Working Groups Session A: who is our community?

IT/Big Data Group
General Discussion Themes

- Who is our community?
- What cyberinfrastructure are in place?
- Cyberinfrastructure challenges
- Education and training
- Funding and sustainability
Who is our community?

- A broad collection of organizations (e.g. LSA, Max Planck, UT Austin, U Hawaii, SIL, others)
- While there are orgs. and a community, individual somewhat isolated scholarship is still common; “fragmented”
- Organizations exist that can help to link institutes, individuals etc.
- This group/project has potential to facilitate CI linkages re. publishing and citation
Cyberinfrastructure challenges

- Basic issues such as **character encoding and formats**
- Data volumes can be “large” relative to available resources
- **Data discovery**
- Knowing what is possible and feasible... then making people aware
Education and training

- Need for education and training on value of publication/citation and the methods involved
- Need to raise awareness re. what is possible - get people excited about how publication, citation, CI in general can enhance their work
- **Workshops**, “hackathons”
- Spread the word through department chairs
- Integrate into coursework
Funding and sustainability

- A real challenge
- Good data management costs; some components are less but humans are important and expensive
- Many are self-fund; grants help but not for long-term preservation
- Subscriber model is popular but makes resources less accessible to individuals/small programs
- Model: **combine grants with small contributions**; taking payment can be a challenge
Working Groups Session B: journals and our community

IT/Big Data Group
Journals and Our Community

- Developing and communicating clear models
- Establishing value propositions
- Incentives (carrots)
- Citation tracking
- Submission
Developing and communicating clear models

- In many cases “supplemental materials” should actually be a data paper
- Need to ensure that there are clear models the define what publication, data publication, supplemental etc. are
- Will help researchers to better understand value of publication/citation and publishers to more easily manage
Establishing value propositions

What are value propositions:

- Practical values (jobs, promotion etc.)
- Inherent value (right thing to do)
- Avoid fraud
- Funding opportunities emerging
Incentives (carrots)

What are the (possible) incentives?

- Opportunities to experiment, play - experimental data/application platform
- Part of hiring/promotion process... still not part of culture
- Look to fields like medicine - accepted intellectual contribution
- Need to actively publish and cite - metrics/indexes (doi tracking) will then indicate publication record and citation stats - legitimizes
- Educate chairs at annual LSA meeting
Citation tracking

- Until fairly recently, linguistics didn’t do much with citation tracking. no impact factors;
- Clear that DOIs are the key to plugging into the global system
- LSA’s Committee for Scholarly Communication in Linguistics working on report – could be a platform for this topic in general and specifics like citation tracking model
Submission

- Easy data submission is key...
- ... formats are important for management and preservation
- Need a balance between easy and effective (including cost)
- Challenge: Word is common but proprietary; ODT potentially good but little traction
ACTION ITEMS from the IT/BigData Folks

Ideas for next steps

- Workshops/training at LSA 2017: talk about top-level things with key messages → spread the word; could even present citation guidelines (ACTION: Develop training materials and organize session) (Owner = ?)
- Establish clear standards about how you add a data collection in a bibliography and how you cite it in text (style): Start by surveying other citation standards and using results to create our own set of standards. (ACTION: review existing standards and propose draft standard for use by journals) (Owner = ?)
Approaching the discovery issue: existing archives, data centers could meet and compare their discovery systems and see which types of fields for discovery should be standard for the entire field → (ACTION: Organize OLAC meeting to develop draft standard; promote i.e. LSA) - Maybe have some kind of presentation about OLAC at our April workshop as they already have a foundational system in place? (Owner = OLAC, ?)

Then use technology to create “Union Catalogue” (OAI-PMH etc.) (Owner = OLAC, ?)
Scholarly misconduct regarding citations, bibliographies hasn’t been discussed much here: Maybe we could do something on exploring the ethics and norms of data citation and attribution? See the Collaborators’ Bill of Rights in digital humanities, which was mentioned by Clement. *(ACTION: Write draft statement of acceptable conduct re. Citation and Attribution; review with community; promote i.e. LAC) (Owner = ?)*

Following basic steps above, start to work on on details such as an inventory of data types and formats in preparation for dealing with more detailed interoperability and archiving solutions. (Owner = ?)