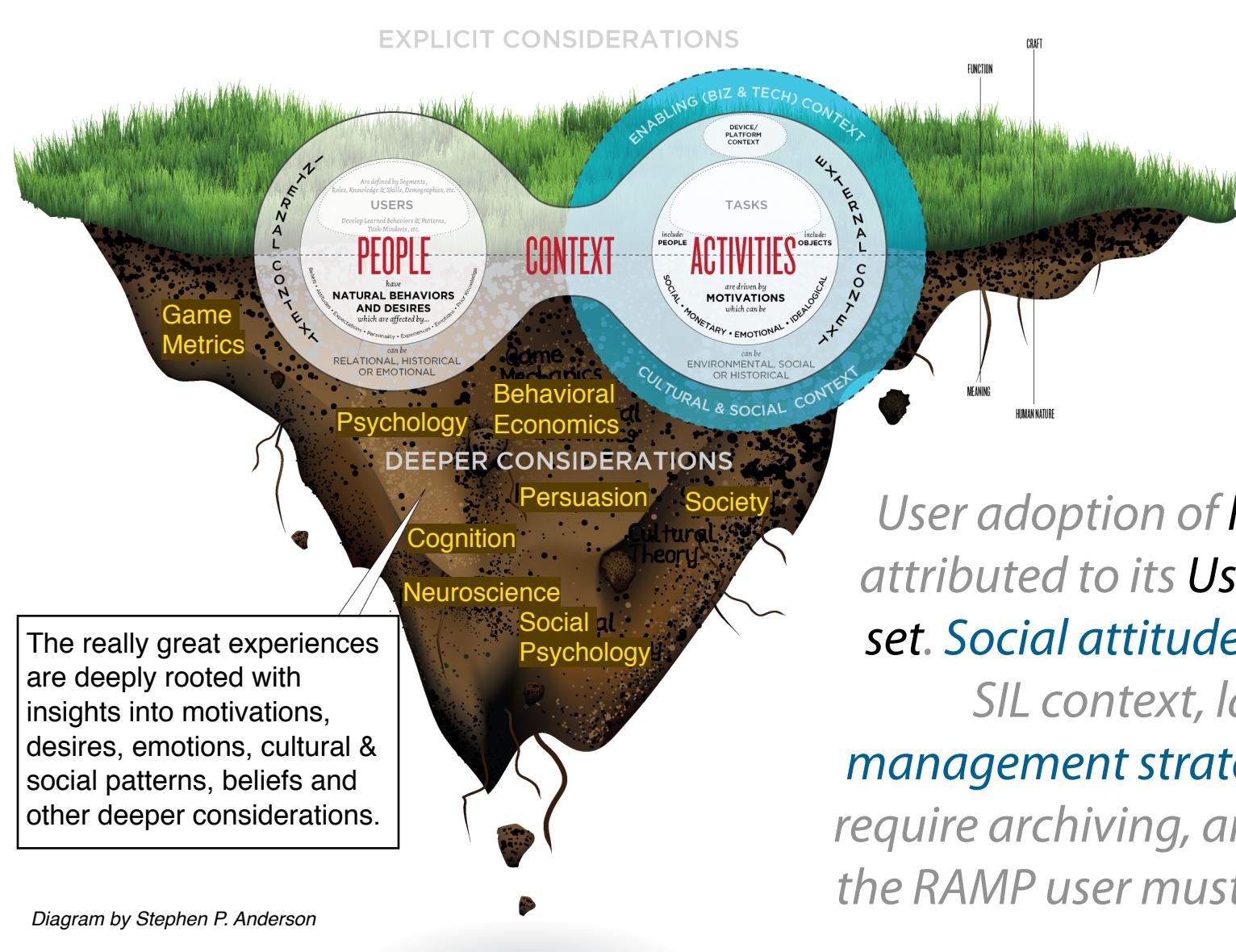


Challenges of implementing a tool to extract metadata from linguists: The use case of RAMP

Hugh Paterson III & Jeremy Nordmoe

1 The Archiving Experience

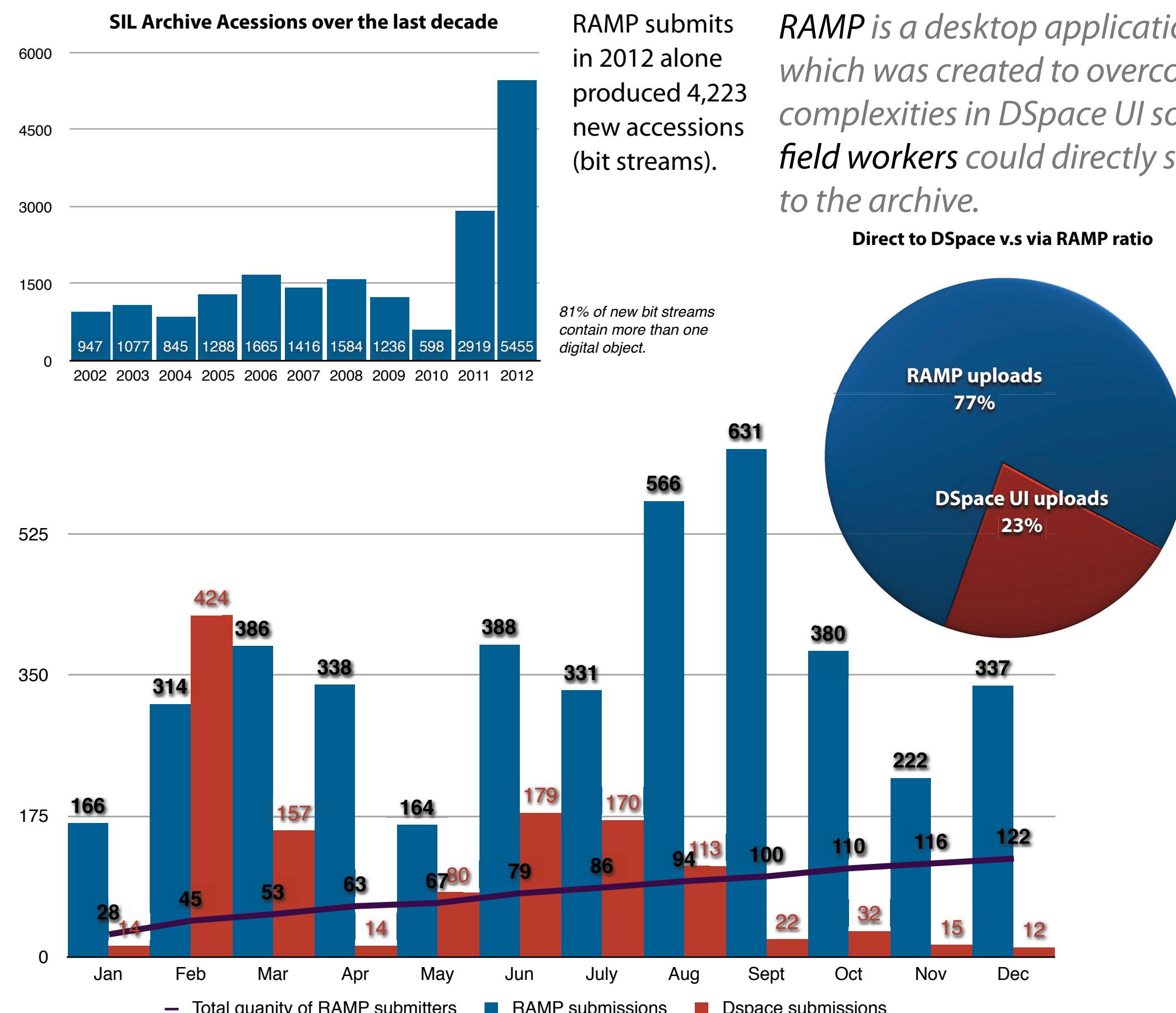
"Designing for experiences is fundamentally about people, their activities, and the context of those activities..."



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

Number of RAMP submissions	Number of bit streams	Number of people submitting a work	Class of SIL staff working	Percentage of total submissions	Percentage of total submitters
2,341	3409	36	Archive	62.49%	25.53%
734	762	2	Archive-temp	13.97%	1.42%
598	610	50	Field worker	11.18%	35.46%
83	86	23	Training	0.55%	2.84%
2	30	1	Publishing	0.55%	0.93%
14	14	14	Media publishing	5.61%	9.93%
5	6	2	Media services	0.11%	1.42%
238	246	10	Admin	4.51%	7.09%
Total number of submissions	5455	141	Total submitters of any kind		
Percentage of RAMP submissions submitted by field workers	4223				
Percentage of RAMP submissions submitted by field workers / DSpace Upload	1232				

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).

Of the 36 users:

- 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).

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 field worker. - 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.

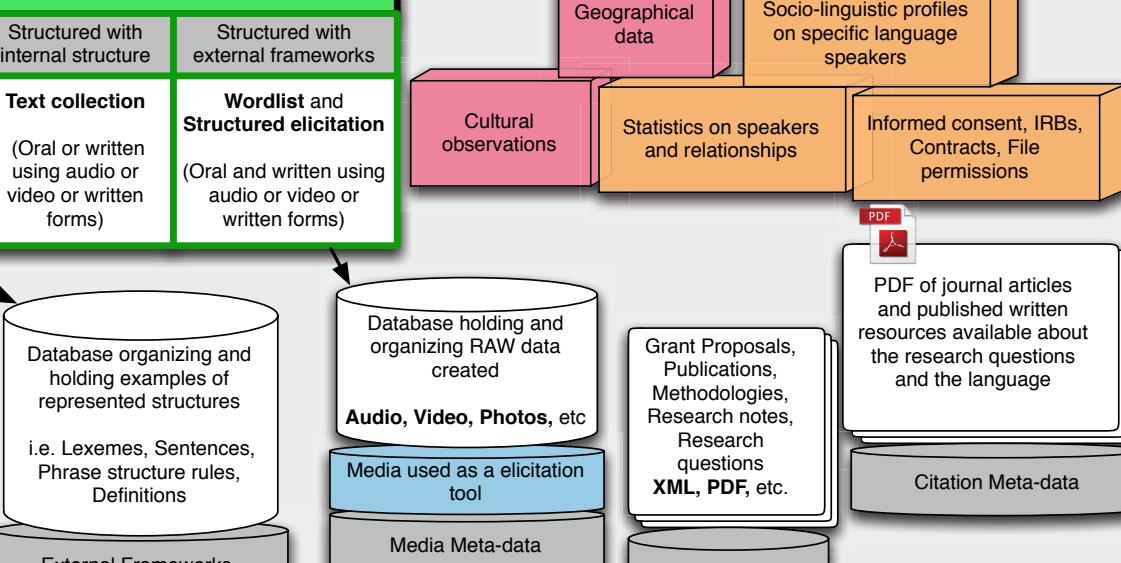
2 Social Attitudes of Linguists

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 attitudes encountered:

- The archiving institution:
 - can be very large institutions
 - can will not restrict content 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 linguist does not fit into the institution's metadata schema.
 - Confusion about the organizational structure of the archiving institution.
 - Opinions that archiving should only take place in the country where the language is indigenous, specifically:
 - 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."



From when do these items come?

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

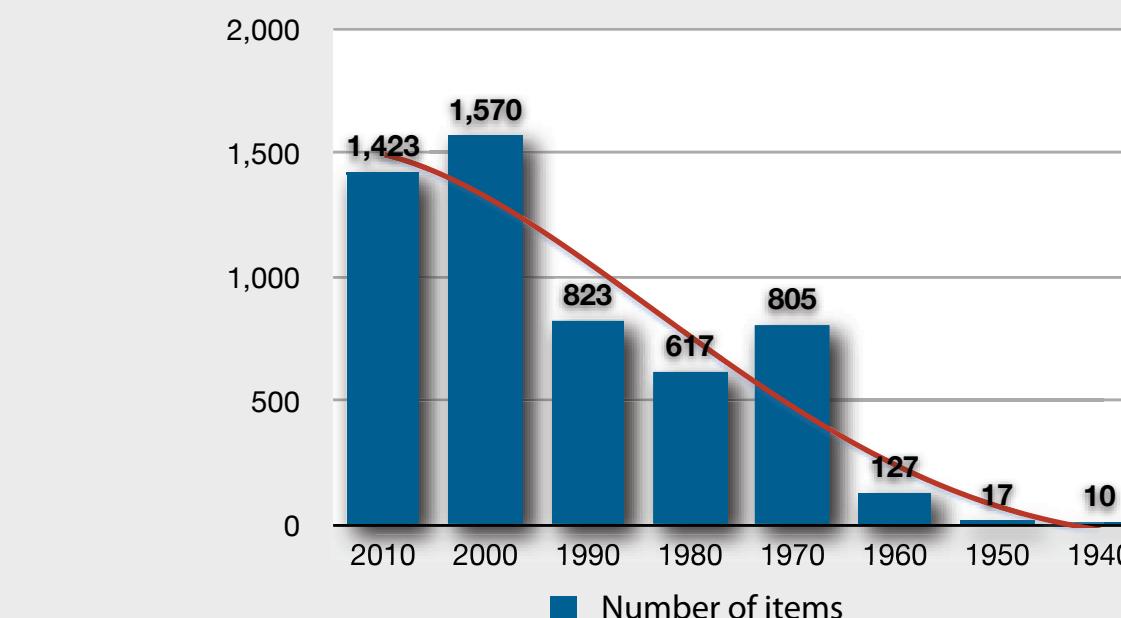
SIL has nearly 80 years of history working with minority language communities.

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?

Archived Items in 2012 with a creation or issue date



3 Task perception

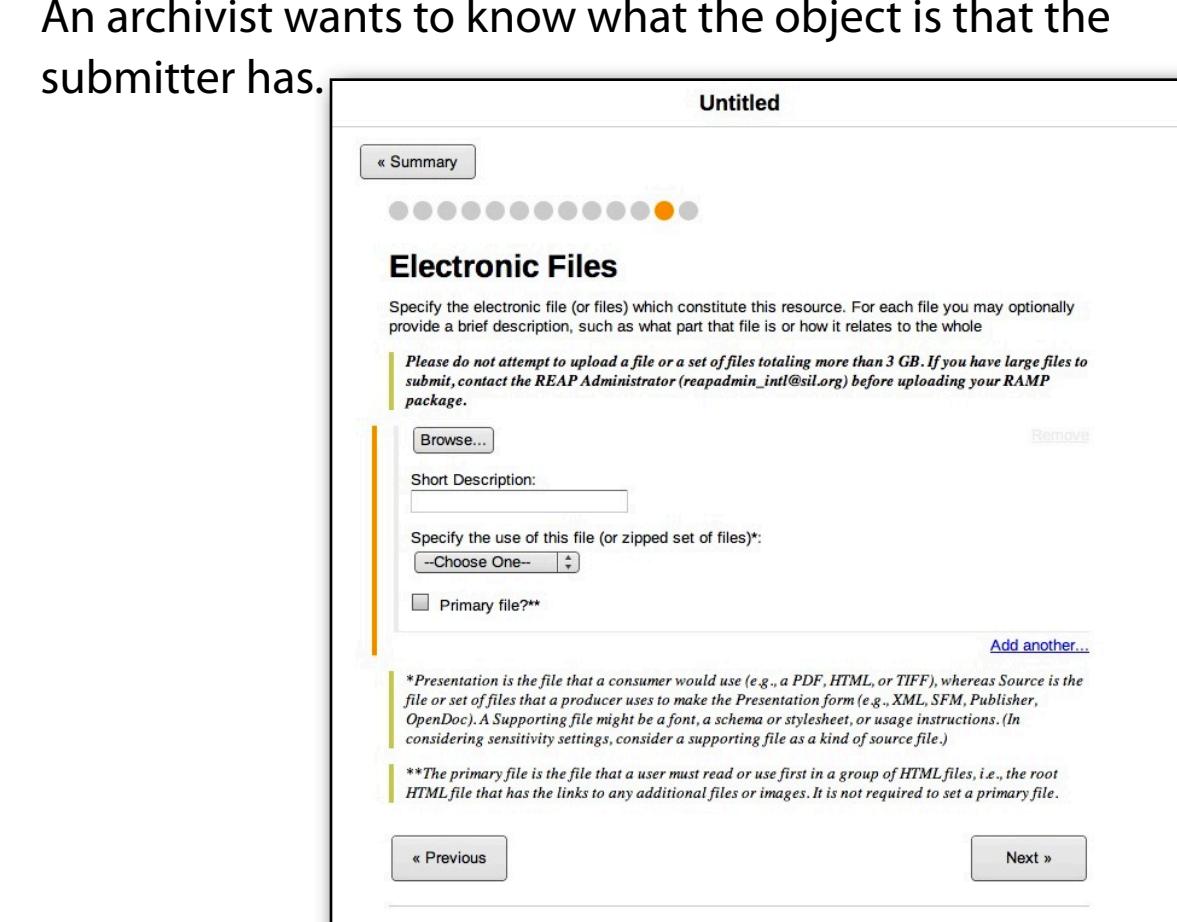
Linguist:

How does RAMP relate to my other data in my workflow?

How does RAMP enable me to keep the promises I made for myself?

Task perception becomes a major issue in user interface design. The user interface also has a major role in setting the mood for the entire interaction.

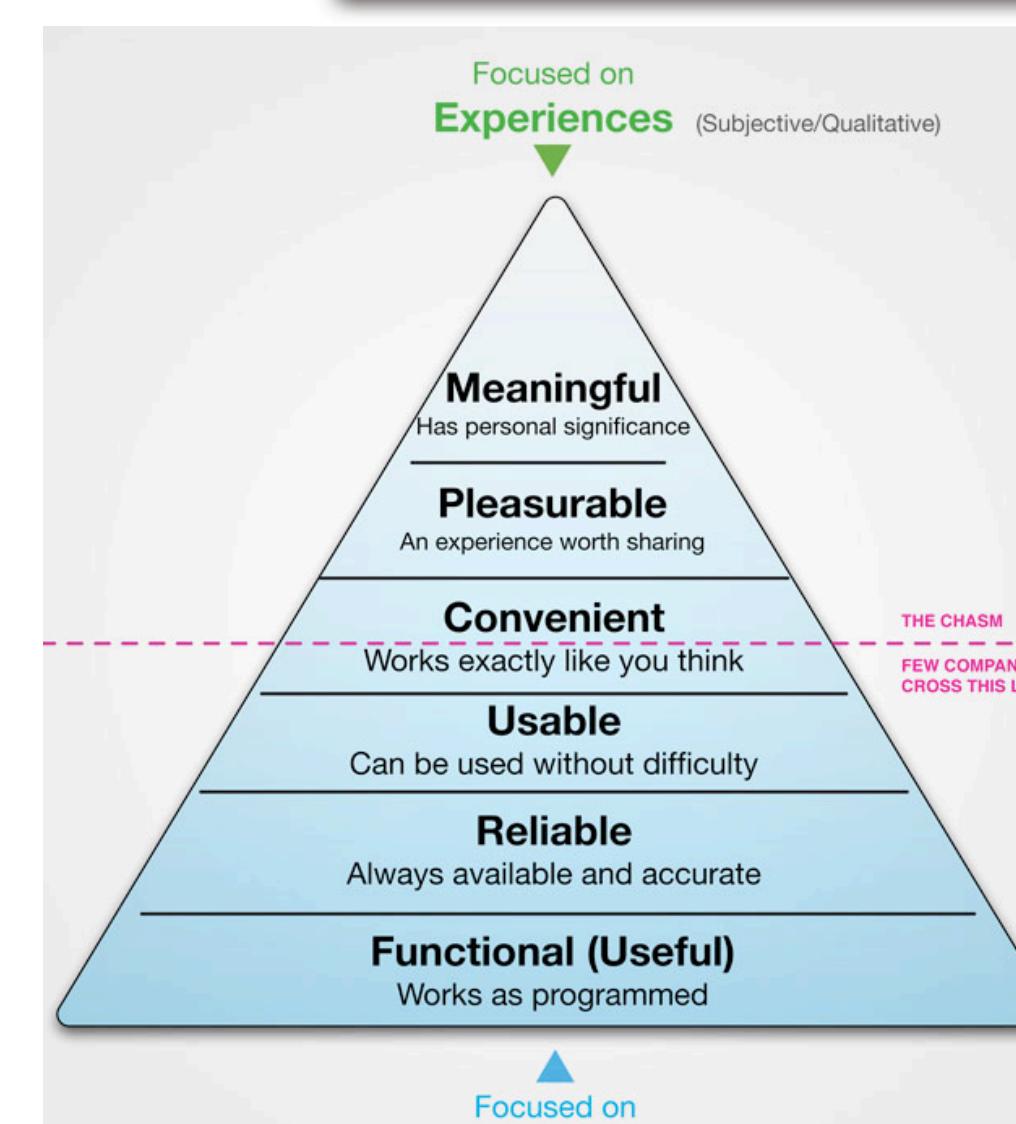
An archivist wants to know what the object is that the submitter has.



Archivist:

How does RAMP tell me what the item is so that I know which "shell" to put it on?

The perception of where RAMP is situated in the entire eco-system is foundational to widespread user adoption.



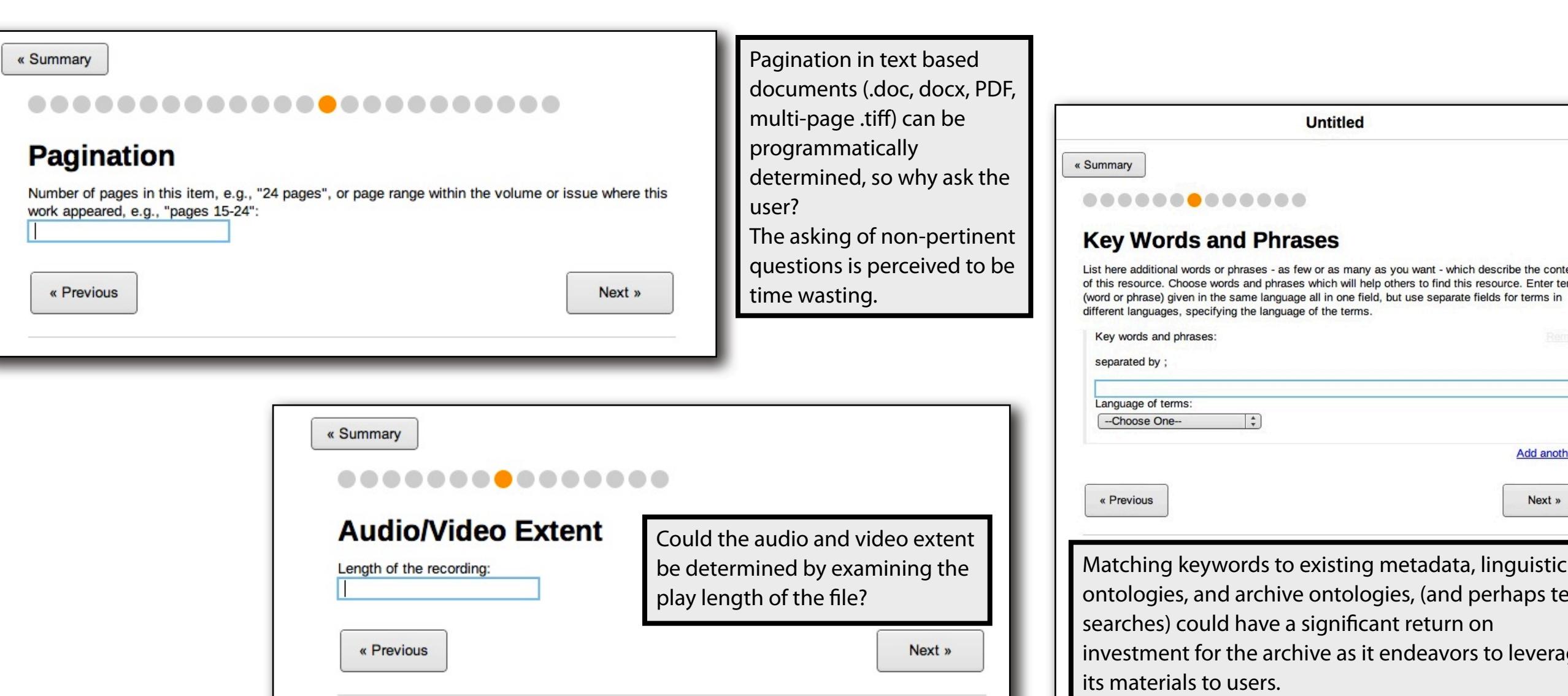
"Emotion and cognition conjointly and equally contribute to the control of thought and behavior." (Gray 2002) Often the design of linguistics based software is focused on specific tasks, not creating meaningful experiences.

Does RAMP cross the chasm and become meaningful to its users? If it did, would we expect to see lateral spread (peer to peer) in the user group, rather than organizational tree based spread?

What is the emotional impact on the RAMP user's attitudes towards archiving? Is it the meta-schema which is too difficult or is it the relationship through the software? - To the RAMP user, is the experience worth repeating and telling their friends about?

Working with existing metadata

Most media files have metadata of some kind embedded in them. Working with this metadata (and allowing the user to verify it) rather than completely ignoring it would save the user effort, and in some cases time. The perceived gain would make the user appreciate the software more. They feel like they have to answer fewer useless or obviously answered questions. But these questions are and can be pertinent in accessioning of non-digital objects.



What kinds of digital objects are being submitted?

Type of Items Submitted in 2012	Total Number of Items across all submitters
Textual based objects (presentations, papers, PDFs)	5,598
Image based objects (.psd, .jpg, raw, tiff)	2,104
Unknown (obscure) object formats, fonts, (ISO, zip)	1,484
Audio based objects (.mp3, .aiff, wav)	1,003
Text-based objects (toolbox files, FLEX, xls)	67
Web formatted files (.html, css)	45
Video based objects (.mov, .vob, .mp4)	32
Total Digital Objects	10,333
60% of images may be part of text based scanning of old documents to archival formats.	

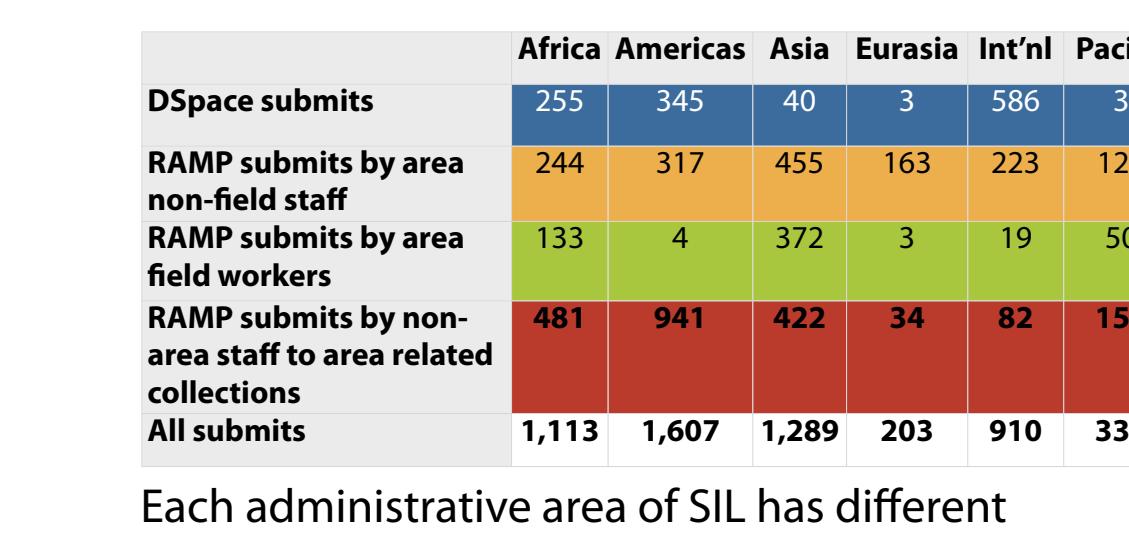
In 2012 there were 475 active Felix clients in the FLEX Google Group yet there were only 4 instances of the FLEX data set archive. These were submitted by 2 contributors and one instance was a version of a previous instance. (Not all FLEX users are in the Google Group, nor are all 475 members Sil staff.)

Another way to assess clumping and dividing is through relationships like X has part Y or Y is a part of X.

1,821 of the items added in 2012 have relationships to other items in the archive.

4 Management Strategies

Submission methods used



Each administrative area of SIL has different strategies for archiving content. These management strategies affect which tools are presented to various sets of linguists and therefore also who does the work related to submission to the archive.

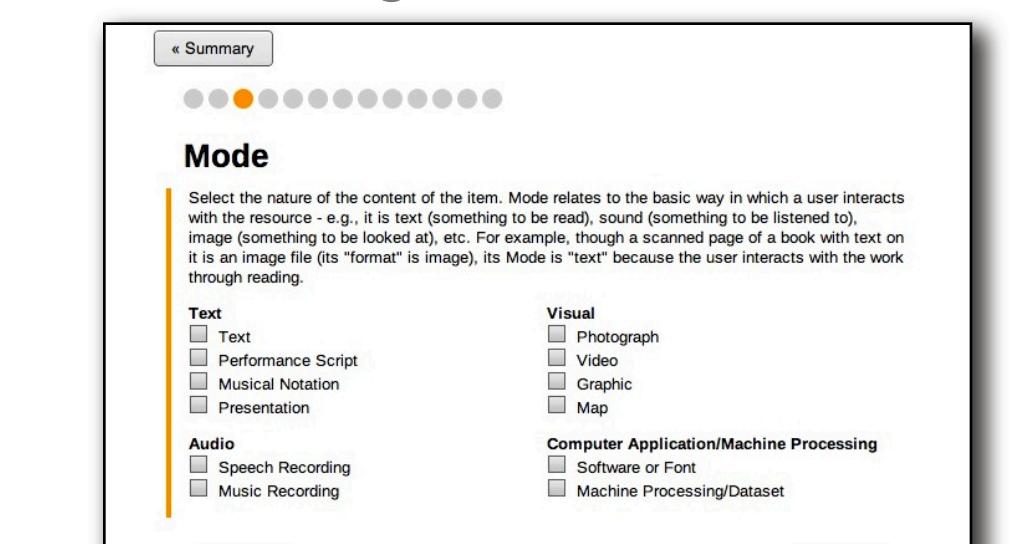
Roughly speaking, higher rates of RAMP usage mean that the area staff is more self-sufficient in terms of submission to the archive. Compare Yellow/Green to Red/Blue.

Though Americas Area leads SIL in total contributions to the archive, it also leads the areas in not encouraging its field workers to submit content directly to the archive via RAMP.

The tool or the fit

Fit is the way that a tool interacts with the entire eco-system. Fit is not just the relationship of the tool to any other part of the eco-system, but also it is how the entire system breaths together to create needs and solutions for users. Pressures or benefits in one part of the system can drive users to use the tool less or more.

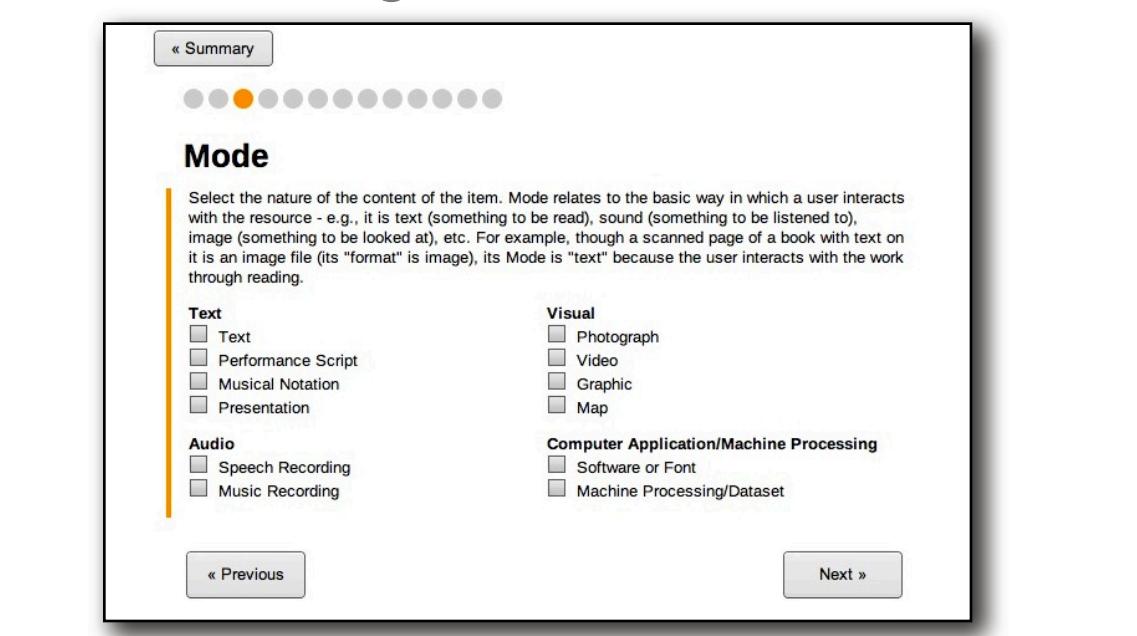
Inappropriate feel in the User Interface or insufficient detail to features can lead to bad report for a tool. Insufficient detail to the overall economy of the larger eco-system can lead to abandonment of the tool, even if it is well designed.



The tool is simply a component of the eco-system designed to involve users in a particular experience.

In user experience analysis we must be careful to not attribute faults of the tool to the fit, and vice versa.

Because of the organizational economics of archiving, there is relatively little return on investment for linguists to archive.



Archives, a dispensable service

The core business of any archive is the marketing of its relevance, often via content promotion and curation services, to both submitters and content users. The more it can convince each group of its value, the more valued it becomes in the eco-system.

Unlike most two sided markets (Parker &