HOW SEMANTIC WEB DIFFERS FROM TRADITIONAL DATA PROCESSING

Slides by Tom Baker and Karen Coyle

#dc2014
characteristics of Semantic Web

- AAA - anyone can say anything about anything
- OWA - open world assumption
- NUNA - non-unique naming assumption
“AAA”: Anyone can say anything about anything

- RDF designed for the Web - to accommodate multiple sources of information reflecting multiple points of view
- Like the “regular” web, anyone can add data, and anyone can link to anything else
“OWA”: Open World Assumption

- Principle: information available at any given time may be incomplete
  - New discoveries, new points of view…
- Closed-World Assumption: information on hand defines the boundaries of what is known
  - Example: personnel database
CWA vs OWA

• CWA: appropriate for capturing all known facts (e.g., personnel databases)

• OWA: appropriate where knowledge or scholarly opinion is a moving target and can be expected to evolve, change, or even to contradict itself.
FRBR

Closed world

Open world

W
E
M
I

E

?
NUNA: Non-Unique Naming Assumption

- Principle: things described in RDF data can have more than one name
- URIs used in RDF as names, and anything may be identified by more than one URI
- Two things are not assumed to be different because they have different names
RDF & OWL & SKOS

• RDF defines basics of Semantic Web:
  ◦ triple: subject – predicate – object
  ◦ classes & properties
  ◦ domains & ranges

• OWL – Web Ontology Language
  ◦ “OWL is a computational logic-based language”
  ◦ adds axiomatic language, e.g. cardinality, equality/inequality, complement/intersection of, disjointness

• SKOS (Simple knowledge organization system)
  ◦ for KOS, like thesauri
  ◦ broader, narrower, preferred & alternate labels
Properties define class of subject (classes cannot define properties)

- In XML, OO, there are “things” with properties
- In RDF, classes add semantics that aid inferencing
  - any subject can be an instance of more than one class, e.g. “father” “teacher” “employee”
Domains and ranges

- RDF domain does not “bind” property to class; it adds semantics to the subject being described
- Any subject can belong to more than one class (if those classes not disjoint)
- RDF properties cannot be limited for use only with members of a specific class
- RDF property is not defined “for a class”
RDF/OWL do not provide data constraints

- RDF/OWL do not themselves specify integrity constraints for data validation and quality control
- Lack of validation mechanism is widely recognized as a gap in the RDF stack
- W3C is addressing this in a new working group
BIBFRAME
RDA in RDF

Properties associated with FRBR classes

Parallel set w/o FRBR

All properties
Example 1: “Orthodox” data using FRBRer properties

``` PREFIX frbrer: <http://frbrer.org/ns#> 
EX:ResourceA
 frbrer:P2001 EX:ResourceB ; # isRealizedThrough
 frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
    # isCreatedByPerson
 frbrer:P3001 "The adventures of Tom Sawyer" .
    # hasTitleOfTheWork

EX:ResourceB
 frbrer:P2002 EX:ResourceA ; # isRealizationOf
 frbrer:P2003 EX:ResourceC ; # isEmbodiedIn
 frbrer:P3011 <http://id.loc.gov/vocabulary/iso639-2/eng>
    # hasLanguageOfExpression

EX:ResourceC
 frbrer:P2004 EX:ResourceB ; # isEmbodimentOf
 frbrer:P3020 "The adventures of Tom Sawyer" ;
    # hasTitleOfTheManifestation
 frbrer:P3055 "1996" . # hasDateOfPublicationOrDistribution
```
Example 1: “Orthodox” data using FRBRer properties

```
ex:ResourceA
  frbrer:P2001 ex:ResourceB ; # isRealizedThrough
  frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ; # isCreatedByPerson
  frbrer:P3001 "The adventures of Tom Sawyer" . # hasTitleOfTheWork
```

ResourceA is a frbrer:Work because frbrer:P2001 has a domain of work.
(So do P2009 and P3001)
Example 1: “Orthodox” data using FRBRer properties

```reasoning
ex:ResourceA
  frbrer:P2001 ex:ResourceB ;  # isRealizedThrough
  frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
    # isCreatedByPerson
  frbrer:P3001 "The adventures of Tom Sawyer" .
    # hasTitleOfTheWork

ex:ResourceB
  frbrer:P2002 ex:ResourceA ;  # isRealizationOf
  frbrer:P2003 ex:ResourceC ;  # isEmbodiedIn
  frbrer:P3011 <http://id.loc.gov/vocabulary/iso639-2/eng>
    # hasLanguageOfExpression
```

ResourceB is a frbrer:Expression because frbrer:P2002 has a domain of Expression as do P2003, P3011.
Example 1: “Orthodox” data using FRBRer properties

ex:ResourceA
    frbrer:P2001 ex:ResourceB ; # isRealizedThrough
    frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
        # isCreatedByPerson
    frbrer:P3001 "The adventures of Tom Sawyer" .
        # hasTitleOfTheWork

ex:ResourceB
    frbrer:P2002 ex:ResourceA ; # isRealizationOf
    frbrer:P2003 ex:ResourceC ; # isEmbodiedIn
    frbrer:P3011 <http://id.loc.gov/vocabulary/iso639-2/eng>
        # hasLanguageOfExpression

ex:ResourceC
    frbrer:P2004 ex:ResourceB ; # isEmbodimentOf
    frbrer:P3020 "The adventures of Tom Sawyer" ;
        # hasTitleOfTheManifestation
    frbrer:P3055 "1996" . # hasDateOfPublicationOrDistribution

ResourceC is a frbrer:Manifestation because frbrer:P2004 has a domain of Expression as do P3020, P3055.
Example 2: “Unorthodox” data with a FRBRer Expression related to more than one FRBRer Work

ex:ResourceB  # FRBRer Expression
  frbrer:P2002 ex:ResourceA ;  # isRealizationOf
  frbrer:P2002 ex:ResourceA1 ;  # isRealizationOf
    # hasLanguageOfExpression

ResourceB is a frbrer:Expression because frbrer:P2002 has a domain of Expression and a range of Work.

FRBRer in OWL gives a maximum cardinality of 1 for frbrer:P2002 – each expression expresses one and only one work. This, however, is not interpreted as an error. The axioms of OWL conclude: because there is only one work, then ResourceA and ResourceA1 are the same work.

This is inferencing vs. constraining.
Example 3: “Unorthodox” data with clashes between disjoint FRBRer domain classes

ex:ResourceA

frbrer:P2001 ex:ResourceB ; # isRealizationOf
frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
    # isCreatedByPerson
frbrer:P3020 "The adventures of Tom Sawyer" .
    # hasTitleOfTheManifestation

ResourceA is a frbrer:Work because frbrer:P2001 has a domain of work.

Resource A is a frbrer:Manifestation because frbrer:P3020 has a domain of manifestation.

Because the owl:Classes frbrer:Work and frbrer:Expression are disjoint, this is declared “inconsistent” by a reasoner, and it may not be possible to process queries against this data. However, nothing prevents anyone from creating and disseminating such data.
Example 4: “Orthodox” BIBFRAME data

ex:ResourceA
  bf:hasInstance ex:ResourceB ;
  bf:creator <http://id.loc.gov/authorities/names/n79021164> ;
  bf:title "The adventures of Tom Sawyer" ;

ex:ResourceB
  bf:instanceOf ex:ResourceA ;
  bf:instanceTitle "The adventures of Tom Sawyer" ;
  bf:providerDate "1996" .

ResourceA is a bf:Work because bf:creator has a domain of work. (ditto the other properties there)

ResourceB is a bf:Instance because bf:instanceOf has a domain of instance (ditto... )
Example 5: “Unorthodox” BIBFRAME data, with a resource that is an instance of two different classes

This example does not violate the BIBFRAME vocabulary

```
ex:ResourceA
    bf:language <http://id.loc.gov/authorities/names/n79021164> ;
    bf:title "The adventures of Tom Sawyer" ;
    bf:instanceTitle "The adventures of Tom Sawyer" ;
    bf:providerDate "1996" .
```
Linking Bibframe and FRBR

BF

Work

Instance

FRBR

Work

Expression

Manifestation
Linking Bibframe and FRBR

BF

Work

Instance

FRBR

Work

Expression

Manifestation
Example 6: Linking FRBRer and BIBFRAME data

ex:ResourceA  # FRBRer/Work
  frbrer:P2001 ex:ResourceB ;
  frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
  frbrer:P3001 “The adventures of Tom Sawyer” .
ex:ResourceA owl:sameAs ex:ResourceX .

ex:ResourceB  # FRBRer/Expression
  frbrer:P2002 ex:ResourceA ;
  frbrer:P2003 ex:ResourceC ;
ex:ResourceB owl:sameAs ex:ResourceX .

ex:ResourceC  # FRBRer Manifestation
  frbrer:P2004 ex:ResourceB ;
  frbrer:P3020 “The adventures of Tom Sawyer” ;
  frbrer:P3055 “1996” .

ex:ResourceX  # BF/Work
  bf:hasAuthority <http://id.loc.gov/authorities/names/n79021164> ;
  bf:workTitle “The adventures of Tom Sawyer” ;

ex:ResourceY  # BF/Instance
  bf:instanceOf ex:ResourceX ;
  bf:instanceTitle “The adventures of Tom Sawyer” ;
  bf:providerDate “1996” .
Example 6: Linking FRBRer and BIBFRAME data

```
ex:ResourceA  # FRBRer/Work
  frbrer:P2001 ex:ResourceB ;
  frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
  frbrer:P3001 "The adventures of Tom Sawyer" .
ex:ResourceA owl:sameAs ex:ResourceX .

ex:ResourceB  # FRBRer/Expression
  frbrer:P2002 ex:ResourceA ;
  frbrer:P2003 ex:ResourceC ;
ex:ResourceB owl:sameAs ex:ResourceX .

ex:ResourceC  # FRBRer Manifestation
  frbrer:P2004 ex:ResourceB ;
  frbrer:P3020 "The adventures of Tom Sawyer" ;
  frbrer:P3055 "1996" .

ex:ResourceX  # BF/Work
  bf:hasAuthority <http://id.loc.gov/authorities/names/n79021164> ;
  bf:workTitle "The adventures of Tom Sawyer" ;

ex:ResourceY  # BF/Instance
  bf:instanceOf ex:ResourceX ;
  bf:instanceTitle "The adventures of Tom Sawyer" ;
  bf:providerDate "1996" .
```
Example 6: Linking FRBRer and BIBFRAME data

```turtle
ex:ResourceA # FRBRer/Work
  frbrer:P2001 ex:ResourceB ;
  frbrer:P2009 <http://id.loc.gov/authorities/names/n79021164> ;
  frbrer:P3001 "The adventures of Tom Sawyer" .
ex:ResourceA owl:sameAs ex:ResourceX .

ex:ResourceB # FRBRer/Expression
  frbrer:P2002 ex:ResourceA ;
  frbrer:P2003 ex:ResourceC ;
ex:ResourceB owl:sameAs ex:ResourceX .

ex:ResourceC # FRBRer Manifestation
  frbrer:P2004 ex:ResourceB ;
  frbrer:P3020 "The adventures of Tom Sawyer" ;
  frbrer:P3055 "1996" .

ex:ResourceX # BF/Work
  bf:hasAuthority <http://id.loc.gov/authorities/names/n79021164> ;
  bf:workTitle "The adventures of Tom Sawyer" ;

ex:ResourceY # BF/Instance
  bf:instanceOf ex:ResourceX ;
  bf:instanceTitle "The adventures of Tom Sawyer" ;
  bf:providerDate "1996" .
```
Linking Bibframe and FRBR

BF

Work

Instance

FRBR

Work

owl:disjoint

Expression

Manifestation
Example 7: Orthodox data using official RDA properties

```turtle
ex:ResourceA
  rdaw:P10078 ex:ResourceB ; # expressionOfWork
  rdaw:P10065 <http://id.loc.gov/authorities/names/n79021164> ;  # creator
  rdaw:P10088 "The adventures of Tom Sawyer" .  # titleOfTheWork

ex:ResourceB
  rdae:P20231 ex:ResourceA ;  # workExpressed
  rdae:P20059 ex:ResourceC ;  # manifestationOfExpression

ex:ResourceC
  rdam:P30139 ex:ResourceB ;  # expressionManifested
  rdam:P30156 "The adventures of Tom Sawyer" ;  # titleProper
  rdam:P30011 "1996" .  # dateOfPublication
```
Example 8: “Unorthodox” (but formally consistent) RDA data with a resource belonging to multiple WEMI classes

```
ex:ResourceA
  rdaw:P10065 <http://id.loc.gov/authorities/nr>
    # creator
  rdaw:P10088 "The adventures of Tom Sawyer" ;
    # titleOfTheWork
  rdae:P20006 <http://id.loc.gov/vocabulary/iso639>
    # languageOfExpression
  rdam:P30156 "The adventures of Tom Sawyer" ;
    # titleProper
  rdam:P30011 "1996" .  # dateOfPublication
```
Summary

- Open world does not constrain data
- Care must be taken in using RDF & OWL to assure interoperability in the Open World
- Yet, there is an obvious need for quality control at the point of data creation and use
- ANSWER: application profiles
Validation (and application profiles)

- Data quality
  - valid dates/numbers
  - selection of values from a list

- Cardinality (CWA)

- Definition of ‘things’
  - binds sets of properties

- Definition of ‘complete’ description (CWA)
THANK YOU

Slides by Tom Baker and Karen Coyle