JATS Conference
National Library of Medicine, NIH,
Bethesda, MD, USA
16 October 2012

From Markup to Linked Data: Mapping NISO JATS v1.0 to RDF using the SPAR Ontologies

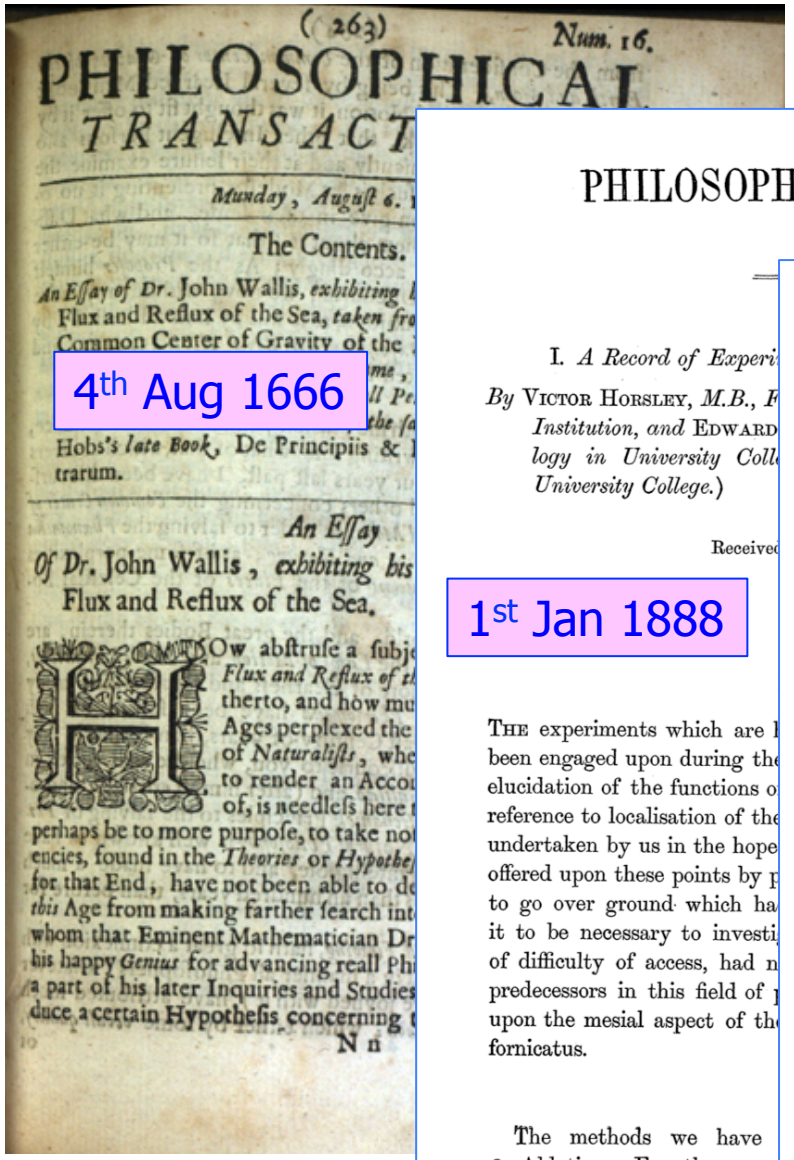
David Shotton

Research Data Management and
Semantic Publishing Research Group
Department of Zoology
University of Oxford, UK

e-mail: david.shotton@zoo.ox.ac.uk

Scholarly communication today

Scholarly articles haven't really changed much in 346 years



(263) Num. 16.

PHILOSOPHICAL TRANSACT

Munday, August 6.

The Contents.

An Essay of Dr. John Wallis, exhibiting

Flux and Reflux of the Sea, taken fro

Common Center of Gravity of the

Hobs's late Book, De Principiis &

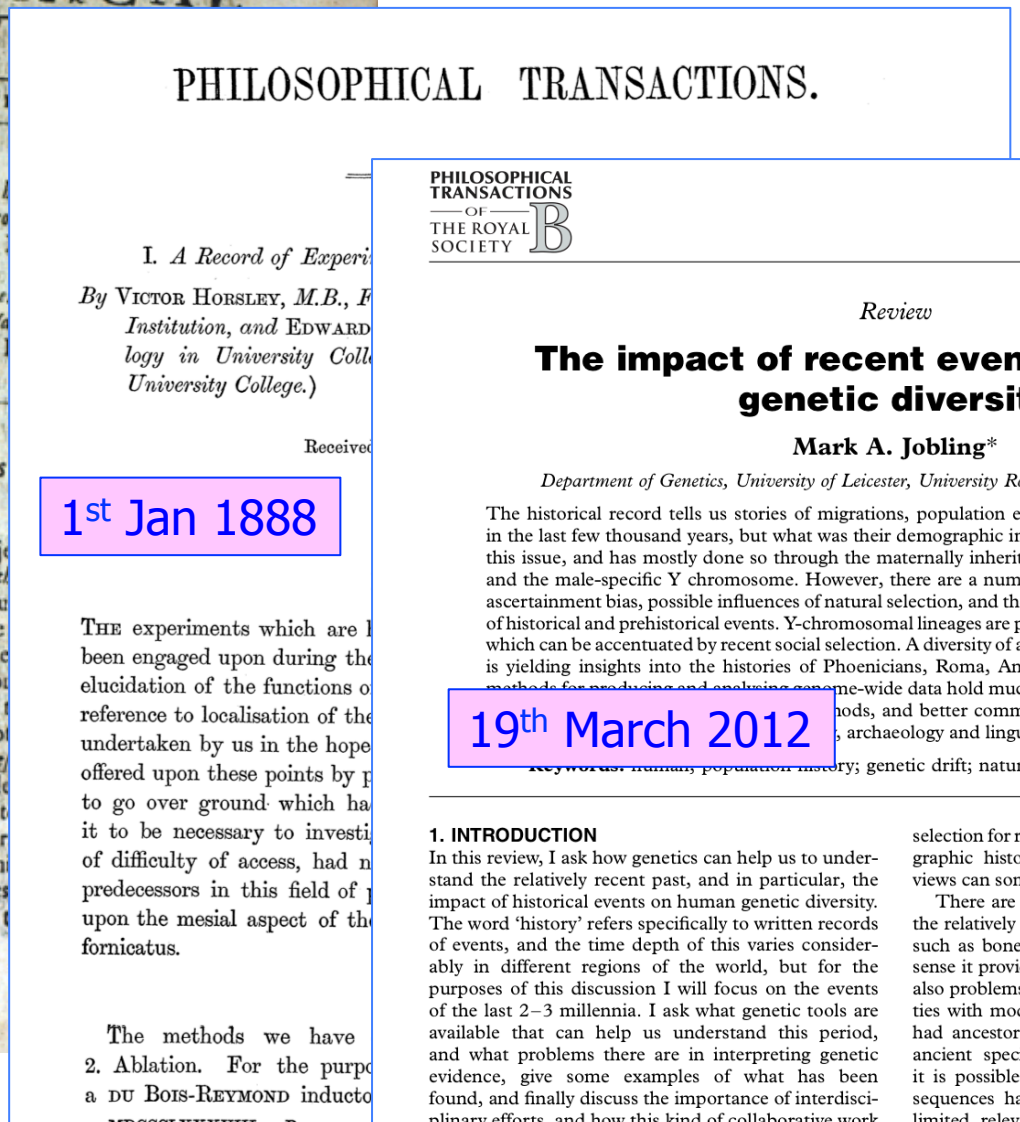
trarum.

An Essay

Of Dr. John Wallis, exhibiting his

Flux and Reflux of the Sea.

4th Aug 1666



PHILOSOPHICAL TRANSACTIONS.

I. A Record of Experi

By Victor Horsley, M.B., F

Institution, and Edward

logy in University Coll

University College.)

Received

THE experiments which are

been engaged upon during the

elucidation of the functions o

reference to localisation of the

undertaken by us in the hope

offered upon these points by

to go over ground which ha

it to be necessary to investi

of difficulty of access, had n

predecessors in this field of

upon the mesial aspect of the

fornicatus.

The methods we have

2. Ablation. For the purpo

a du Bois-Reymond inducto

1st Jan 1888

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2012) **367**, 793–799
doi:10.1098/rstb.2011.0297

Review

The impact of recent events on human genetic diversity

Mark A. Jobling*

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

The historical record tells us stories of migrations, population expansions and colonization events in the last few thousand years, but what was their demographic impact? Genetics can throw light on this issue, and has mostly done so through the maternally inherited mitochondrial DNA (mtDNA) and the male-specific Y chromosome. However, there are a number of problems, including marker ascertainment bias, possible influences of natural selection, and the obscuring layers of the palimpsest of historical and prehistorical events. Y-chromosomal lineages are particularly affected by genetic drift, which can be accentuated by recent social selection. A diversity of approaches to expansions in Europe is yielding insights into the histories of Phoenicians, Roma, Anglo-Saxons and Vikings, and new methods for producing and analysing genome-wide data hold much promise. The field would benefit [illegible] methods, and better communication between geneticists and [illegible], archaeology and linguistics.

Keywords: human; population history; genetic drift; natural selection; genetic diversity

1. INTRODUCTION

In this review, I ask how genetics can help us to understand the relatively recent past, and in particular, the impact of historical events on human genetic diversity. The word 'history' refers specifically to written records of events, and the time depth of this varies considerably in different regions of the world, but for the purposes of this discussion I will focus on the events of the last 2–3 millennia. I ask what genetic tools are available that can help us understand this period, and what problems there are in interpreting genetic evidence, give some examples of what has been found, and finally discuss the importance of interdisciplinary efforts, and how this kind of collaborative work

selection for resistance to infectious disease—and demographic history the nuisance. Reconciling these two views can sometimes be difficult.

There are two ways to read DNA's message. One is the relatively direct means of studying ancient samples, such as bones, themselves. This has advantages—in a sense it provides 'real' information about the past—but also problems: sample sizes are small; there are difficulties with modern DNA contamination; even skeletons had ancestors, so it is still necessary to consider the ancient specimens' own demographic histories; and it is possible that the bearers of these ancient DNA sequences have no living descendants, and therefore limited relevance to today's diversity. Novel

19th March 2012

Scholarly communication – an analogy

- Scholarly communication is, at this mid-point in the digital revolution, in an ill-defined transitional state — a ‘horseless carriage’ state — that lies somewhere between the world of print and paper and the world of the web and computers, with the former still exercising significantly more influence than the latter
- We started here:

- We’re now here (online):
- Great – that’s a significant start

Scholarly communication – an analogy

- . . . but this is really where we need to be!

Publishers' metadata standards are out of date

- Scholarly publishing is in the throes of the digital revolution, as the full potential of on-line publishing is explored
- But publishers still employ a variety of proprietary XML-based informational models and document type definitions (DTDs) to annotate manuscripts
  - . . . as we heard from the American Institute of Physics this morning
  - These now appears anachronistic, since publications and their metadata from different sources are incompatible, requiring hand-crafted mappings to convert from one to another
- In contrast, modern Web information management techniques employ global standards such as RDF (the Resource Description Framework) and OWL 2 (the Web Ontology Language)
- These encode information in ways that permit computers to query metadata and integrate information from multiple resources in an automated manner
- Since the processes of scholarly communication are central to the practice of science, we believe that it is essential that publishers now adopt Web standards to permit inference over the entire corpus of scholarly communication

A bluffer’s guide to RDF and linked data

RDF, ontologies and linked data

- The principles are very simple
  - All entities (classes) and their relationships (properties) are identified by unique URIs, and thus are defined on the Web
  - The URIs reference publicly available and commonly accepted structured vocabularies (ontologies), so that the meaning of terms is unambiguous
  - Each relationship is expressed as a subject – predicate – object 'triple'
  - The syntax is defined by W3C's Resource Description Framework (RDF)
- Examples:

```
:my-article rdf:type fabio:JournalArticle .
:my-article dc:creator "Shotton, David" .
:my-article dc:title "CiTO, the Citation Typing Ontology" .
```

- Such statements can be combined into interconnected information networks (RDF graphs) – forming 'linked data'
  - thereby creating a web of knowledge, the Semantic Web,

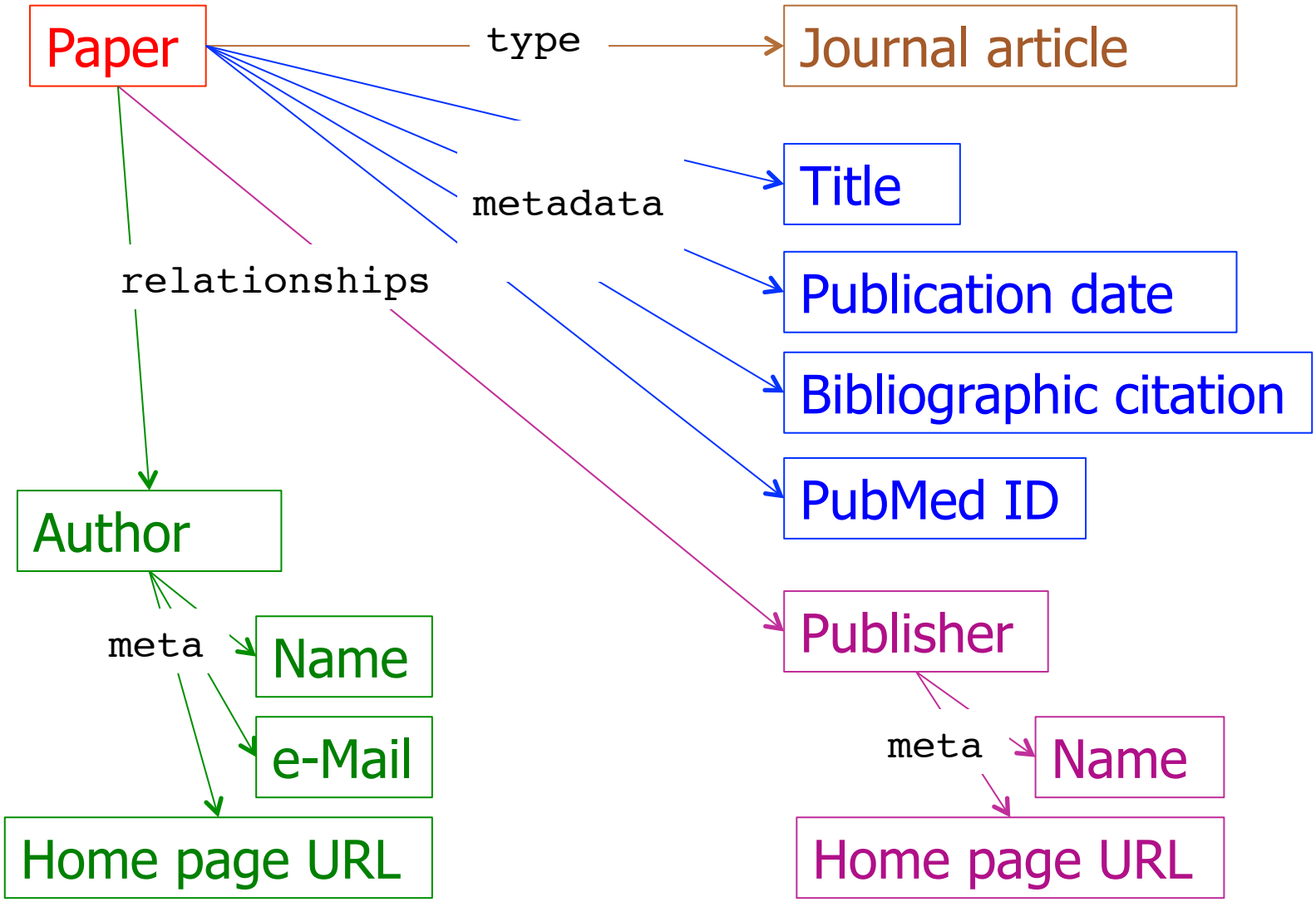
    in which the truth content of each original statement is maintained

A traditional bibliographic record card for my CiTO paper

```
TI   - CiTO, the Citation Typing Ontology.
AU   - Shotton D
DP   - 2010
PHST- 2010/06/22 [aheadofprint]
PT   - Journal Article
TA   - J Biomed Semantics
JT   - Journal of biomedical semantics
SO   - J Biomed Semantics. 2010 Jun 22;1 Suppl 1:S6.
VI   - 1 Suppl 1
PG   - S6
PMID- 20626926
AID - 10.1186/2041-1480-1-S1-S6 [doi]
FAU - Shotton, David
```

- Simple PubMed tag–value pairs
- No relationships
- No hierarchical structure

A generic RDF graph to encode this information



The RDF graph for this CiTO record (written in Turtle format)

```
<http://dx.doi.org/10.1186/2041-1480-1-S1-S6> # URI of the
          CiTO paper in Journal of Biomedical Semantics

  rdf:type fabio:JournalArticle ;

  dc:title "CiTO, the Citation Typing Ontology" ;
  fabio:hasPublicationDate "2010-06-22" ;
  dcterms:bibliographicCitation "Shotton D (2010). CiTO, the
    Citation Typing Ontology. J. Biomed. Semant. 1,S1: S6." ;
  fabio:hasPubMedId "20626926" ;

  dcterms:publisher
    [ rdf:type foaf:Organization ; foaf:name "BioMed Central";
    foaf:homepage <http://www.biomedcentral.com/> ] ;

  dcterms:creator
    [ rdf:type foaf:Person ; foaf:name "David Shotton " ;
    foaf:mbox <mailto:david.shotton@zoo.ox.ac.uk> ;
    foaf:workplaceHomePage
 <http://www.zoo.ox.ac.uk/staff/academics/shotton_dm.htm> ] .
```

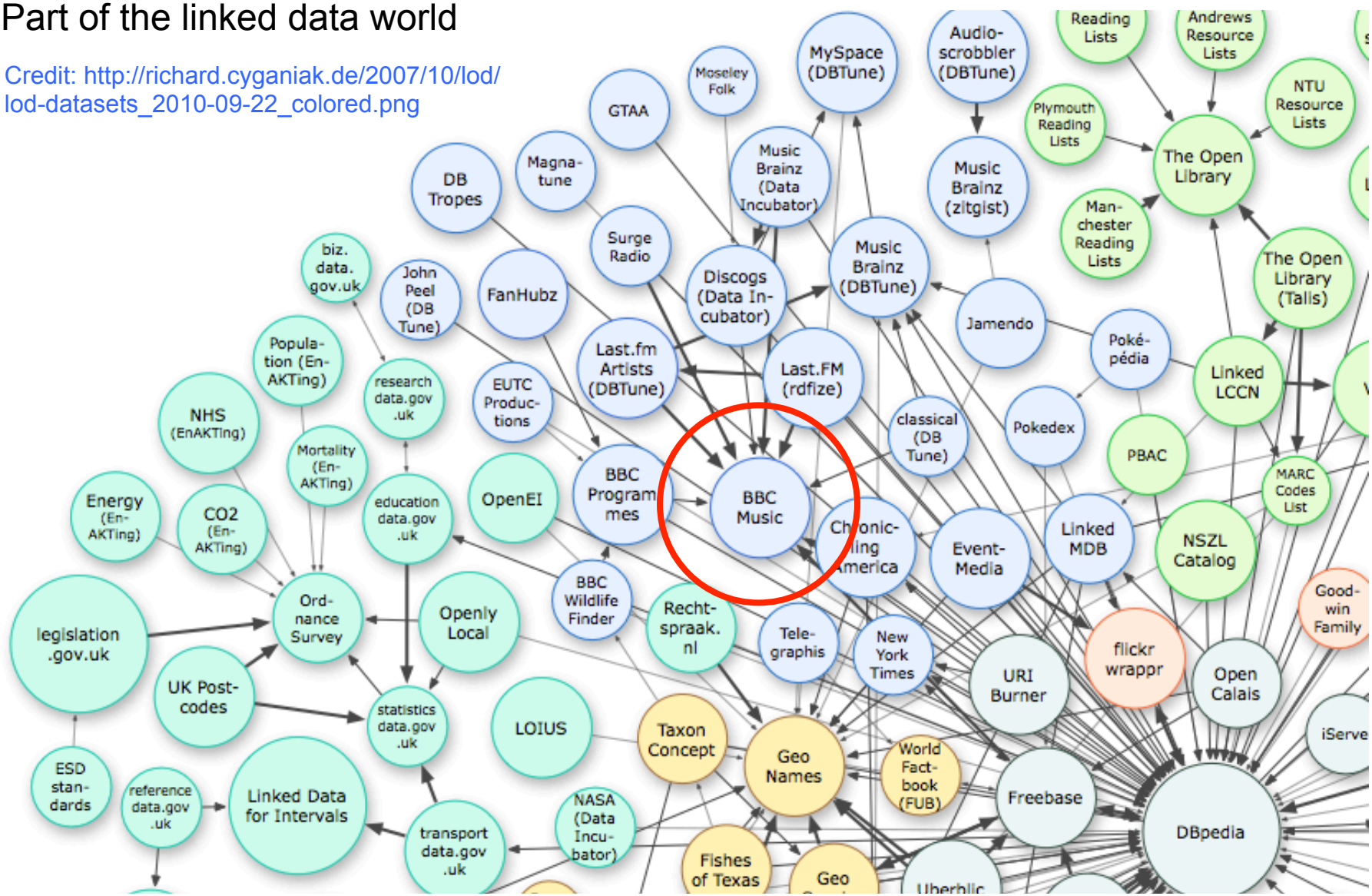
A major user of RDF linked data – the BBC

- The BBC World Cup 2010 website used a high-performance dynamic semantic publishing framework underpinned by RDF and appropriate ontologies
  - This provides far deeper and richer use of content than can be achieved through traditional CMS-driven publishing solutions
- The BBC Music website is built on lot of Linked Data and RDF goodness. BBC Music provides a truly RESTful API for querying its data
  - For example, each artist in BBC Music has an RDF representation
- And the entire BBC Natural History web site is powered by RDF, with its own Wildlife Ontology
- The BBC got to this place by hiring bright people who has relevant semantic web skills

Such services become global resources to which others link

Part of the linked data world

Credit: http://richard.cyganiak.de/2007/10/lod/lod-datasets_2010-09-22_colored.png



The SPAR
(Semantic Publishing and Referencing)
Ontologies

SPAR

Semantic Publishing And Referencing

The SPAR Ontologies

- These SPAR ontologies are described at http://purl.org/spar/ and in my blog Open Citations and Semantic Publishing at http://opencitations.wordpress.com

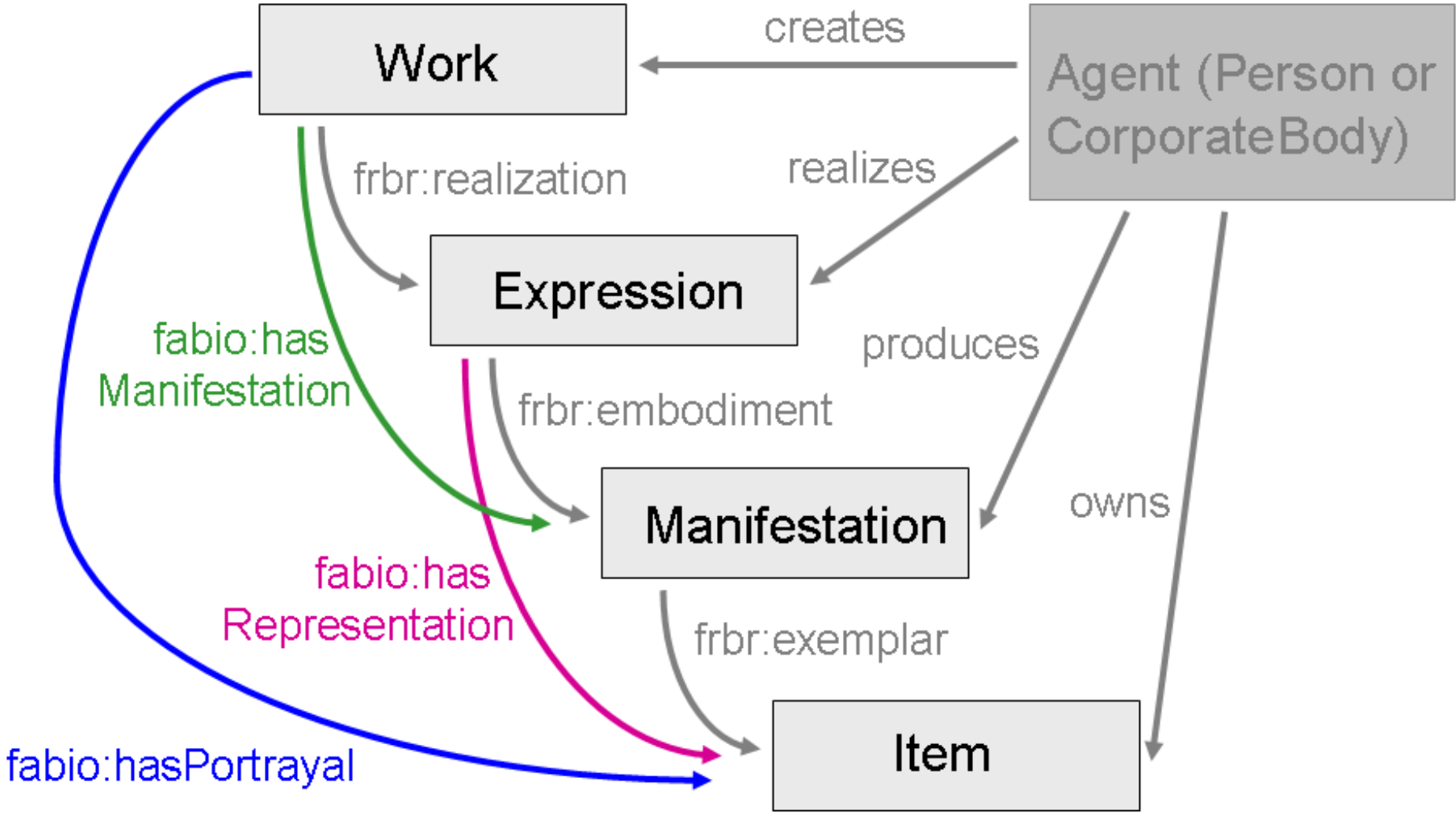
CiTO, the Citation Typing Ontology http://purl.org/spar/cito
enable characterization of the nature or type of citations, both factually and rhetorically

FaBiO, the FRBR-aligned Bibliographic Ontology http://purl.org/spar/fabio
is an ontology for describing bibliographic entities (books, articles, etc.)

BiRO, the Bibliographic Reference Ontology http://purl.org/spar/biro
is an ontology to define bibliographic records and references, and their compilation into bibliographic collections and reference lists, respectively

FaBiO and BiRO classes are structured according to the FRBR model of *Works*, *Expressions*, *Manifestations* and *Items*

New Work-Expresson-Manifestation-Item relationships in FaBiO



Reciprocals: frbr:realizationOf frbr:embodimentOf frbr:exemplarOf
fabio:isManifestationOf fabio:isRepresentationOf fabio:isPortrayalOf

The SPAR Ontologies, continued

C4O, the Citation Counting and Context Characterization Ontology http://purl.org/spar/c4o allows the characterization of bibliographic citations in terms of their number (both locally and globally), and their textual context

DoCO, the Document Components Ontology http://purl.org/spar/doco provides a structured vocabulary of document components, both structural (e.g. heading, paragraph) and rhetorical (e.g. Abstract, Introduction)

PRO, the Publishing Roles Ontology http://purl.org/spar/pro is an ontology for the roles of agents (e.g., author, editor, publisher, librarian) in the publication process, and the times during which those roles are held

PSO, the Publishing Status Ontology http://purl.org/spar/pso is an ontology for the temporal status of a document (e.g. draft, under review, published, Version of Record) during the publication process

PWO, the Publishing Workflow Ontology http://purl.org/spar/pwo describing the steps in the workflow associated with the publication of a document or other publication entity

Bibliographic information encoded in RDF using SPAR

```
<http://dx.doi.org/10.1371/journal.pntd.0000228>      # The citing paper, Reis et al., 2008
    a fabio:JournalArticle ;      # expression
    frbr:realizationOf [ a fabio:ResearchPaper ] ;      # work
    pso:holds [ a pso:StatusInTime ; pso:withStatus pso:peer-reviewed ] ;
    cito:cites <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;      # Reference [6]; Ko et al., 1999
    cito:obtainsBackgroundFrom  <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
    cito:usesDataFrom <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
    cito:confirms <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
    cito:extends <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
    cito:sharesAuthorsWith <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
    frbr:part [ a biro:BibliographicReference ;
        biro:references <http://dx.doi.org/10.1016/S0140-6736(99)80012-9> ;
        c4o:hasInTextCitationFrequency "10"^^xsd:nonNegativeInteger ] .

<http://dx.doi.org/10.1016/S0140-6736(99)80012-9>      # Reference [6], the cited paper, Ko et al., 1999
    dcterms:bibliographicCitation "Ko AI, Reis MG, Ribeiro Dourado CM, Johnson WD Jr, Riley LW (1999). Urban
      epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. Lancet 354: 820-825." ;
    prism:publicationDate "1999-09-04"^^xsd:date ;
    cito:isCitedBy <http://dx.doi.org/10.1371/journal.pntd.0000228> ;      # The citing paper, Reis et al., 2008
    c4o:hasGlobalCitationFrequency [ a c4o:GlobalCitationCount ;
        c4o:hasGlobalCountValue "309"^^xsd:integer ; c4o:hasGlobalCountDate "2011-09-07"^^xsd:date  ;
        c4o:hasGlobalCountSource <http://scholar.google.com>  ] .
```

Mapping JATS, the Journal Article Tag Suite, to RDF

Metadata for describing bibliographic entities – next steps

- The National Library of Medicine DTD has become a *de facto* standard for many publishers and PubMed Central to create XML mark-up for journal articles
- The most recent version of the NLM DTD is the Journal Article Tag Suite (JATS), published on 22 August 2012 as ANSI/NISO Z39.96-2012, JATS: Journal Article Tag Suite (version 1.0)
- In July, guided by Deborah Lapeyre of Mulberry who created JATS as to its meaning, Silvio Peroni and I mapped to RDF the key metadata elements of the ANSI/NISO JATS Journal Publishing Tag Library
- For this, we used the SPAR Ontologies and other appropriate ontologies (DataCite, Dublin Core, FOAF, FRBR, PRISM, SKOS, vCard)
- The mapping document is available from http://purl.org/spar/JATS2RDF/

What we mapped

- The JATS Journal Publishing Tag Library Version 1.0 specification is large, containing 246 elements and 134 attributes
- We chose to map the JATS **metadata entities** that describe an article
  - e.g. <journal-meta> for metadata about the journal in which the article was published
- We left aside (for a possible later mapping exercise using DoCO, the Document Components Ontology) those entities describing the textual and graphical structure and content of the article (e.g. <title>, <body>, <fig>, <table>)
- The principle metadata elements that we chose to map are
  - <article> <article-meta> <journal-meta> <contrib> <ref-list>

    and their key component elements and attributes
- In all, 242 separate XML to RDF mapping statements have been made
- Translated titles, name alternatives and alternate languages accommodated
- Using the Collections Ontology, we can also encode ordered lists (e.g. authors)

JATS2RDF: The first four items in the <ref-list> mapping table

Element/attribute name	XML example	RDF translation

ref-list	`<ref-list> ... </ref-list>`	`:textual-entity frbr:part :ref-list .`
		
		`:ref-list a biro:ReferenceList .`

ref	`<ref-list>`
	`  <ref id="XXX"> ... </ref>`
	`  <ref id="YYY"> ... </ref>`
	`  ...`
	`</ref-list>`
		`:ref-list co:item :iref-XXX .`
		
		`:iref-XXX a co:ListItem`
		`  co:itemContent :ref-XXX ;`
		`  co:nextItem :iref-YYY`
		`  co:index "{$count_references}" .`
		
		`:ref-XXX a biro:BibliographicReference .`

element-citation	`<ref id="XXX">`
	`  <element-citation>`
	`  ...`
	`  </element-citation>`
	`</ref>`
		`:ref-XXX biro:references :textual-entity-XXX .`
		
		`:textual-entity cito:cites :textual-entity-XXX .`

chapter-title	`<element-citation>`
	`  ...`
	`  <chapter-title>XXX</chapter-title>`
	`  ...`
	`</element-citation>`
		`:textual-entity-XXX a fabio:BookChapter ;`
		`  frbr:partOf :textual-entity-XXX-collection ;`
		`  dcterms:title "XXX" .`
		
		`:textual-entity-XXX-collection a fabio:Book .`

Challenges encountered when mapping JATS to RDF

Differing philosophical viewpoints of XML and RDF

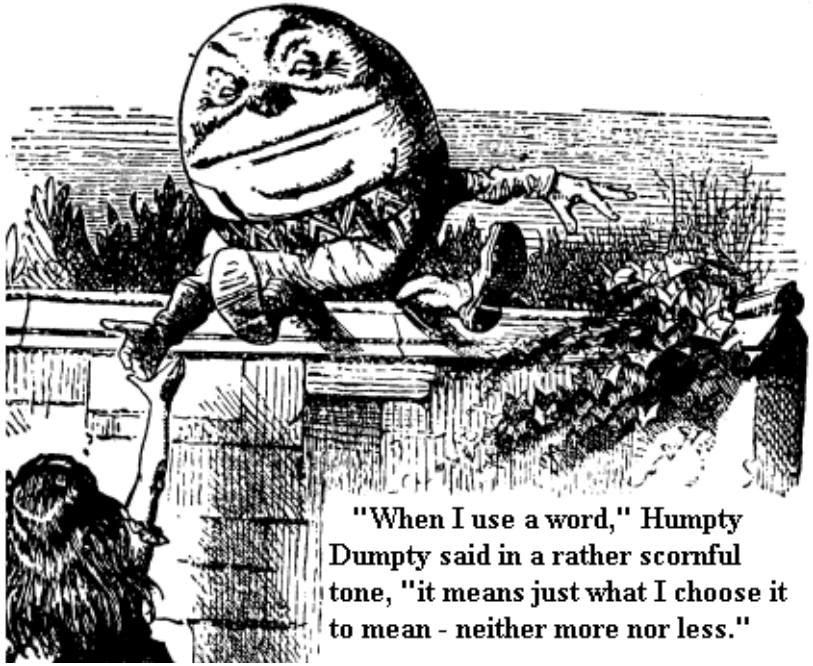
- Closed and open worlds
- Loose and precise semantics
- Hierarchies versus triples
- ‘Flat’ versus FRBR categories
- Mapping publication formats
- Mapping roles

Closed and open worlds

- The first thing to understand is that Semantic Web technologies are underpinned by an an 'open world' philosophical viewpoint
- This is commonly contrasted with the 'closed world' of database technologies
  - If an item of information is not present in a database, its converse is assumed to be true
  - For example, if a journal article is not recorded as being open access, it is assumed not to be
- But in the open world view of RDF, if an article is not describes as being open access, one has to keep an open mind — it might be, or it might not
- There is thus a difference in the assumed meanings of unstated assertions
- The JATS documentation for the attribute @publication-format suggests the value "online-only”
- However, in the open world of RDF, we would not wish to state that a publication is "online-only", since we cannot read the future
  - someone might come out with a print edition later on, and we wish our RDF encoding to be as true in the future as it is in the present

Precise semantics for mark-up terms

- A second interesting contrast arises when comparing RDF and XML descriptions is between the semantic meanings of markup terms
- A cornerstone of the Semantic Web is the use of open published ontologies to give terms precise and universally available definitions, so that RDF statements are unambiguous in their meaning
- This is *not* the case in the world of XML, where markup terms can take on different meanings, depending upon who is using them
- This is reminiscent of Humpty Dumpty's statement in Alice's Adventures in Wonderland:
  - 'When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.'



Semantics in JATS

- For JATS, this is by design. JATS is a descriptive, not a prescriptive model, that endeavours to capture and document the actual practice of current publishing
- The JATS standard is *deliberately* vague about the meaning of terms, because there is no intention to tell any publisher what they should call their content
- Furthermore, suggested values for JATS entities are just that — suggested
- E.g. JATS <article>:
  - "This element can be used to describe not only typical journal articles (research articles) but also much of the non-article content within a journal, such as book and product reviews, editorials, commentaries, and news summaries."
- Thus JATS <article> may be used to describe
  - a research article,
  - *or* another kind of journal content, e.g. an obituary, a quiz, an interview
  - *or* even a non-published document. e.g. a preprint
- This goes beyond what the average person means by "journal article"

Accommodating the loose semantics of JATS in RDF

- We have seen that what JATS means by <article> is most frequently what is defined as a fabio:JournalArticle, but it can also mean
  - fabio:JournalEditorial, fabio:JournalNewsItem, fabio:BookReview, etc.
- All these *could* be mapped in RDF as follows:

```
:periodical-entity a fabio:PeriodicalItem ;
    frbr:partOf [ a fabio:JournalIssue ] .
```

- However, since JATS <article> can also mean fabio:Preprint, even this use of fabio:PeriodicalItem is too specific
- Thus we have to map the various entities described by JATS <article> to

```
:textual-entity
```

  a resource name that is broad enough to include all relevant possibilities

- This achieving semantic accuracy, if not detailed specificity!

Hierarchies versus triples

- XML uses nested statements to define hierarchical relationships, e.g.
  - <article-meta> ...
    <article-categories>
    <subj-group>
    <subject>XXX</subject>
    <subj-group>
    <subject>YYY</subject>
    </subj-group>
    </article-categories>
    … </article-meta>
- However, RDF triples are themselves ‘flat’, not hierarchical, and thus to indicate this subject hierarchy we use SKOS (http://www.w3.org/2004/02/skos/) to say that YYY is a more specific term than XXX:

```
:textual-entity fabio:hasSubjectTerm  :term1 , :term2 .

:term1 a fabio:SubjectTerm ; rdfs:label “XXX” ;
   skos:narrower :term2 .

:term2 a fabio:SubjectTerm ; rdfs:label “YYY” .
```

Mapping JATS to the FRBR categories

- In creating FaBiO, we aligned our ontology to the FRBR model, deciding that
  - some things (e.g. an Opinion, a Research Paper, a Novel) were Works
  - while others (e.g. an Editorial, a Journal Article) were Expressions, and
  - yet others (e.g. a Reprint) were clearly Manifestations
- We adopted this same methodology in creating the mapping from JATS to RDF, using the following general resource names as appropriate:
  - :conceptual-work (the Work from which the JATS article derives)
  - :textual-entity (the Expression bearing the JATS XML markup)
  - :digital-embodiment (a digital Manifestation of the article, e.g. a PDF)
  - :digital-item (an Item, a single ownable copy of the JATS article)
- The following FRBR relationships exist between these entities:

```
:textual-entity a fabio:Expression ;
frbr:realizationOf :conceptual-work ;
frbr:embodiment :digital-embodiment ;
fabio:hasRepresentation :digital-item .
```

Resulting mappings – revisions and retractions

- The FRBR Work layer is the only one that may change during time, from the first draft to the final published version or subsequently corrected version
- The individual Expression at each stage is a static document that does not change, while every revision of the Work results in a new Expression
- Thus the @date-type attributes “rev-request”, “rev-rec”, “ecorrected” and “pcorrected” are mapped to :conceptual-work, rather than :textual-entity:

```
:conceptual-work fabio:hasRevisionRequestDate
        “YYYY-MM-DD”^^xsd:date .
```

- Conversely, retractions apply to published Expressions (here :textual-entity) or their Manifestations (:digital-manifestation). You cannot retract a Work.

```
:textual-entity
        fabio:hasRetractionDate “YYYY-MM-DD”^^xsd:date .
```

Media types - mapping the attribute @publication-format

- The JATS documentation for this optional attribute that defines the format of a publication suggests the following values: "print", "electronic", "video", "audio", "ebook", "online-only”
  - "Online" and "web" are additional possibilities
- PROBLEM!! This grouping of terms betrays ‘loose’ thinking, since it conflates the following independent categories:
  - the nature of the information, e.g. text, image, sound
  - the nature of the storage medium, e.g. paper, digital tape, DVD, Web
  - the analogue or digital file format, e.g. PDF, XML, VHS and JPEG
- Each of these categories is encoded independently in FaBiO
- The exact nature of our mapping thus depends on what is precisely intended, as explained in detail in our mapping paper
- Reminiscent of the American Institute of Physics conclusion this morning –
  - “How to treat problems? Change JATS, change XSLT or manual fix”

Mapping roles

- Someone can be the editor of one paper, and an author of another
- Because we wish to make statements in RDF that are independently and universally true, we cannot just say

```
:this-person a pro:Editor
```

  since that relationship holds only in a particular context

- To map publishing roles such as editor or contributor to RDF, we thus use PRO, the Publishing Roles Ontology, that permits the context of the role to be specified (and also, where necessary, the temporal extent of that role), e.g.

```
:this-person pro:holdsRoleInTime [ pro:withRole
  pro:author ; pro:relatesToDocument :conceptual-work ] .
```

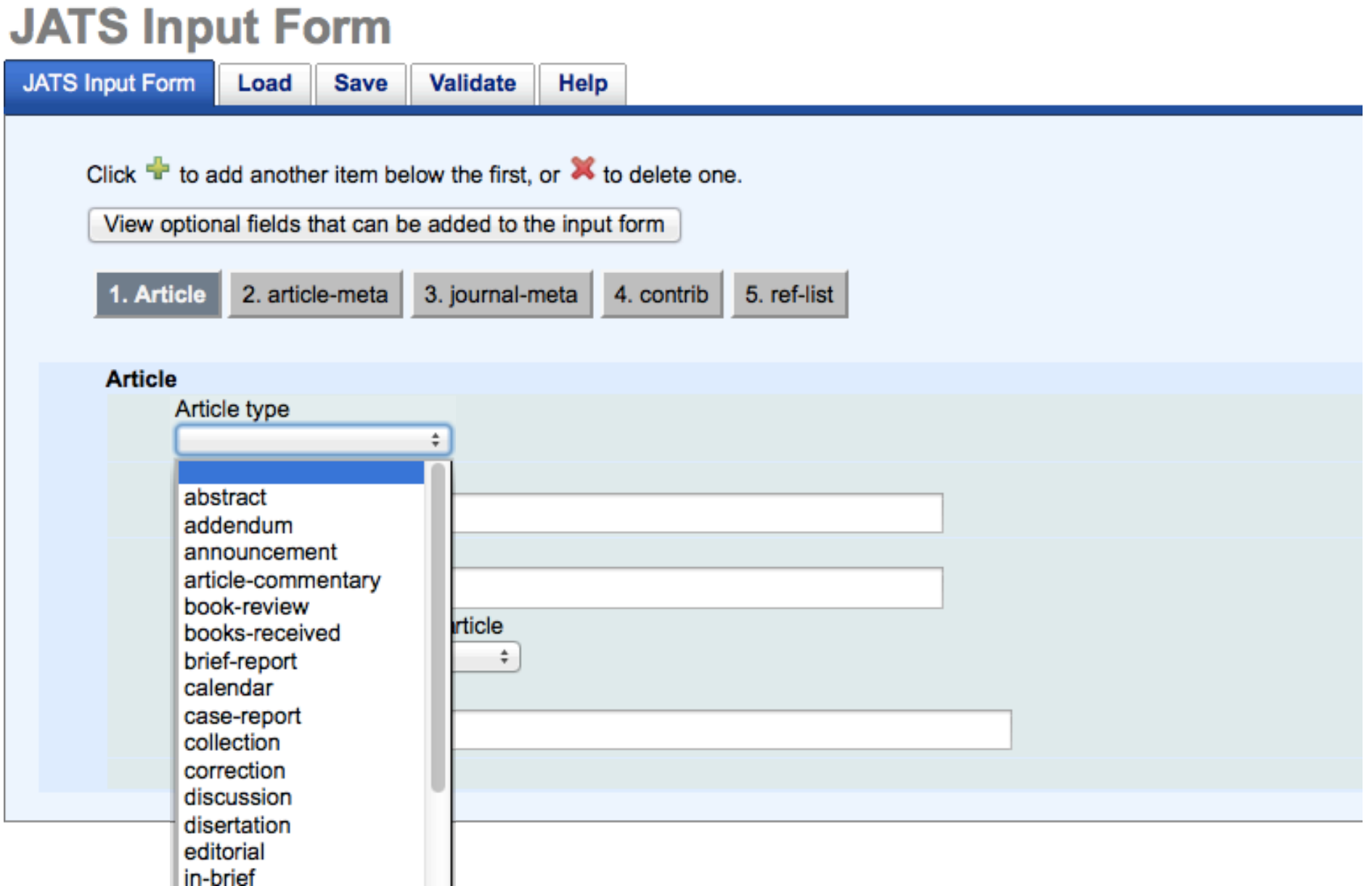
- For other, non-publishing, roles (e.g. photographer), we use the SCORO, the Scholarly Contributions and Roles Ontology, e.g.

```
:this-person pro:holdsRoleInTime [
    pro:withRole scoro:photographer ;
    pro:relatesToDocument :conceptual-work ] .
```

Automating the conversion from JATS to RDF

- Last month, while I was away from work for family reasons, Silvio Peroni created an Extensible Stylesheet Language Transformation (XSLT) transform that permits the metadata elements of a document marked up in JATS XML to be transformed automatically into RDF
- This XSLT is available at http://purl.org/spar/jats2rdf/xslt
- Our JATS2RDF mapping and this XSLT transform together now permit
  - the JATS metadata elements and their attributes
  - from documents marked up in XML using the NISO-JATS Journal Publishing Tag Library v1.0
  - to be converted automatically to RDF
  - enabling this information to be published to the Semantic Web as linked open data in a manner that is unambiguous and universally understood

JATS metadata input form - <article>



JATS metadata input form - <article-meta>

JATS Input Form

JATS Input Form | Load | Save | Validate | Help

Click + to add another item below the first, or ✖ to delete one.

Reset form

View optional fields that can be added to the input form

1. Article | 2. article-meta | 3. journal-meta | 4. contrib | 5. ref-list

article-meta

article-id

pub-id-type

article-categories

1. **subj-group**

subject

title-group

article-title

subtitle

trans-title-group

language

trans-title

alt-title

pub-date

date-type

publication-format

calendar

pub-type

Day of the month

Month

Year (yyyy: four-digit numeral)

Season

issue

seq

issue-id

issue-title

issue-sponsor

issue-part

JATS metadata input form - <journal-meta>

JATS Input Form

JATS Input Form | Load | Save | Validate | Help

Click + to add another item below the first, or ✖ to delete one.

Reset form

View optional fields that can be added to the input form

1. Article | 2. article-meta | 3. journal-meta | 4. contrib | 5. ref-list

journal-meta

journal-id

journal-id-type

journal-title-group

journal-title

journal-subtitle

trans-title-group

language

trans-title

trans-subtitle

abbrev-journal-title

abbrev-type

issn

issnl

isbn

publisher

publisher-name

publisher-loc

JATS metadata input form - <contrib>

JATS Input Form

JATS Input Form | Load | Save | Validate | Help

Click + to add another item below the first, or ✕ to delete one.

Reset form

View optional fields that can be added to the input form

1. Article | 2. article-meta | 3. journal-meta | 4. contrib | 5. ref-list

contrib

contrib-type

corresp

deceased

equal-contrib

anonymous

contrib-id

contrib-id-type

1. **collab**

collab-type

language

collab-alternatives

1. **collab**

collab-type

name

surname

initial(s)

given name(s)

initial(s)

prefix

suffix

name-alternatives

1. **name**

surname

initial(s)

language

Whole name

degrees

JATS metadata input form - <ref-list>

JATS Input Form

JATS Input Form | Load | Save | Validate | Help

Click + to add another item below the first, or ✖ to delete one.

Reset form

View optional fields that can be added to the input form

1. Article | 2. article-meta | 3. journal-meta | 4. contrib | 5. ref-list

ref-list

1. **ref** +

id

Please click on a button to display the corresponding input fields:

element-citation | mixed-citation

element-citation

publication-type

chapter-title

part-title

source

edition

gov

person-group

person-group-type

etal

patent

std

annotation

date-in-citation

content-type

std-organization

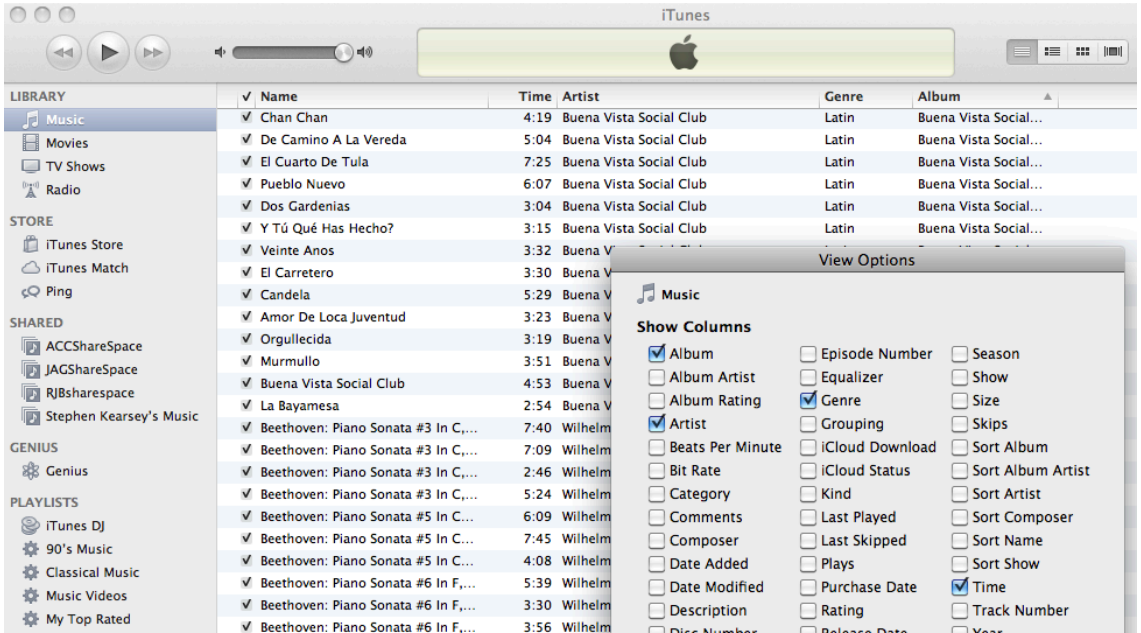
Recommendation 1

Think Web, not print

In the Web era . . .

- Electronic resources have already become the dominant norm, far more important than physical ones

- Cataloguing paradigms based on the index card will become replaced by faceted browse and semantic search over rich linked metadata



Recommendation 2

Adopt Semantic Web technologies

RDF markup is becoming increasingly used

- Last October, the US Library of Congress agreed to replace MARC, adopting
  - a new bibliographic framework focused on Web environment
  - linked Data principles and mechanisms
  - use of the Resource Description Framework (RDF) as the basic data model
- "RDF will enable the integration of library data... on the Web for more expansive user access to information"

It is difficult to change a publisher's direction – but not impossible!

- By creating the SPAR (Semantic Publishing and Referencing) Ontologies, and by providing the ability to convert JATS metadata automatically to RDF, we have made it easier for publishers, and for PubMed Central, to publish their bibliographic metadata about journal articles as linked open data
- We hope that this ability to express in RDF the JATS Journal Publishing Tag Library metadata descriptions will promote the use of JATS to a wider community

Acknowledgements

- Semantic exemplar: Katie Portwin, Alistair Miles and Graham Klyne
- SPAR ontologies, DataCite2RDF and JATS2RDF mappings: Silvio Peroni
- Web forms for JATS metadata entry: Tanya Gray

- Funding: Joint Information Systems Committee

end