Wikifying Research

Michael Cysouw, MPI-EVA (cysouw@eva.mpg.de)
Jeff Good, University at Buffalo (jcgood@buffalo.edu)

Taking back the web...

- The world wide web was originally developed in an academic context

- However, it has strayed from its origins

  ‣ Commercial uses

  ‣ Social uses

- How do we adapt it to turn it into a tool for cutting-edge research?

The social web

- Recent "Web 2.0" developments have begun to illustrate the web's potential as a place for collaboration and sharing

  ‣ Wikipedia

  ‣ Flickr

  ‣ YouTube

  ‣ ...

The scholar's web

- But none of these are well designed to deal with issues crucial to academic research
    - ▸ Citation
    - ▸ Provenance
    - ▸ "Publication"
    - ▸ Contested interpretations

New knowledge

- Sites like Wikipedia have shown the Web's power for collecting and organizing textbook knowledge

- But support for pushing the boundaries of knowledge is lacking

- How do you bring together the required knowledge to write an article on a previously unknown topic?

Multiple Interpretations

- A unifying feature of much of the work to be discussed here is that it involves data that is subject to multiple interpretations

- Often, it is impossible to replicate a particular datum (historically bound, expensive data)

- In that case researchers interpret the datum again and again rather than trying to re-collect it

Wikifying research

- We use the term "wikifying" to be more provocative then precise

- Wikis give us a model of collaborative authorship which can

  ‣ Can develop high-quality content

  ‣ Track the process through which this content is developed

- How can we adapt this model in the creation of new knowledge?

Some desiderata

- Data and interpretations should be citable at fine levels of granularity

- The provenance of a data point or an interpretation should readily recoverable

- All collection of data and stages of interpretation should be recorded

- Capability of marking high-quality data and interpretations in a way equivalent to (peer-reviewed) publishing

Support for collaboration

- Various social models of collaboration should be supported

  ‣ Laboratory style: Head researcher, different levels of assistants

  ‣ Humanities style: Small number of collaborators, roughly equal levels

  ‣ Wiki style: many participants, no central organisation

  ‣ …

What's needed

- In order to wikify research, we need at least the following things

  ▸ A conceptual data model

  ▸ An implementation of the model

  ▸ Content

  ▸ Tools to exploit content using the model

The data model

- Data as a labeled graph

- Nodes in the graph are unique

- This model is borrowed from the W3C's *Resource Description Framework* (RDF)

- The ultimate *relational* database

- Graph-based models give flexibility for encoding cutting-edge knowledge
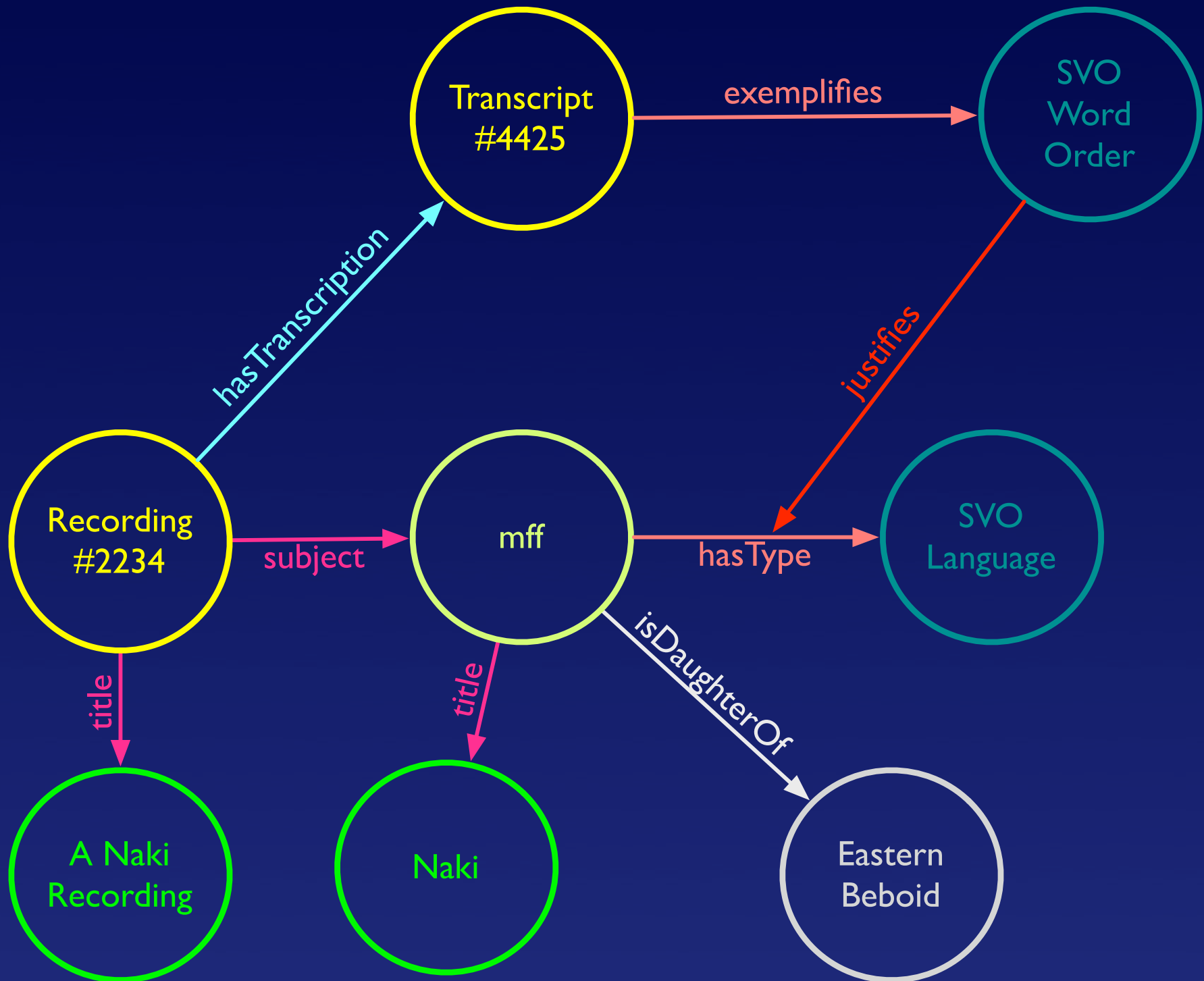
Implementation

- RDF has not yet been extensively adopted, but its use is growing

- It is a key component of the Semantic Web—also known as "Web 3.0"

- Tool support for working with RDF is increasing

Content

- But the data model on its own will not get us very far

- Ultimately, what we want is research-quality content encoded in the model

- This requires

  ‣ Systems for managing unique identifiers

  ‣ Systems for managing relations

Domain-specific content

- Some issues involving content will be domain-specific

- Notions like "SVO Language" are of interest primarily to linguists, for example

- Managing the encoding of such notions should be the job of the relevant academic community

General research

- But other issues are the domain of the whole academic community

  ‣ Managing citation

  ‣ Provenance

  ‣ Publication

  ‣ (Peer) Review

Moving forward

- Our goals here

  ‣ Share ideas regarding content management using new technologies

  ‣ Gain an understanding of the domain-specific and general challenges to wikifying research

  ‣ Establish worthwhile areas of collaboration across disciplines

This workshop

- This workshop, being organized by linguists, has a linguistics bias

- However, we have deliberately tried to also bring together historians, biologists, and others here

- We believe that building the necessary infrastructure for linguistics is best served by first understanding what is needed for a wide range of disciplines

Workshop program

- **Jürgen Renn, Urs Schoepflin, Simone Rieger**
  MPI for the History of Science, Berlin
  "Scholarly workbench for the history of science"

- **Sigrid Weigel, Falko Schmieder**
  ZFL Berlin
  "The gap between semantics and data. What kind of facilities do Humanities need?"

Workshop program

- **Lars Vogt, Peter Grobe**
  FU Berlin
  "Morph.D.Base and MorphOntology: What can a RDF ontology provide for a Morphological Description Database?"

- **Janet Kelso, Robert Hoehndorf**
  MPI-EVA Leipzig
  "Collaborative knowledge management for biological data"

Workshop program

- **Balthasar Bickel**
  University of Leipzig
  "Workflow management in large teams: beyond the hermit linguist"

- **Martin Haspelmath**
  MPI-EVA Leipzig
  "Long-distance collaboration in the creation of cross-linguistic databases"

Workshop program

- **Brian Fuchs**
  (*Imperial College London*)
  "Service sharing: a service composition toolkit for collaborative online research"

- **Laurent Romary** *MPDL, Berlin*
  Lee Gillam *University of Surrey*
  "Linking open spaces and standards: the case of language codes and language description"

Workshop program

- **Sebastian Nordhoff**
  University of Amsterdam
  "Tracing building blocks of typological hypotheses through the grammar authoring system GALOES

- **Arienne M. Dwyer**
  University of Kansas
  "Diverse Research Teams as Proto-Wikis: The challenges of multiple ontologies and metadata inconsistencies"

Workshop program

- **Dafydd Gibbon**
  University of Bielefeld
  "Fast food for thought: on truth, negotiation and prosody"