Ontology Applications for NSF Records Management

Frank Olken
National Science Foundation CISE/IIS/III
folken@nsf.gov

Presentation to
CEW/GSA/NSF Workshop on
Towards Stable Meaning and Records Preservation in Information-Sharing: Building the Way Forward Together
at
National Science Foundation
Arlington, VA
July 18, 2007

Disclaimer

These remarks are NOT NSF management views or policies.

They are the personal views of Frank Olken.

This talk is about potential future applications of ontologies for NSF Record Management.

None of this exists today.

Overview

- NSF Activities
- NSF Record Management
- Ontology Applications
  - searching
  - clustering, routing
  - reporting
  - workflow management
- Ontologies Needed

NSF Activities

- Receive proposals
- Review proposals
- Either decline or fund proposals
- Transfer funds
- Track award progress
- Track funds

NSF Records Management

- Every decision must be justified and documented
- All money must be tracked, documented
- Research progress must be documented
- Declines have approx. 10-20 records each
- Awards will have several dozen records each
- Active lifetime of award is approx. 4-8 years
- NSF currently transitioning from paper to fully electronic records (currently both).

NSF Archival Records Management

- Lifetime of (online) Archival Records
  - For awards: lifetime of agency ?
  - For declines: ???
- Uses of (online) Archival Records
  - Search for similar proposals
  - To find / evaluate reviewers
  - To evaluate novelty of proposals
  - To evaluate principal investigators
  - To answer management queries

NSF Record Types

- Research proposals
- Reviews
- Awards processing
- Monetary (budgets, expenditures, ...)
- Progress reports

NSF Records

- Policy Manual
- Solicitations
- Proposals
- Reviews, panel summaries, review analyses
- IRB Approvals for human subjects, animal subjects
- Budgets (proposal, NSF)
- Award letters
- Award budgets
- Award abstracts
- Progress reports

NSF Ontology Applications

- Searching
- Clustering / Routing
- Reporting
- Workflow Management
- Entity resolution

Searching

- Ontologies can enhance searching
  - Synonyms
  - Taxonomies
  - via query expansion
  - via taxonomy based searching

Clustering / Routing

- Want to cluster similar proposals
- For routing proposals to:
  - divisions (for NSF wide solicitations)
  - clusters within divisions (division solicitations)
  - program officers within clusters
  - panels
  - reviewers
- Note: routing must avoid conflict-of-interest issues

Reporting

- Some built-in reports
  - by program, NSF organizational units
- Many requests for ad hoc reports
  - from NSF mgt, auditors, Congress, OMB
  - taxonomies would help with classification and aggregation
  - current methods: keyword search, manual classification
- Ontologies could help automation of financial reporting (cf. XBRL, W. McCarthy (MSU))

Workflow Management

- NSF uses multiple, stove piped WFM applications
  - communicate via file transfers, shared DB
  - hardwired WF processes
  - inflexible, lack auto escalation, limited reporting, slow,
- Administrative ontologies could help with:
  - formalization/automation of WF processes
  - automation of conflict-of-interest checking

Entity Resolution

- Ontological categories not enough
- Need instances: PIs, schools, reviewers
- Currently many of these appear multiple times in DB (variant spellings, duplicates)
- Duplicate persons, schools:
  - difficult to cluster, aggregate records
  - difficult for reporting
  - difficult for COI detection
- Need entity resolution (duplicate removal)

Ontologies Needed

- Administrative:
  - NSF record types, entities, actions, states
  - Will require NSF internal development
  - Relatively small (1K – 10K concepts)
- Scientific
  - Taxonomies of scientific research areas
  - Mostly not NSF specific
  - Could use mostly external ontologies
  - Very large (100K concepts? more?)

Future NSF Records Management

- Bibliographic database of NSF supported publications
  - citations first
  - abstracts to follow ?
  - full text ???
- Fully electronic records
  - digital signatures
  - document formats ??? .pdf, XML, .odf ? ....

Future NSF Records Management

- Electronic organization charts
  - to support workflow management
  - automate signature/authorization checking
  - automate escalation, rerouting
  - additional storage, maintenance, software requirements
- Additional security requirements
  - Encryption, access control, extract tracking

Acknowledgements

- I would like to thank George Strawn (NSF CIO), and Maria Zemankova (NSF CISE IIS Program Director) for many discussions of these topics.