

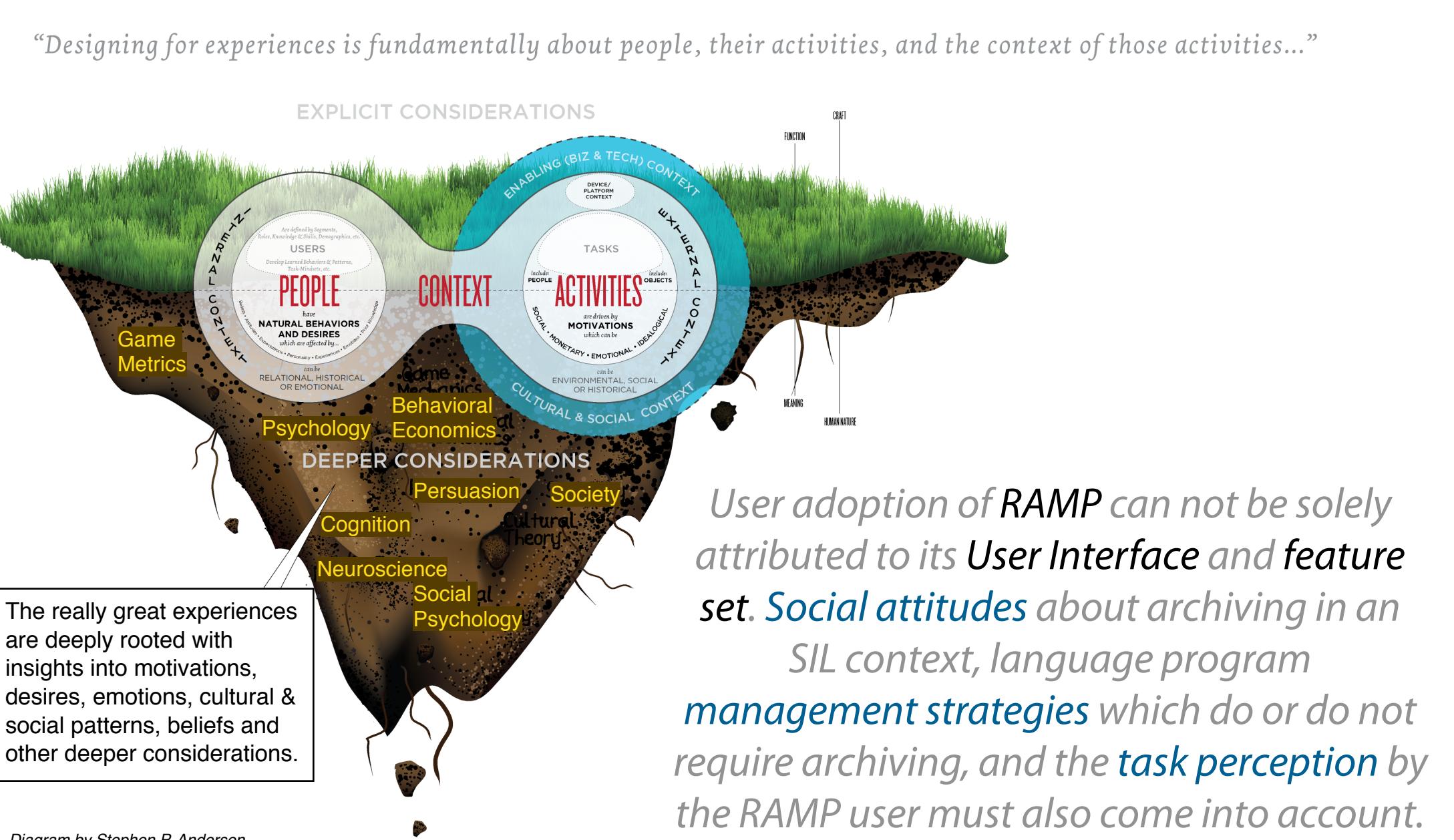


# Challenges of implementing a tool to extract metadata from linguists: The use case of RAMP

Hugh Paterson III & Jeremy Nordmoe



## 1 The Archiving Experience



### The DSpace "effect"

Of the 37 DSpace submitters in 2012, 13 of them have an archiving role.

- 60% of DSpace submitters are non-archivists.

This stands in contrast to all 141 submitters, of whom 38 have an archiving role and 13 of these used DSpace.

- 60% of SIL archivists don't use DSpace for submissions.

Of the top 12 repeat DSpace submitters, 10 of them have roles in archiving or publishing. Others were, 1 each: training, and fieldworker. - Those SIL archivists who use DSpace use it a lot (or for batches).

For 12 DSpace submitters (none of whom have an archiving role), the DSpace experience was the only digital interaction with the archive that they had in 2012 (they were non-RAMP users). All of these users only contributed once. They were all contributing their own content.

- 4 of these 12 would go on to contribute materials to the archive via another method (non-digital, or through another individual submitting the content). But again not as a user of any digital system for archiving with SIL.

- 5 of these 12 had already submitted something to the archive either through another individual or through non-digital means when they tried DSpace for the first time.

They have a belief that archiving is important, or their works are in corporate publishing workflows which result in archiving.

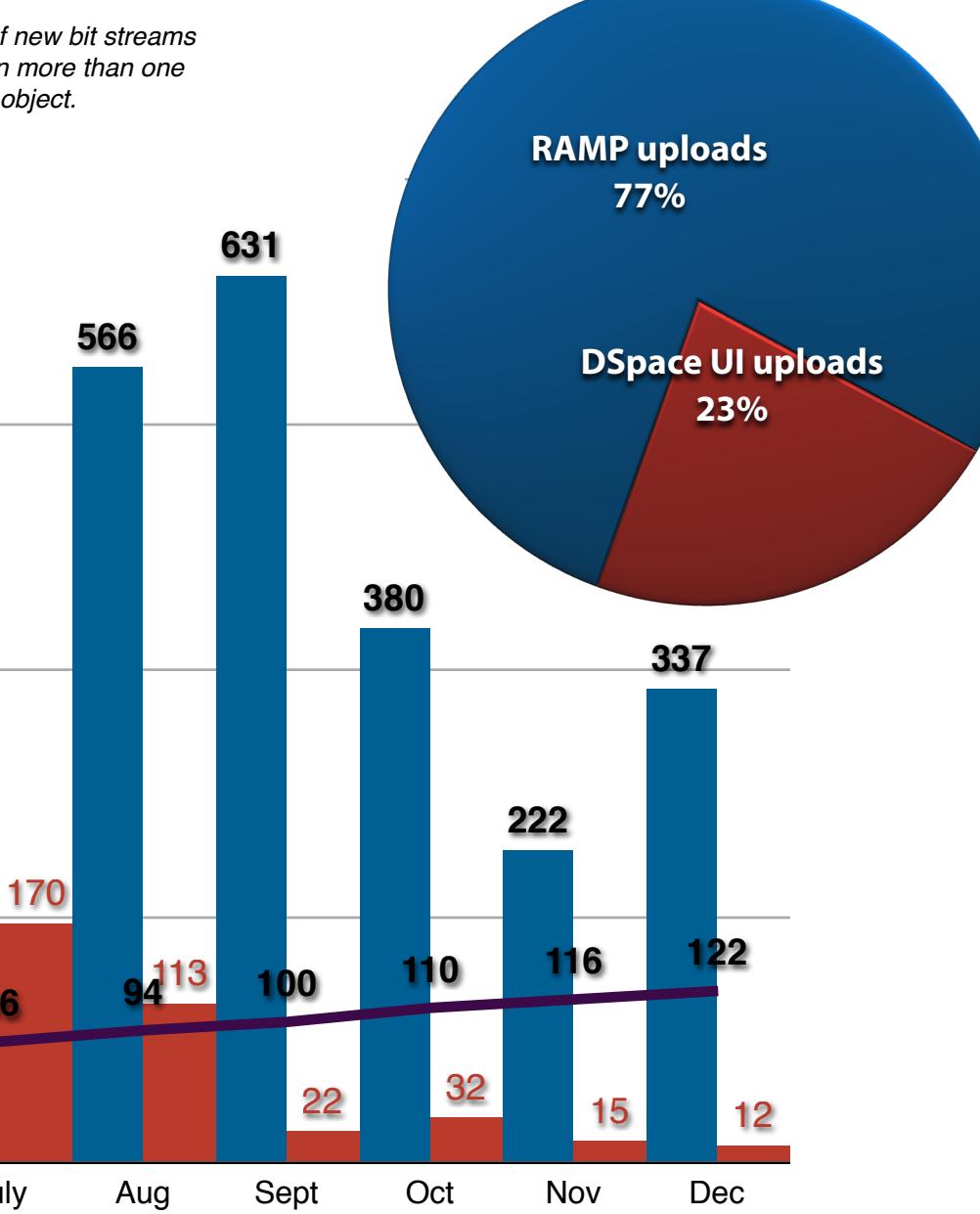
User adoption of RAMP can not be solely attributed to its User Interface and feature set. Social attitudes about archiving in an SIL context, language program management strategies which do or do not require archiving, and the task perception by the RAMP user must also come into account.

### The RAMP "effect"

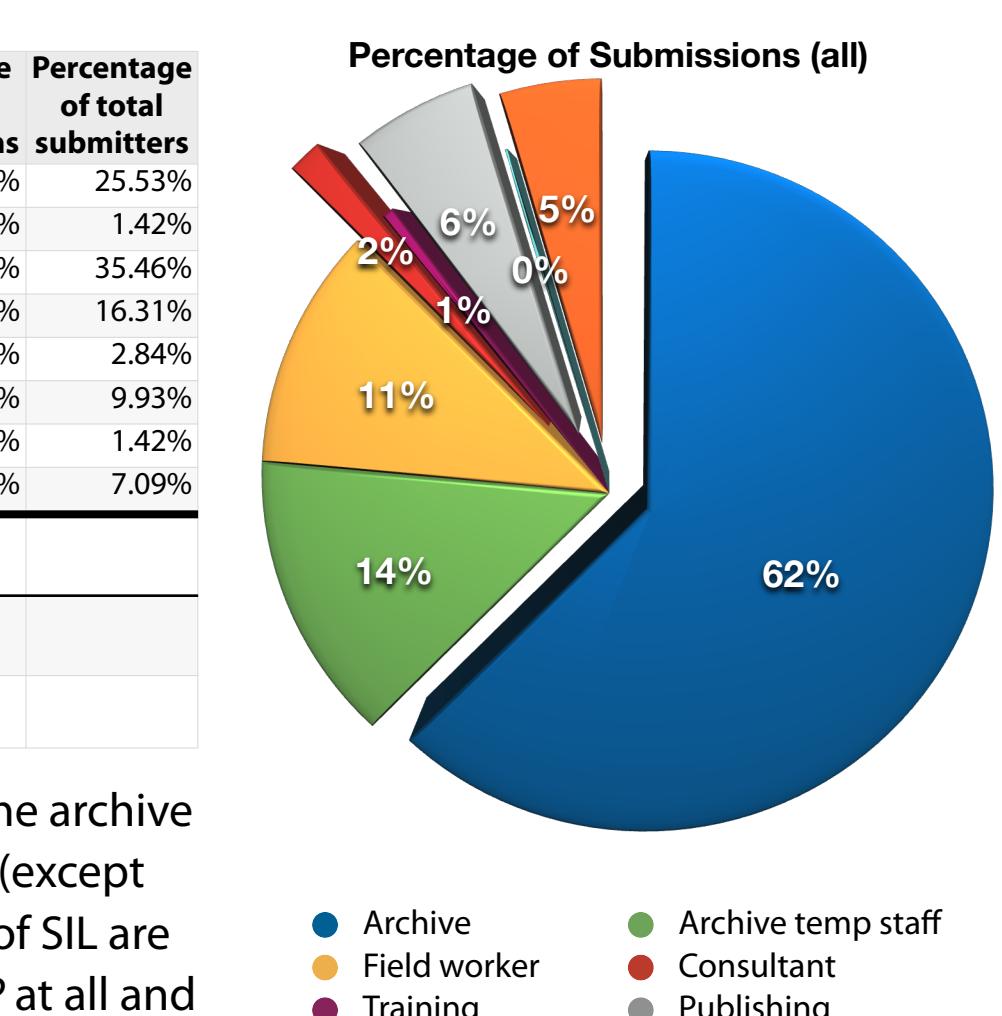
Prior accession rates over the last 10 years have averaged between 1,500 and 2,000 items per year.

RAMP submits in 2012 alone produced 4,223 new accessions (bit streams). RAMP is a desktop application which was created to overcome complexities in DSpace UI so that field workers could directly submit to the archive.

Direct to DSpace v.s via RAMP ratio



RAMP contributed an increased capacity (200%) to accession materials, but was it equally successful in reaching its intended audience?



Of the 141 people who have made submissions to the archive in 2012, only 122 of them used RAMP. Each job type (except archive-temp) and all six major administrative units of SIL are represented by those 19 users who did not use RAMP at all and made submissions to the archive; 12 of those 19 only made 1 DSpace submission.

78% of all submissions to the archive in 2012 were made by SIL staff with a specific role in archiving. In 2012:

- 2.3% of all SIL staff globally made RAMP submissions
- 3.5% of language development staff with roles in active projects made submissions via RAMP
- Repeat DSpace submitters, who have never used RAMP tend to be in publishing roles.

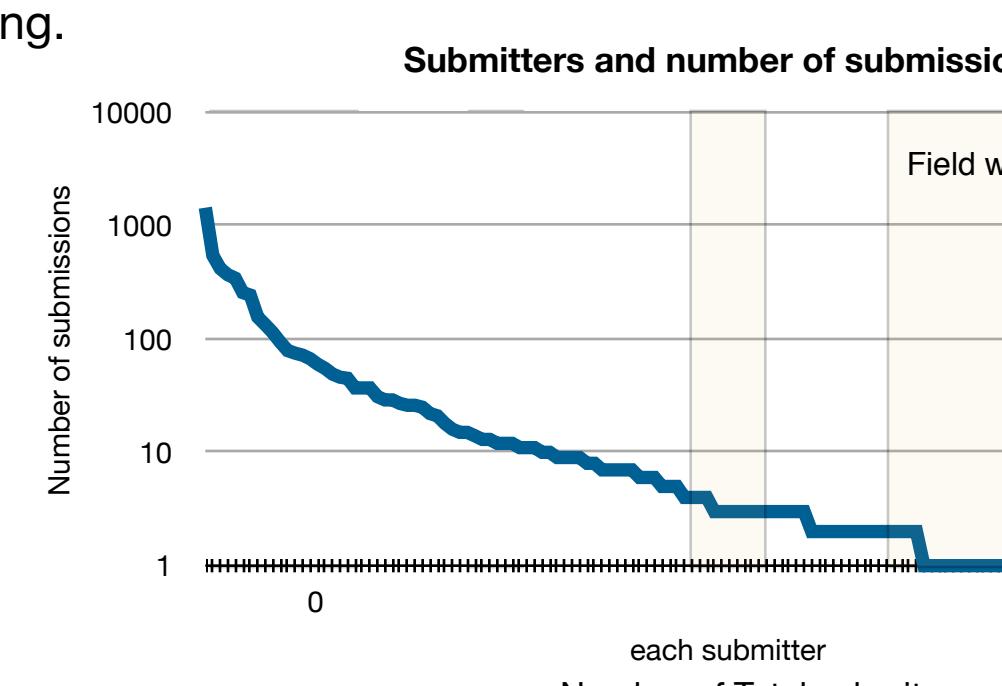
Based on the population of SIL staff working in language projects, it is not unreasonable to expect the user base of RAMP to exceed 2000 unique users per year.

Of the 122 RAMP submitters in 2012, 36 of them did not use DSpace and also only submitted one item. - 30% of RAMP users chose not to use the software again (having never compared it with DSpace).

- 12 were submitting objects on which they were not contributors, e.g. not author, not composer
- 12 continued to submit materials to the archive, but chose to do so through another person, or via a non-digital means.
- 6 individuals who had previously submitted items to the archive through another person, or non-digital means, chose to attempt to use RAMP, but had no desire to continue to use RAMP (or possibly further opportunity to use RAMP, the archive does not know).

8 of the top 10 RAMP users have a role in archiving.

Of the non-single use submitters, 38% (32 users), had an archiving role. If persons with a publishing role are added then it goes to 45% (39 users).



## 2 Social Attitudes of Linguists

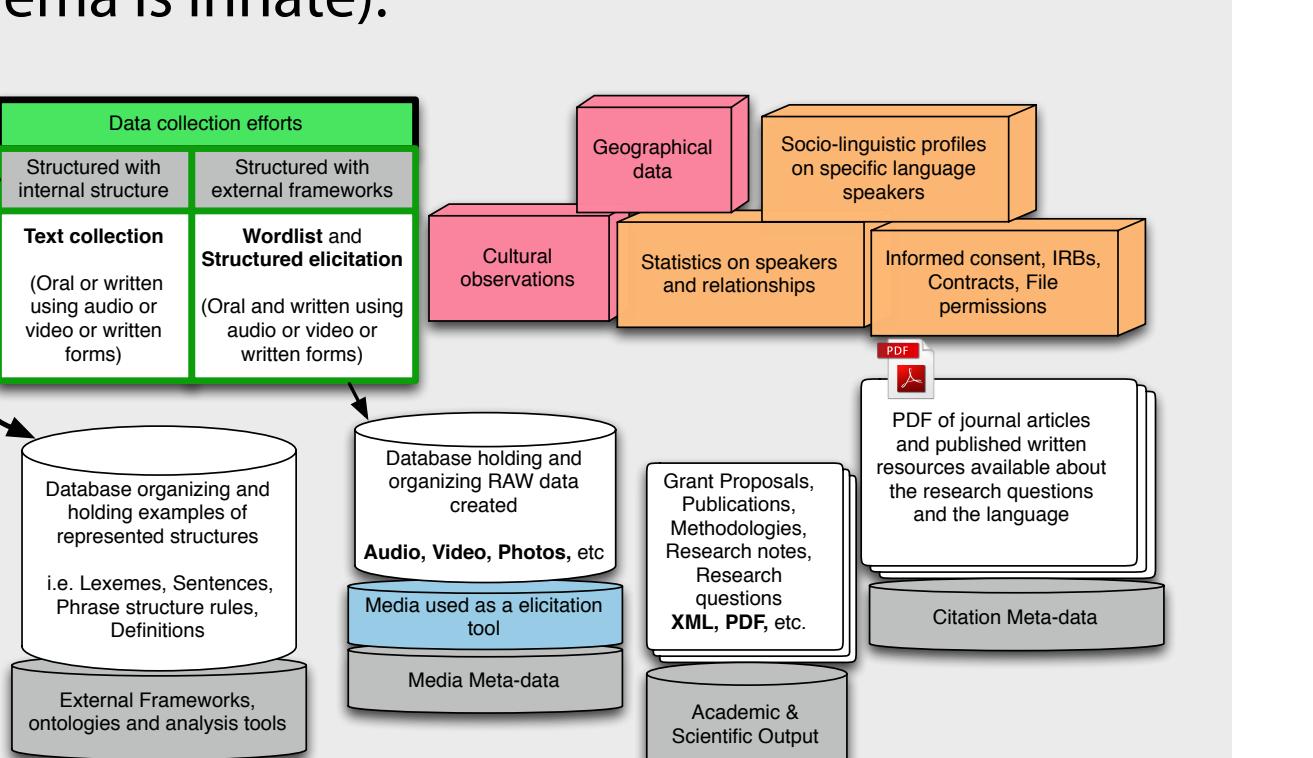
Nordmoe (2011) claims that archiving meta-schemas remain too complex for linguists... We find this objection un-grounded coming from linguists who devise meta-schemas for describing language... (Though we make no claim that any schema is innate).

Linguists use a variety of complex metadata schemas during their working day - though some linguists may be unaware of them. The user experience challenge for archivists is: can archivists access these data at the point of first use?

Other difficulties encountered:

- The archiving institution:
  - loses control
  - can not restrict consent access appropriately
  - can not publish content to open access in a timely manner
  - does not value certain types of content or will charge for access
  - does not maintain accurate records because all the information provided by the institution does not fit into the institution's metadata schema.
  - Confusion about the organizational structure of the archiving institution.
  - Opinions that archivists should only take place in the country where the language is indigenous spoken.
  - I have heard a linguist say "I hate Metadata". For a linguist to value the archive more, more than discovery metadata must be exposed about the data in the archive.
  - The I don't care attitude: "I'll just turn it over to the archive to do whatever they do."

Diagram conceptually showing Data collection effects.



### From when do these items come?

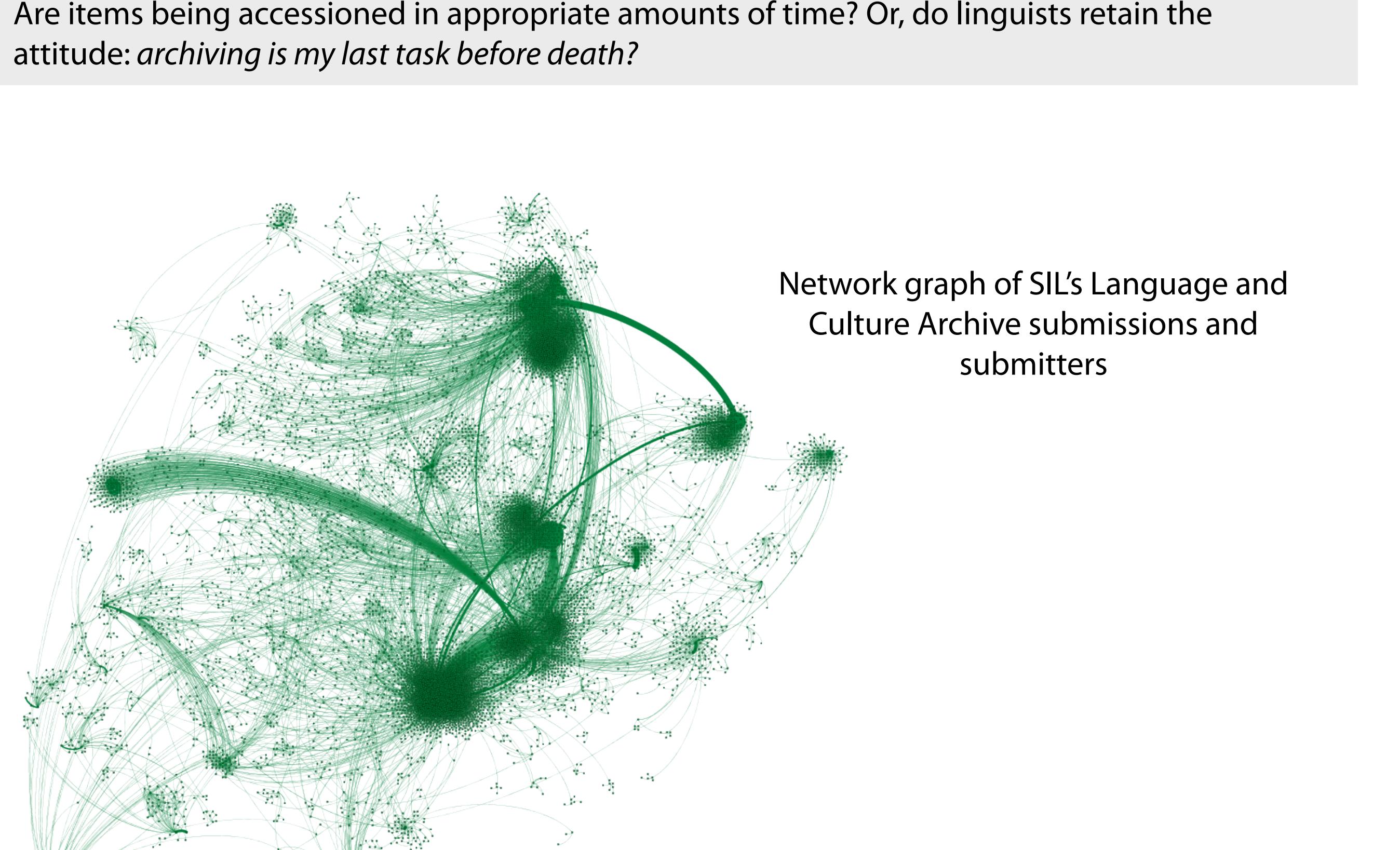
604 items archived in 2012 are known to have been created or published in 2012.

SIL has nearly 80 years of history working with minority language communities.

About 1 million relevant non-digital objects are estimated to exist in SIL networks.

About 50 million relevant digital objects are estimated to exist in SIL networks.

Are items being accessioned in appropriate amounts of time? Or, do linguists retain the attitude: archiving is my last task before death?



Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams: 13.97%. Percentage of staff: 11.98%. Percentage of people: 1.58%. Percentage of bit streams: 16.31%. Percentage of people: 2.84%. Percentage of bit streams: 0.55%. Percentage of people: 9.93%. Percentage of bit streams: 4.11%. Percentage of people: 4.51%. Percentage of bit streams: 7.09%. Percentage of people: 7.09%.

Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams: 13.97%. Percentage of staff: 11.98%. Percentage of people: 1.58%. Percentage of bit streams: 16.31%. Percentage of people: 2.84%. Percentage of bit streams: 0.55%. Percentage of people: 9.93%. Percentage of bit streams: 4.11%. Percentage of people: 4.51%. Percentage of bit streams: 7.09%. Percentage of people: 7.09%.

Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams: 13.97%. Percentage of staff: 11.98%. Percentage of people: 1.58%. Percentage of bit streams: 16.31%. Percentage of people: 2.84%. Percentage of bit streams: 0.55%. Percentage of people: 9.93%. Percentage of bit streams: 4.11%. Percentage of people: 4.51%. Percentage of bit streams: 7.09%. Percentage of people: 7.09%.

Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams: 13.97%. Percentage of staff: 11.98%. Percentage of people: 1.58%. Percentage of bit streams: 16.31%. Percentage of people: 2.84%. Percentage of bit streams: 0.55%. Percentage of people: 9.93%. Percentage of bit streams: 4.11%. Percentage of people: 4.51%. Percentage of bit streams: 7.09%. Percentage of people: 7.09%.

Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams: 13.97%. Percentage of staff: 11.98%. Percentage of people: 1.58%. Percentage of bit streams: 16.31%. Percentage of people: 2.84%. Percentage of bit streams: 0.55%. Percentage of people: 9.93%. Percentage of bit streams: 4.11%. Percentage of people: 4.51%. Percentage of bit streams: 7.09%. Percentage of people: 7.09%.

Number of RAMP submissions: 2,341. Number of bit streams: 3,409. Number of people: 734. Number of staff submitting work: 36. Percentage of total submitters: 62.49%. Percentage of total submissions: 25.53%. Percentage of bit streams