Engaging and Connecting Faculty: Research Discovery, Access, Re-use, and Archiving

Janet McCue and Jon Corson-Rikert
Albert R. Mann Library
Cornell University

CNI Spring 2007 Task Force Meeting
April 16, 2007
“An essential process is the joining together of subcultures when a wider common language is needed. Often two groups independently develop very similar concepts, and describing the relationship between them brings great benefits. Like a Finnish-English dictionary, or a weights and measure conversion table, the relations allow communication and collaboration even when the commonality of concept has not (yet) led to a commonality of terms.”

Focus

- NSF Small Grant for Exploratory Research (SGER) to explore collaboration between scientists and research library staff
- Semantic web applications that create a virtual research community
Research questions

• When is laboratory-library collaboration feasible and desirable?
• How should the responsibilities of the library and the laboratory be balanced?
• What significant challenges and costs are associated with various activities?
• What is a conceptual model for collaboration that would allow us to focus on promoting the preservation and discovery of resources valuable for interdisciplinary research?
Library goals

Direct – better information
- More timely, accurate and complete documentation of data
- Formats that can more easily be preserved
- More standard metadata to promote discovery

Indirect – better community
- Promote awareness and exchange of data prior to and after publication
- Promote collaboration within & across disciplines
- Promote new methods of publication that include data
- Keep libraries front and center in the academy
Cornell Language Acquisition Lab (Barbara Lust)

- Large digitization requirement from analog audio tapes
- Collecting and standardizing metadata
Upper Susquehanna Applied Ecology Program

- Multiple investigators
- Born-digital data
“NSF’s data policies will be redesigned as necessary to mitigate existing sociological and cultural barriers to data sharing and access, and to bring them into accord across programs and ensure coherence. This will lead to the development of a suite of harmonized policy statements supporting data open access and usability. NSF’s actions will promote a change in culture such that the collection and deposition of all appropriate digital data and associated metadata become a matter of routine for investigators in all fields. This change will be encouraged through an NSF-wide requirement for data management plans in all proposals. These plans will be considered in the merit review process and will be actively monitored post-award.”

Cyberinfrastructure Vision for 21st Century Discovery (March, 2007)
The approach

• Involve the library early on
• Provide tangible, low-barrier, near-term assistance
• Know the culture (librarians with disciplinary expertise)
• Offer a blend of services through one point of contact
• Offer patterns, tools, and training rather than data processing
• Demonstrate benefits for collaboration and exchange as well as long-term stewardship
Think from the faculty perspective

Facilitate a primary goal: publication
  • Communicates the science
  • Validates the research project
  • Enhances personal and lab reputations (especially for young investigators)
  • Lays the groundwork for future funding

New: help faculty comply with funding agency requirements for data management plans
Provide synergistic services

Faculty know how to collect and analyze data

- Help with collaboration, data sharing, special analysis or display (e.g., GIS), and metadata creation
- Publicize the availability of additional services such as high-performance computing, large dataset storage

Faculty are familiar with the publication process

- May fall short on data formats and metadata to enable data to be re-used and ultimately preserved
- May appreciate help preparing data for a repository
Provide wikis

Goals

• Internal project use – minutes, posters, mini-grants
• Informal data review, comment, and sharing
• Handoffs between data producers and modelers

Results

• Ideally one willing, tech-savvy person on the project to managing and re-organize content
• Asked for more storage for data & documents
• Word gets out -- three other projects now have wikis
Provide portals

Public access to

- Background information
- Research plans
- Participants
- Activities
- Public datasets
  (available now or anticipated)
Provide staging repositories

- Expand from the wiki to support larger data sets and/or data backup
- Provide project-level access and exchange prior to publication
- Coordinate with statistical consulting, computer and information science, and high-performance computing for value-added services
- Provide a platform for appraisal & selection prior to submission to an institutional or disciplinary repository
- Begin to tap the services of the Grid
Provide metadata tools & training

Local expertise geared toward domain standards
  • Open Language Archives Community (OLAC)
  • Ecological Metadata Language (EML)
    • Community has developed tools for metadata creation (Morpho) and management (Metacat)

Library hosts, customizes, and teaches these tools
  • Workshop attended by three of four USAEP PIs
  • Researcher enters metadata only once, in the format closest to their domain
  • Conversion process to create Dublin Core for DSpace
Metadata Challenges

• Varied formats require specialized editing and storage software
• XML DTDs and schemas enforce syntactic correctness but most content is in free-format data literals
• Meaning may be implicitly rather than explicitly encoded, visible only via local transformation tools
• Any cross-references between elements can't be relied on to be consistently interpreted
Promising approaches

Ontologies

- Better support for automated processing
- Capture object relations at creation time through explicit properties, not just free-text values
- Global resource identifiers for interoperability

Store independent statements

- Collectively encode meaning as well as structure
- Transform metadata into necessary output formats on demand
LiLaC Conceptual Framework

Level 1
Search Engines, Harvesters, and Inter-Repository Exchange

Level 2
Domain and Institutional Repositories
- OLAC metadata
- VCLA staging level
- MIT CLAL

Level 3
Staging Repository (library service)
- VCLA staging level
- USAEP staging level

Level 4
Individual Laboratories / Researchers
- Project participants

Institutional Repository
DSpace, etc.
(data and metadata)

Library-Laboratory Collaboration for Research Data
NSF 0437603, “SGER: Planning Information Infrastructure Through a New Library-Research Partnership”
VI VO – integrated discovery

- Provides consistent and highly-visible information on researchers, facilities, grants, publications, and data
- Makes content accessible independently of Cornell's administrative structure
- Provides a Google-like search while adding rich contextual navigation
- Includes information on research in progress
VIVO as harvester and distributor
Barrett, Christopher | Professor | Applied Economics & Management profile | Barrett web page

faculty member in
- Applied Economics and Management (AEM) | Cornell department | AEM web page | Applied Economics & Management courses
- International Programs (IP/CALS) | academic program office | IP/CALS web page

member of graduate field
- International Agriculture and Rural Development | graduate field | Graduate School Field web page | MPS/International Agriculture & Rural Development web page
- International Development | graduate field | Graduate School Field web page | MPS/International Development web page
- Natural Resources | graduate field | Graduate School Field web page | Natural Resources Graduate program web page | MPS (ALS) with Peace Corps Option web page

primary investigator of
- AFRICAN FOOD SECURITY AND NATURAL RESOURCE MANAGEMENT - SECOND COHORT FIELD RESEARCH | Research Grant
- CONCEPTUAL AND STATISTICAL FOUNDATION FOR THE ESTIMATION OF POVERTY TRAPS | Research Grant
- CORNELL AFRICAN FOOD SECURITY AND NATURAL RESOURCE MANAGEMENT PROGRAM | Research Grant
- MODELING SOIL FERTILITY REPLENISHMENT INVESTMENT BEHAVIOR AMONG SMALLHOLDER FARMERS IN THE WESTERN KENYA HIGHLANDS | Research Grant

co investigator of
- CORNELL AFRICAN FOOD SECURITY AND NATURAL RESOURCE MANAGEMENT PROGRAM | Research Grant
- HOMEOSTASIS AND DEGRADATION IN FRAGILE TROPICAL AGROECOSYSTEMS | Research Grant

teaches
- Contemporary Controversies in the Global Economy | 3 credit course | AEM 200
- Development Microeconomics Graduate Research Seminar | 1-3 credit course | AEM 785
- Microeconomics of International Development | 3 credit course | AEM 762
taught courses
- Contemporary Controversies in the Global Economy | 3 credit course | AEM 200
- Development Microeconomics Graduate Research Seminar | 1-3 credit course | AEM 765
- Microeconomics of International Development | 3 credit course | AEM 762

research areas
- applied economics | collaborative research area (CALS)
- biocomplexity | collaborative research area (CALS)
- biodiversity | collaborative research area (CALS)
- food system and/or policy | collaborative research area (CALS)
- international agriculture | collaborative research area (CALS)
- 6 more

affiliates in
- African Food Security and Natural Resources Management (AFSNRM) | academic program office | AFSNRM web page
- Center for the Study of Inequality (CSI) | research center | CSI web page
- Institute for African Development | research institute | IAD web page

has geographical research area
- Africa | continent

speaker at
- AEM Roundtable: The Disintegration of the Doha Development Deliberations | AEM Seminar | Wed 11/01/2006 1:30pm
- Agricultural Technology, Productivity, Poverty and Food Security in Madagascar | CIIFAD Seminar | Wed 02/15/2006 12:20pm
- Social Network Capital, Economic Mobility and Poverty Traps | ISS Networks Seminar | Fri 10/20/2006 10:30am
- Stagnant Smallholder Agriculture? Rice Yield Dynamics in the Highlands of Madagascar | CIIFAD Seminar | Wed 03/14/2007 12:20pm

Chris Barrett teaches and conducts research in the areas of international development, environmental and resource economics, international trade, markets and price analysis, agricultural production and distribution, and applied econometrics. Professor Barrett has published more than 125 journal articles, books, and book chapters, and has been principal investigator or co-principal investigator on research grants totaling more than $15 million.

There are three basic, interrelated thrusts to Professor Barrett's research program. The first concerns poverty, hunger, food security, economic policy, and the structural transformation of low-income societies. The second considers issues of individual and market behavior under risk and uncertainty. The third revolves around the interrelationship between poverty, food security, and environmental stress in developing areas.

Keywords: development economics, conservation, sustainable development, international agriculture, poverty reduction, agricultural economics, environmental and resource economics, markets analysis, food aid, food insecurity
Africa | continent

includes
- Angola | country
- Benin | country
- Botswana | country
- Burkina Faso | country
- Burundi | country
- 10 more

associated with research by
- Assie-Lumumba, N'Dri T. | Professor of African Studies | Africaana Studies and Research Center profile
- Barrett, Christopher B. | Professor | Applied Economics & Management profile | Barrett web page
- Blake, Robert W. | Professor | Animal Sciences profile
- Christy, Ralph D. | J. Thomas Clark Professor | Applied Economics & Management profile
- Groen, Vernom | Visiting Professor | Plant Breeding & Genetics profile
- 7 more

associated with activities of
- Africaana Studies and Research Center (AS&RC) | research and educational resource center | AS&RC web page | AS&RC courses

affected by project
- Biofortified Crops for Improved Human Nutrition (HarvestPlus) | Impact statement | Miller, Dennis Dean
- Building collaborative educational opportunities for conservation and sustainable development | Impact statement | Cassole, James Philip
- Development and application of advanced technologies for monitoring rare and endangered species in marine and terrestrial environments | Impact statement | Clark, Christopher W
- Development of Gramene, a comparative mapping database for cereals | Impact statement | McCouch, Susan Rutherford
- Food Security in Freetown: The Role of Urban/Peri-urban Agriculture (UPA) | Impact statement | Kiroma, Margaret M
- Inference of biological function and mechanism of evolutionary change from comparative genomic data | Impact statement | Aquadro, Charles
- Mineral requirements in vulnerable pediatric and adult population groups | Impact statement | O'Brien, Kimberly C
- Phylogeny and ecology of bees in New York state | Impact statement | Dendforth, Bryon Nicholas
- Recombinant commensal bacteria as therapeutics | Impact statement | March, John C
- Strategies and Analyses for Growth and Access (SAGA) | Impact statement | Sehn, David

sub-regions
- Central Africa | trans-national region
- Cote D'Ivoire |
- Eastern Africa | trans-national region
Property editing

### Object (entity-to-entity) Properties

<table>
<thead>
<tr>
<th>Domain</th>
<th>Property</th>
<th>Range</th>
<th>actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property:</td>
<td>Person (research areas) Research Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wClass:</td>
<td>Person (co-investigator of) Financial Award</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entity:</td>
<td>Person (collaborating investigator) Financial Award</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualifier:</td>
<td>Person (compilable listings) Enumerated Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save:</td>
<td>Person (research areas) Research Area</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Barrett, Christopher B.</th>
<th>research areas</th>
<th>Food system and/or policy</th>
<th>Edit</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>Rural development</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>Sustainable development</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>International development</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>Resource policy and management</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>Applied economics</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>research areas</td>
<td>Risk analysis and assessment</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>primary investigator of</td>
<td>Conceptual and statistical foundation for the estimation of poverty traps</td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Barrett, Christopher B.</td>
<td>primary investigator of</td>
<td>Modeling soil fertility replenishment investment behavior among smallholder farmers in the Western Kenya Highlands</td>
<td>Edit</td>
<td>Delete</td>
</tr>
</tbody>
</table>
VIVO architecture

- Manage content in small units
- Integrate content from multiple sources
  - Import from databases of record (OHR, OSP, events, news)
  - Allow direct entry and update through faculty reporting
- Leverage semantic structure for display, filtering and reporting, including areas of new development
  - Standardizing on OWL data model
  - OWLIM for dynamic updates of inferred classes
  - SPARQL for more complex queries than SQL
Institutional relationships

• Database exposed via web services on multiple websites
• Integrated content offered back to distributed units
• VIVO data model (Vitro) selected for new Cornell-wide faculty reporting database
  – collaborative development project starting up for the faculty self-reporting interface and workflow
  – Colleges keep editorial control while gaining wider exposure for their faculty and research
Next steps

• Collaborate across institutions
• Explore interoperability and semantic integration
• Leverage the value of the grid through improved semantics as well as computation, workflow, and networking
Agents of Integration
Questions?
Citations


