Repository requirements:
Documentary Linguists & Community Members

Arienne M. Dwyer

Co-Director, Institute for Digital Research in the Humanities & Professor of Linguistic Anthropology, University of Kansas Visiting Professor of Digital Humanities, CUNY Grad Center

Research powered by:  
A.M. Dwyer, Repository Reqs
Stakeholders

- indigenous community members
- academics of various stripes (here, documentary linguists)
- academic institutions
- the public
But if archiving were just a technical issue, linguists would be...

- Blogging details of ongoing projects
- Flooding repositories with data sets to make them web-accessible
- Routinely sharing data sets with other scholars and collaborating with them
- Pursuing external funding to do the above.
Barriers to participation

are as much social as they are technical

so...

a focus on maximizing collateral benefits and minimizing collateral damage

A.M. Dwyer, Repository Reqs
Maximizing participation

- Perceived usefulness
- Privacy and Recognition
- Graded Access
- Clear policies (governance etc.)
Reward structures of academia

(for, say, documentary linguists)

- Print publications
- single-authored over collaborative
- rapid publication of theoretical analyses
  (where theory often = maximally generalized)
**Typical docuLinguist trajectory**

(including indigenous community linguists)

- months/years of building consensus on projects
- multiple months/years of documentation (D)
- $D \times 10$ for basic analysis (not just annotation)
- produce preliminary products for community
- meantime: maintain community relationships, continually revisit and renegotiate agreements...

DocuLinguists enjoy this process, but ...

A.M. Dwyer, Repository Reqs
Data analysis/publication are the reward

• After all this effort and intellectual contributions,
  – finally! the transcribed data can be analyzed
  – analytical/theoretical products can be created.
  – These are the main rewards for the docuLinguist.

• If instead: after all this effort and contributions
  – you ask the docuLinguist to donate the data,
  – you are potentially taking away the main reward

• So: what incentives could be created?

A.M. Dwyer, Repository Reqs
The ideology of open access

- Open access/source
- increases knowledge
- everyone benefits
- encourages participation

A.M. Dwyer, Repository Reqs
Culture clash!

- Openness is for suckers
  - what is free isn't worth anything
  - sharing invites intellectual property theft
  - (But IP laws result in less sharing of info)

- Things can go wrong:
  - data theft, data degradation (malicious or other)
  - political, commercial, religious mis-use
3 possible responses:

• Let's share everything right now!
  • This is world knowledge and heritage.
  • The collaboration and tools will speed our work.

• Let's keep the material to work on ourselves!
  • We worked decades to create it, we should benefit.
  • Basic research is always undervalued compared to analysis. If we don't do analysis too, we're sunk.

• Let's work on it for 2 years, get our publications out, and then set the data free!
  • Everyone's happy (?!?!??)
Data embargos

- Short-term embargos may be useful;
- Long-term embargos stifle scholarship.
- But with no embargo, IP likely to be scooped
  - particularly important for dissertators and junior scholars
  - book contracts, future grants often depend on producing the earliest product
  - But this is beginning to change....
Growing scholarly incentives for sharing data
(not just publications)

- NSF, SSHRC require data mgmt plan (NEH, DFG nearly do)
- Digital „products,“ not just publications
  - datasets & software
For instance:

- An individual collected nearly 600 mss. from Chinese Turkestan around 1930.
- Steady stream of pubs (incl. fascimile annotations and translations) 1935 to 1997;
- Entire corpus donated to a library in the early 90s.
- And who is this generous, highly prolific scholar?
The most productive scholar: a diplomat

Gunnar Jarring - Swedish ambassador.

not an academic.

A model of open access, with a terrific publication record!

Implications

• Academia provides disincentives to Open Access
• As a non-academic, Jarring wasn't dependent on the academic evaluation system.
• Linguists can play an active role in creating incentives to en-value data sharing
Participation entails changing private practices

- **Public policy issues**: Repositories...
  - their creation, maintenance, control;
  - valuation of the work in them.
- **Private practices**: Participation in...
  - making data and scholarship public;
  - valuation systems for new forms of scholarship
- **Negotiating and redefining resource value**
  i.e. its social and academic capital

A.M. Dwyer, Repository Reqs
Three ways forward:

1. Build in incentives into data repositories

2. Create individual incentives for data sharing

3. Create institutional incentives
   - licensing arrangements share costs of repositories
1. Built-in incentives:

- Data originers & communities determine access.
- Protecting IP (which means clearly specifying rights and citation)
- Preventing data theft, data degradation
- Altmetrics built into archive (e.g. citation generator)
2. Individual incentives: Altmetrics

- Recognize expanded types of scholarship
  - annotated databases, data sets, programs
  - Via universities: posting boilerplate evaluation metrics (in US „P&T” text)
  - Via scholarly societies
- Web badges, collaboration changing the 'gold standard' of the single-authored monograph
- Use altmetrics & web badges to assess impact of curated data sets, posts, and more formal papers
2. more individual incentives

- **Evaluation systems**
  - recognition of IP-added data, data sets
  - recognition of software and other creative work
  - recognition of alternative publications

- **Carrot/stick: participation in collaborations**
  - What will speed/enhance work?
  - In micro-collaborations (Buddhism dictionary)

- **Creating institutional/career/community rewards**

- **Making early and iterative archiving the norm**
Carrots and sticks
3. Institutional incentives for data sharing

- Recognition of the IP of data sets:
  - encourages collaboration
  - Collaboration provides the basis for successful external funding proposals.

- Internal grant mechanisms > external proposals

- Participation in consortia
  - minimizes costs, shares expertise, develop transnational guidelines
  - Brings prestige.

A.M. Dwyer, Repository Reqs
Summing up

- **Participation**: many barriers are social
- **Security**: perceptual vs. technical
- **Access**: graded
- **Consultation**: with data originators (what rewards would be useful?)
What would it take?

Answering the following questions first

- Who controls the data?
- Who can access which materials?
  - Clear graded access policies
- Are alternative views/perspectives allowed?
  - alternative interpretations
  - alternative text renderings (orthographies, transcriptions)
- Can the annotated data be peer-reviewed?

A.M. Dwyer, Repository Reqs
Access, control, and rewards

• Many of the previous questions are technically solved, but ...

• our challenge is to implement a rewards system
  – based on data originators' suggestions (community members, linguists etc.)
  – to not assume that everyone's model of open data sharing is universal, instead:
  – Data sharing is acceptable under what conditions?

A.M. Dwyer, Repository Reqs