LINKED DATA IN THE CURRICULUM OF THE DILL INTERNATIONAL MASTER

Anna Maria Tammaro, Universita’ di Parma
Vittore Casarosa, ISTI-CNR, Pisa
Carlo Meghini, ISTI-CNR, Pisa

Roma 20 Febbraio 2014
International Master financed for five years (2006-2011) by the Erasmus Mundus Program

Three partners (three Master degrees)
- Oslo Akershus University
- Tallin University
- Parma University

Master thesis in one of the three

Presently being continued without Erasmus funding

What is a Digital Library?
DILL Topics

FIRST YEAR
Epistemology of science
Research Methods

SECOND YEAR
Research methods

Digital document
Information management
Human resources management
Users and uses
Access to digital library
Dissertation
DILL Topics

FIRST YEAR
Epistemology of science Research Methods

SECOND YEAR
Research methods

Topics more related to IT

Digital document
Information management
Human resources management
Users and uses
Access to digital library
Dissertation
Memory institutions are key players in providing knowledge:
- this is their mission
- their knowledge is trusted and of high quality

Nowadays, knowledge is shared on the web
- human consumable knowledge is expressed in natural languages and shared via HTML documents
- machine consumable knowledge expressed in RDF
- shared through Linked Data

Memory institutions have a key role to play in Linked Data

Libraries, in particular, can offer their knowledge to the rest of the world by:
- encoding it in RDF
- using standard vocabularies for classes and properties
- using well-known URIs for naming resources such as people, places, times, concepts, events
- providing URIs for their own resources so that other institutions can use them
Rappresentazione testuale (Turtle):

Rappresentazione a grafo:
Open Data

- make your stuff available on the Web (whatever format) under an open license
- make it available as structured data (e.g., Excel instead of image scan of a table)
- use non-proprietary formats (e.g., CSV instead of Excel)
- use URIs to denote things, so that people can point at your stuff
- link your data to other data to provide context

Five star Open Data (Tim Berners Lee)
Simple basic rules (Tim Berners Lee)
1. Use URIs as names for resources.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things
(meta) Methodology

- Development of a methodology for publishing Library data as Linked Data
- The data gathered by the W3C Incubator Group of Library Linked Data was a starting point for this work.
  - pointed out use cases and issues
- The issues were used in conjunction with project reports to come up with questions
- The questions were used as a guideline for in-depth interviews
  - also books, reports and position papers
Methodology

- A 15 steps recipe
- The steps form the basis for different workflows that can be used to publish Linked Data, depending on purpose, data and context
- Data of interest:
  - knowledge organization systems (classification schemes, thesauri)
  - authority files
  - digital contents and their descriptions
  - catalogue data including circulation data sets.
- All these datasets should have links within themselves and should establish outgoing links to many other web resources, in order to attract many incoming links
- “Web Centric Cataloguing”
The 15 steps

1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation
The 15 steps

1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation
1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation
The 15 steps

1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation
The 15 steps

1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation
The 15 steps

1. Motivation
2. Management approval
3. Sorting out the legal and financial issues
4. Assessment of skills & data available
5. Tools assessment and evaluation
6. Dataset analysis
7. URI assignment
8. Vocabulary Modeling
9. Generation of RDF Data
10. Enriching the data
11. Describing the data set
12. Evaluating the Dataset
13. Publishing
14. Incoming links
15. Curation

Roma, 20 Febbraio 2014
Vittore Casarosa – ISTI-CNR, Pisa
Conclusions

- Adopting linked data technologies allows libraries to
  - improve their presence where today’s information is sought (i.e. the web)
  - improve the services offered to their users
  - promote innovative use of the data that the libraries held

- A small numbers of libraries (and even less archives and museums) have embraced the Linked Data paradigm

- Awareness is raising and knowledge is coming