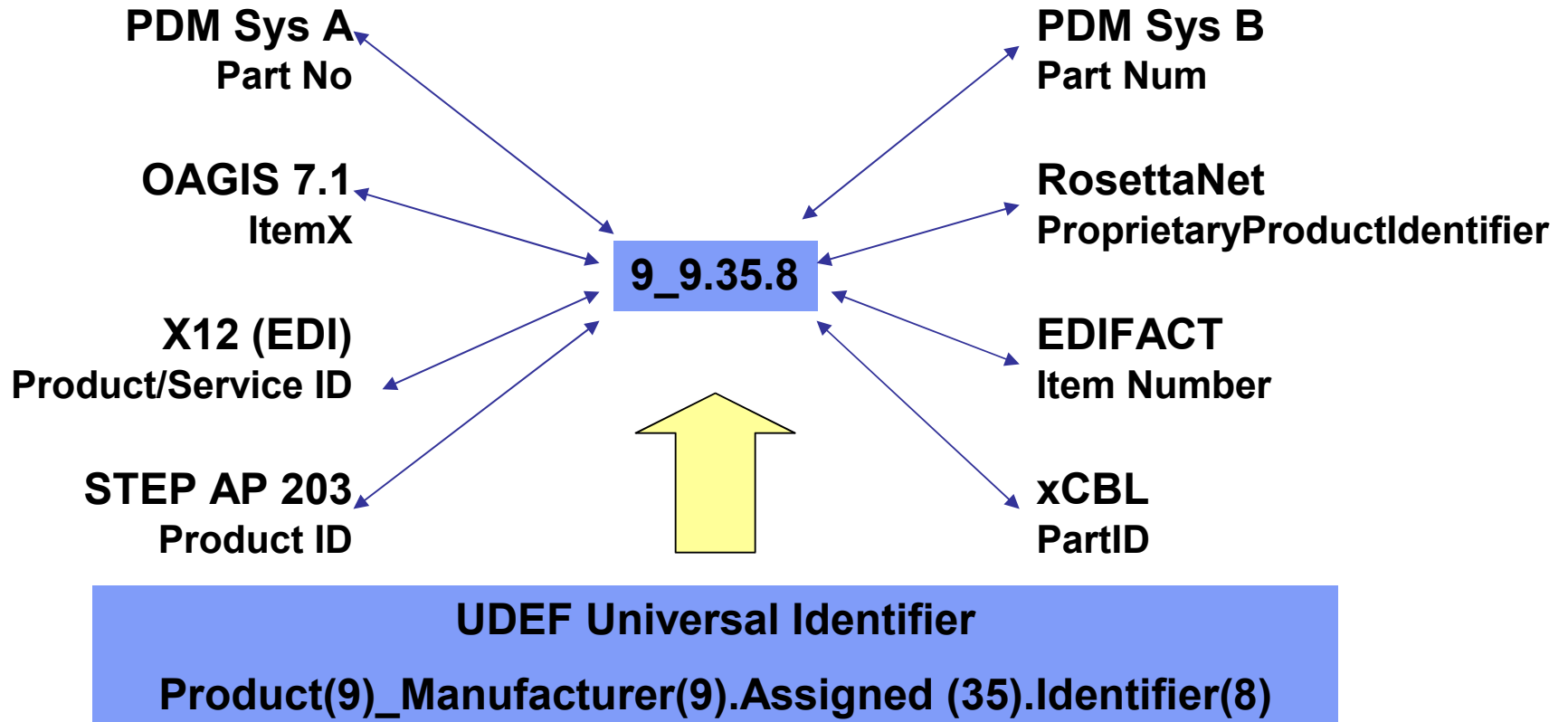6. Derive structured ID based on UDEF taxonomy that carries UDEF inherited indexing scheme, e.g., <CustomerPersonLastName GUID="as.5_5.10">.

A Metadata Managed Architecture

DCMI Tools Working Group

Sample UDEF Mappings

PDM Sys A
Part No

OAGIS 7.1
ItemX

X12 (EDI)
Product/Service ID

STEP AP 203
Product ID

9_9.35.8

PDM Sys B
Part Num

RosettaNet
ProprietaryProductIdentifier

EDIFACT
Item Number

xCBL
PartID

UDEF Universal Identifier

Product(9)_Manufacturer(9).Assigned (35).Identifier(8)

DCMI Tools Working Group

- Emergency response team arrives at a disaster site without batteries to operate GPS system and other communications devices for search and rescue workers.

- Where Semantic DNS comes in: determine whether resources might be available nearby within a manufacturer's inventory.

- Use pair of UDEF tags (IDs) to locate available resources

- Use UDEF tags to support semantic integration of separate procurement applications that use different purchase order semantics

- Leverage OASIS Emergency Management TC defined message wrappers

- Open Group UDEF Forum defines message content

- As of May 2005, UDEF Officially Adopted by The Open Group. Next step is to establish UDEF as the universally-used categorization system for data.
  - Develop, maintain UDEF as open standard
  - Advocate and promote UDEF
  - Put in place a technical infrastructure to support it
  - Implement a Registry for UDEF
  - Set up education programs to train information professionals in its use.

UDEF Trees on the web:

http://www.opengroup.org/udefinfo/defs.htm

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

UNC SCHOOL OF PUBLIC HEALTH

Metadata Research Center <MRC>

DCMI Tools Working Group