

Archives & Data Ecology

Hugh Paterson III

Some Concepts

- Language Artifacts
- Data Ecology
- Archives
- Lexical Databases



Language Artifacts

Language Artifacts



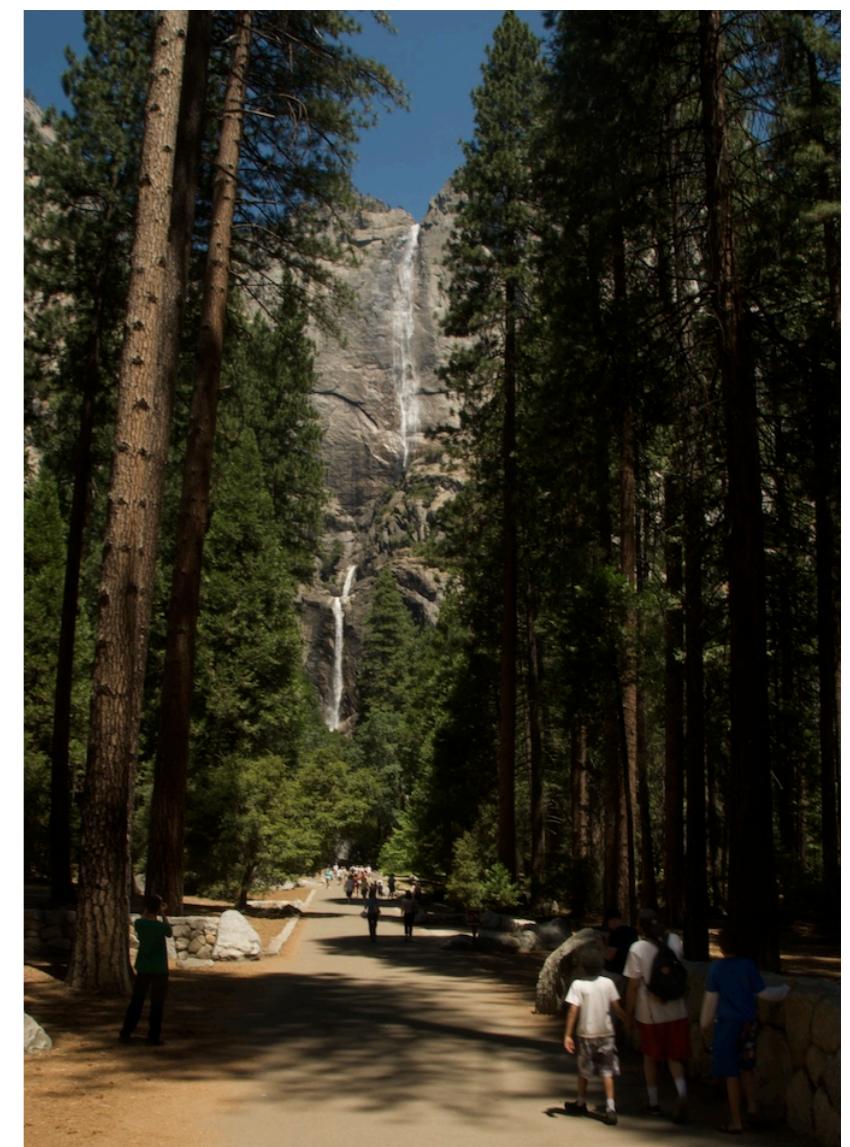
Language Artifacts



Language Artifacts

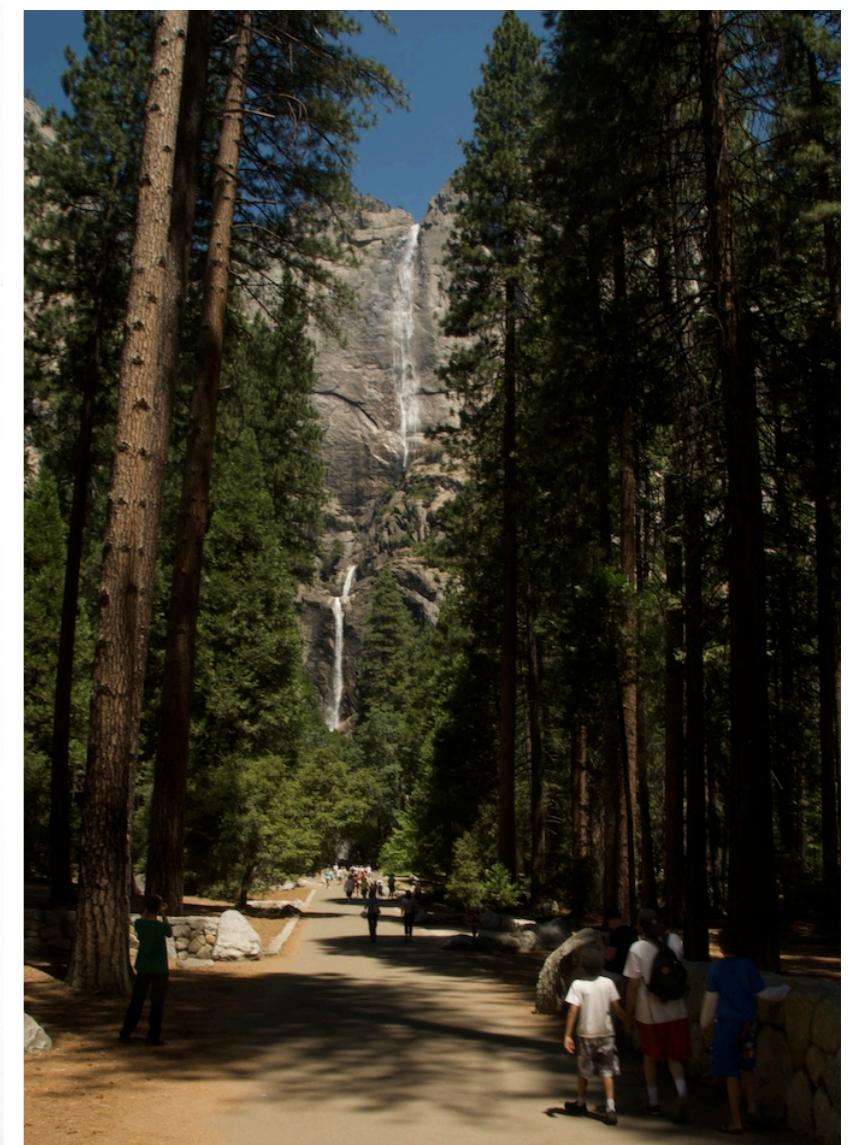
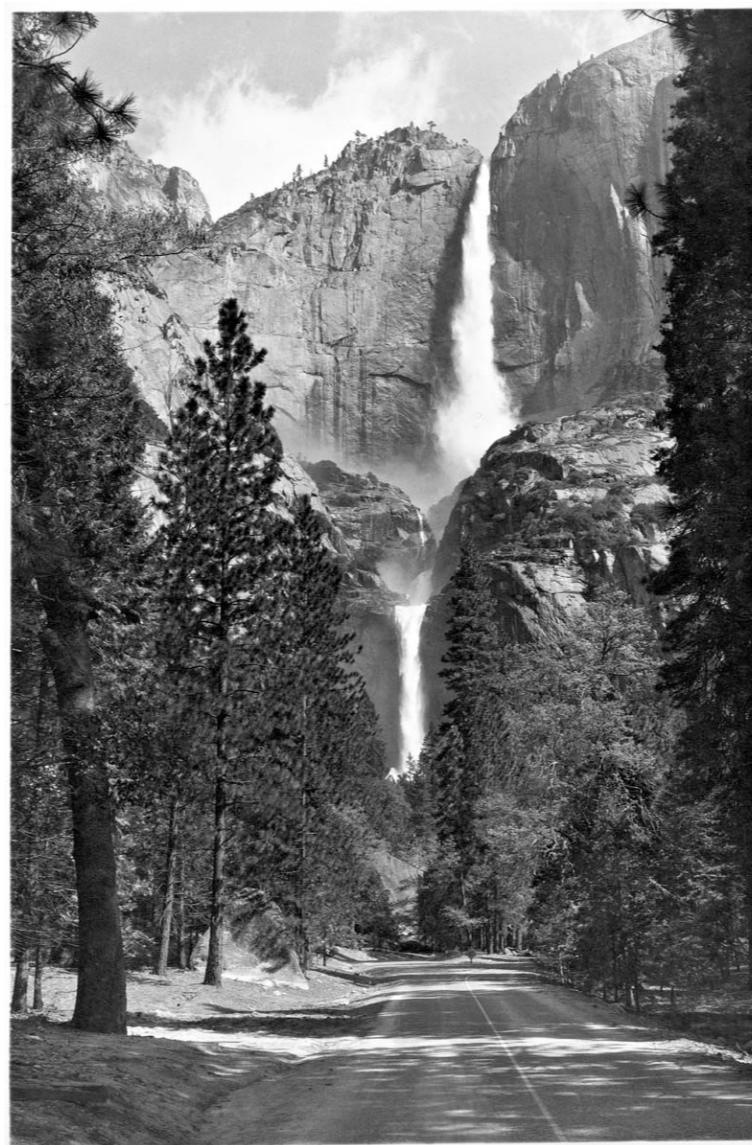
A photograph is usually
looked at — seldom
looked into.

-Ansel Adams



Language Artifacts

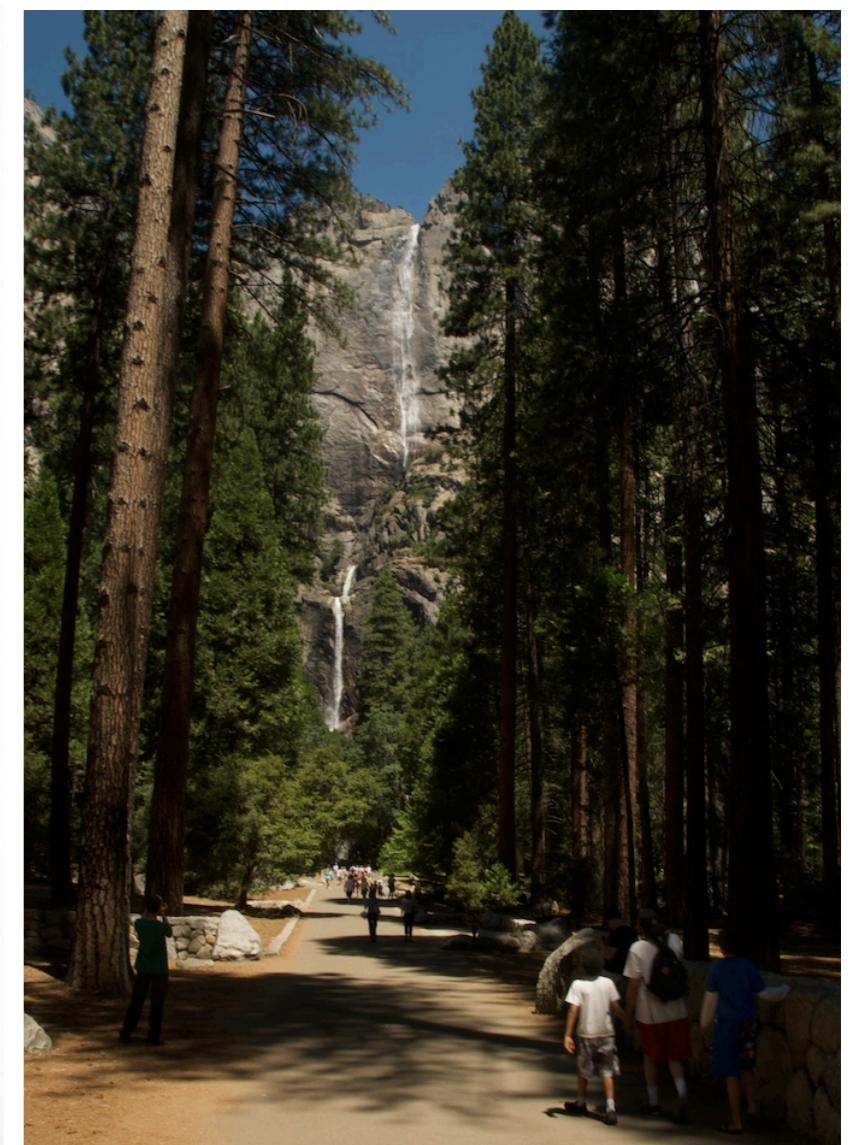
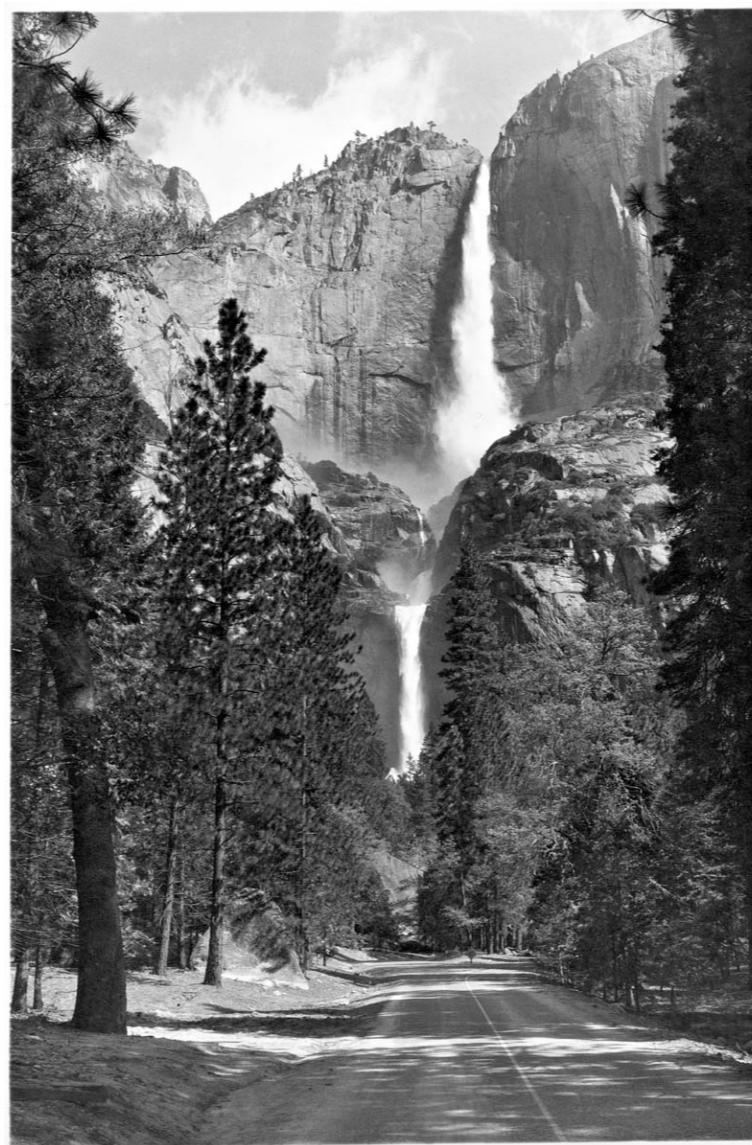
A photograph is usually
looked at — seldom
looked into.
-Ansel Adams



Language Artifacts

A photograph is usually looked at — seldom looked into.
-Ansel Adams

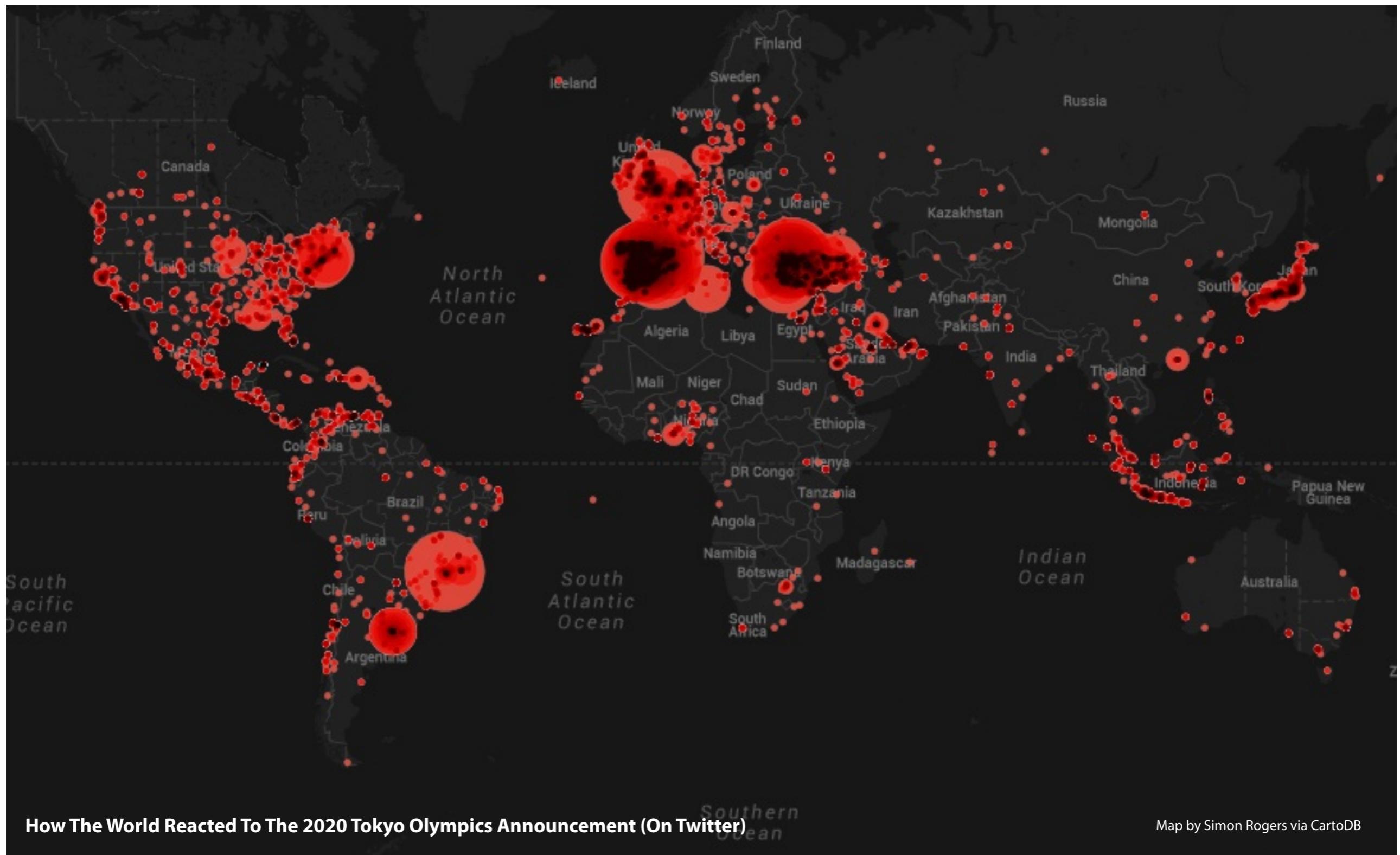
Actually, I'm not all that interested in the subject of photography. Once the picture is in the box, I'm not all that interested in what happens next. Hunters, after all, aren't cooks.
- Henri Cartier-Bresson



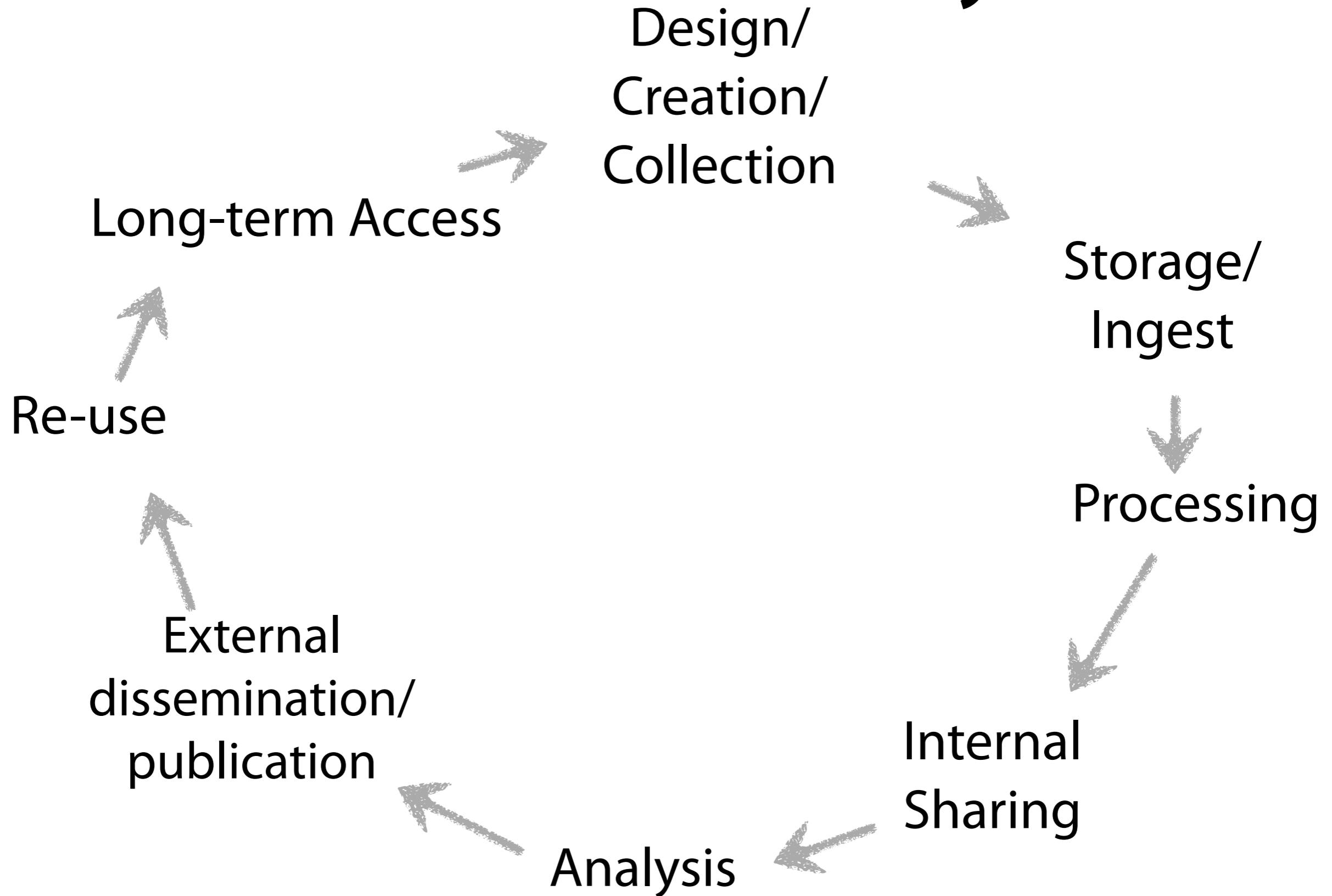
Ecology of Data

- Language artifacts make up a significant portion of the “Linguist’s” data
- Contrast with information ecology
- Life cycle
- Partners

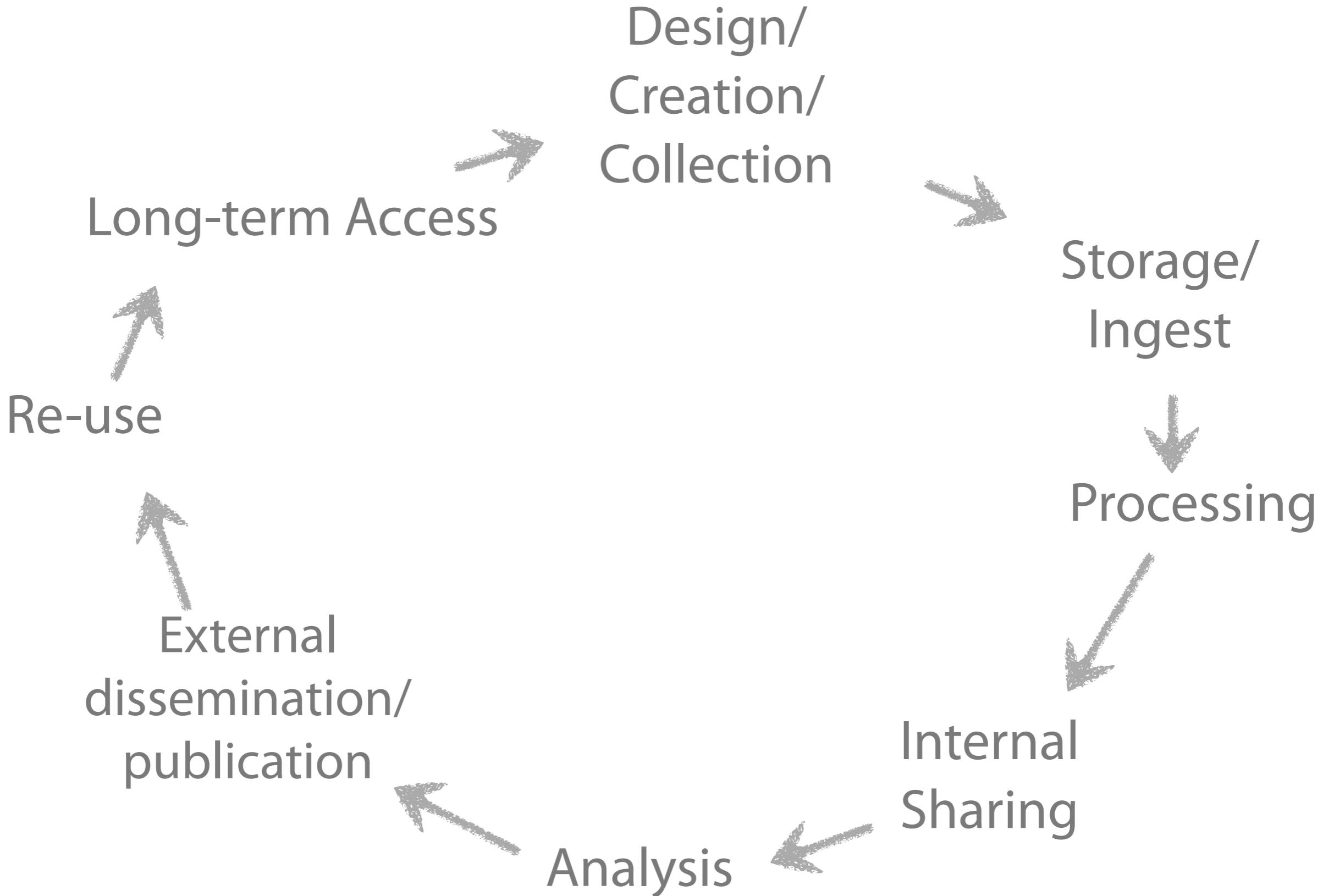
Information Ecology



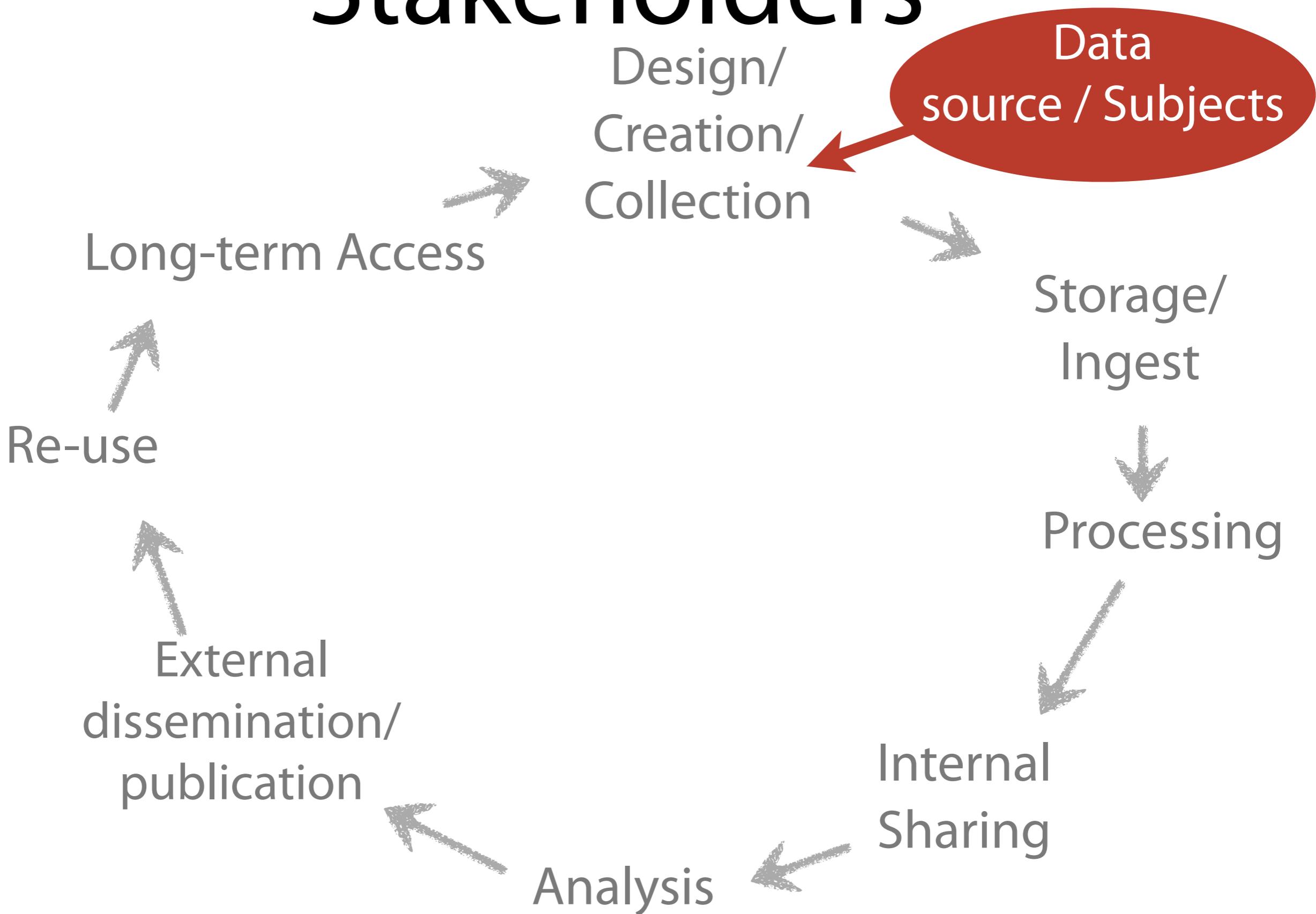
Information Lifecycle



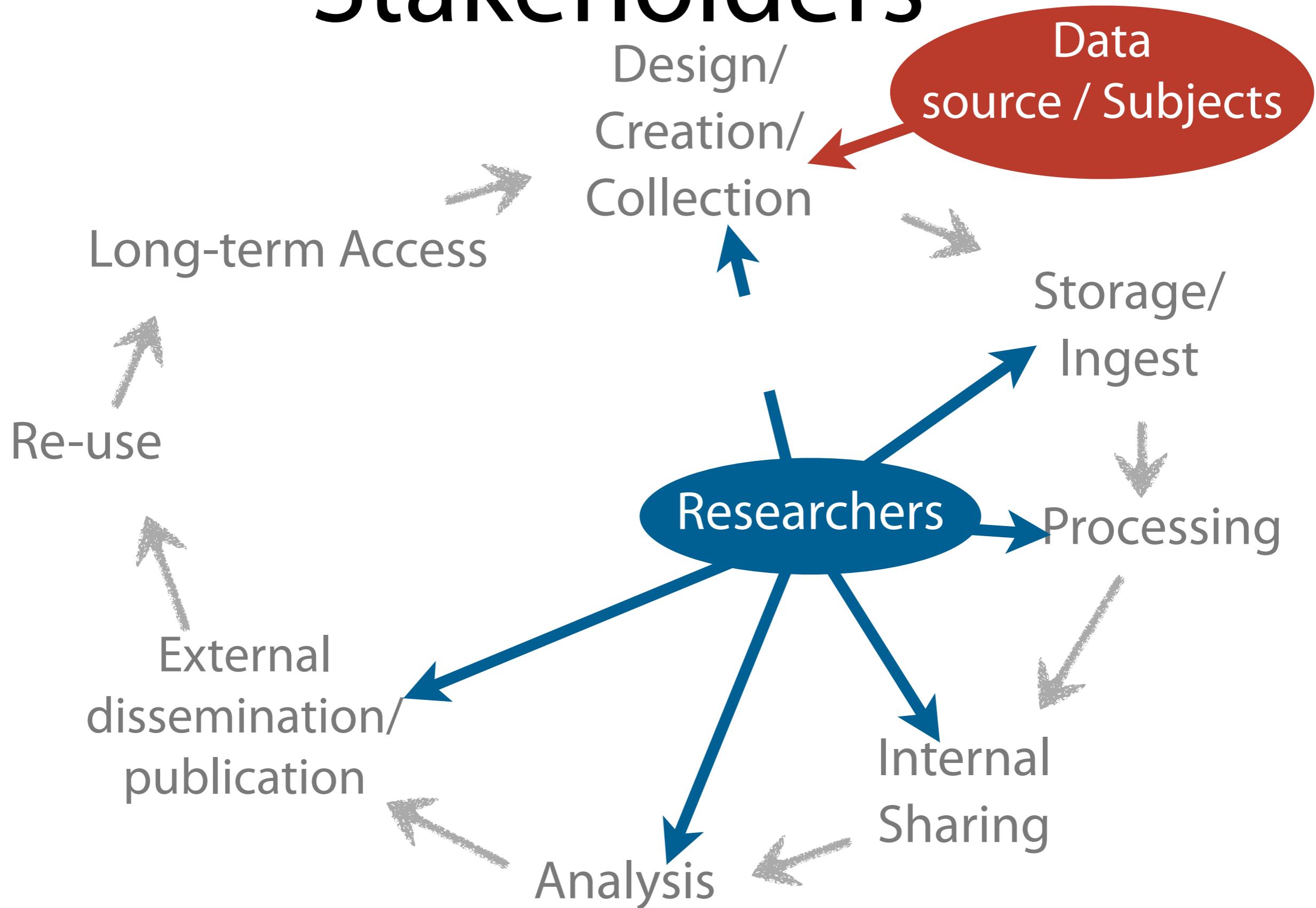
Stakeholders



Stakeholders



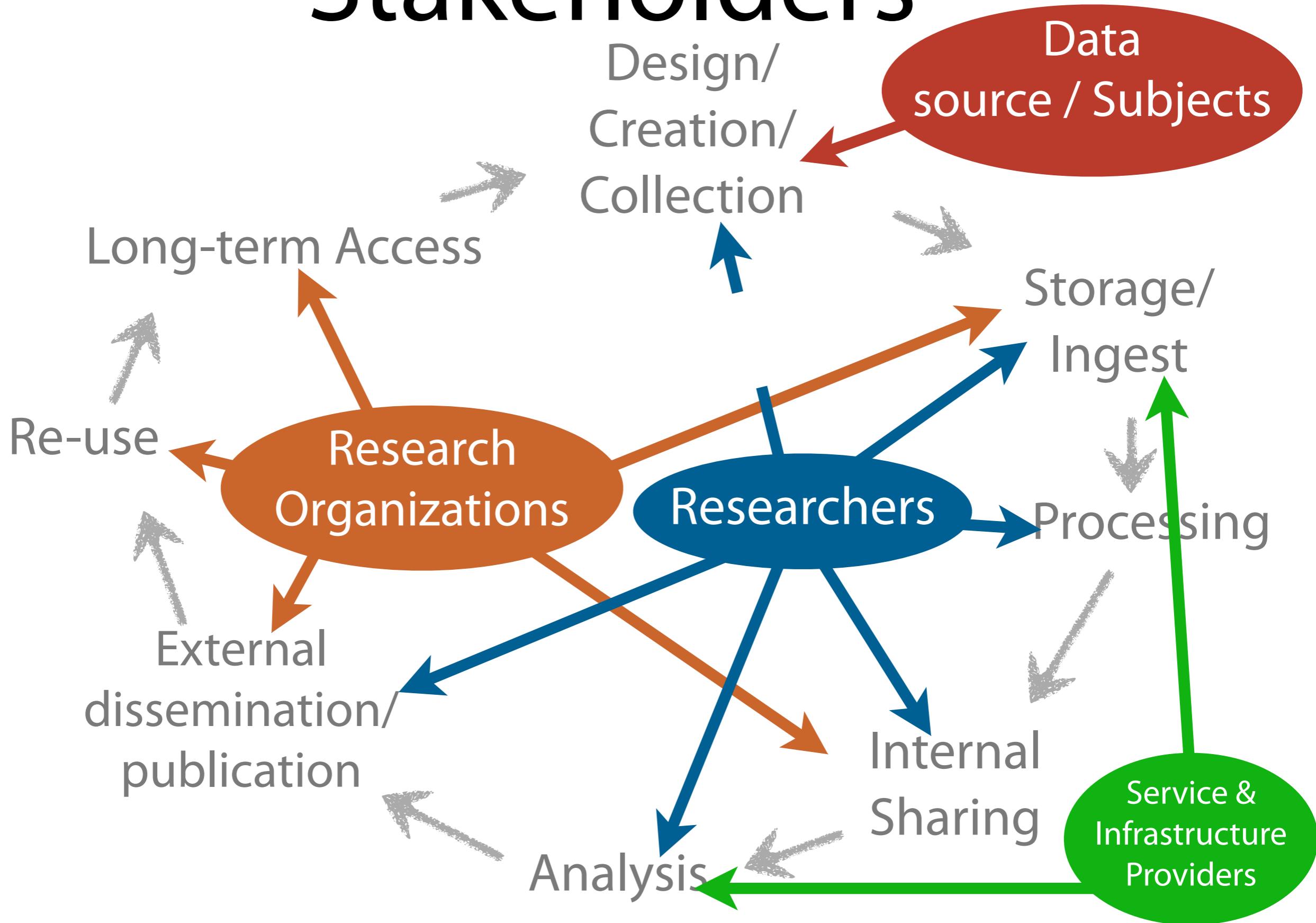
Stakeholders



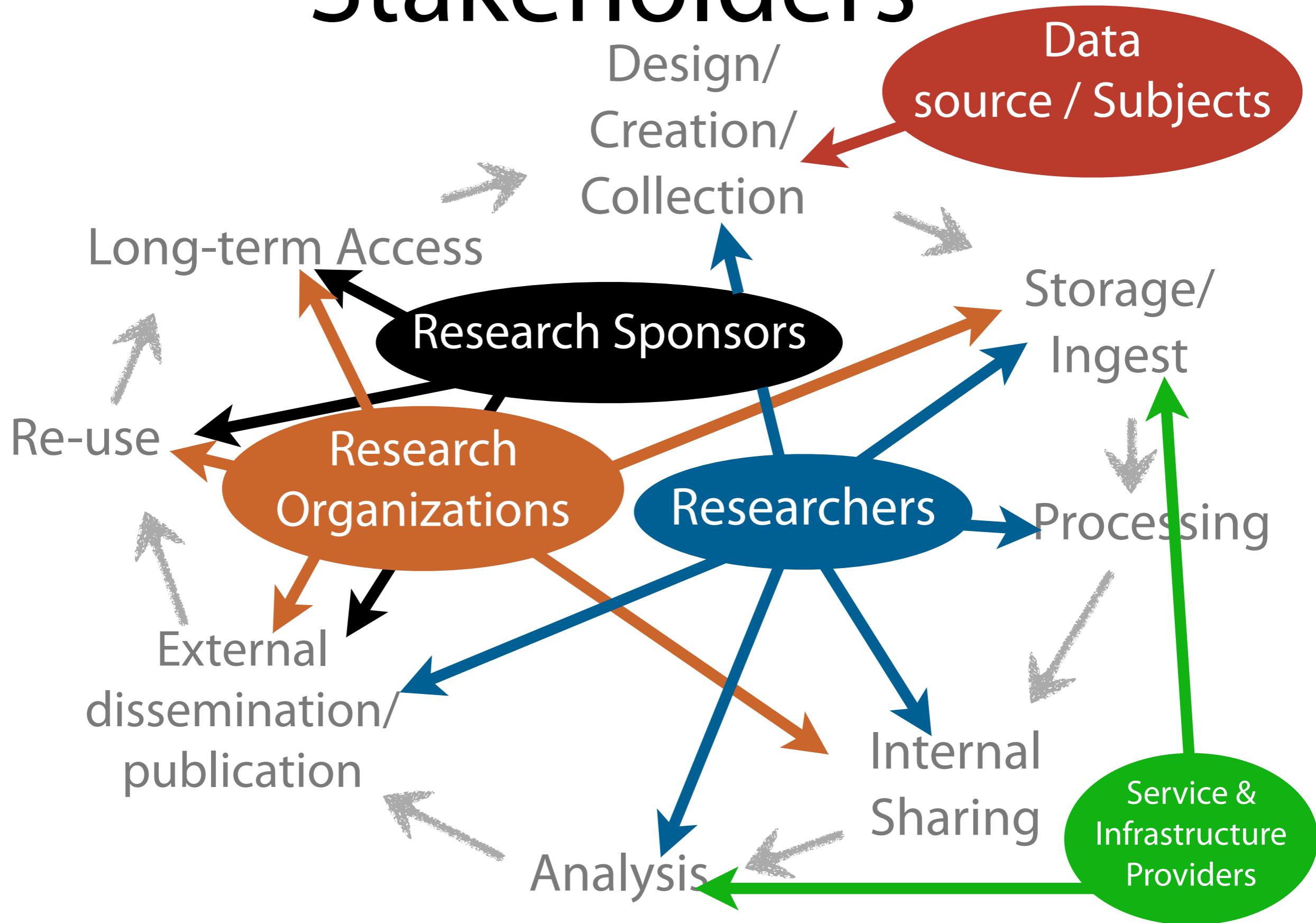
Stakeholders



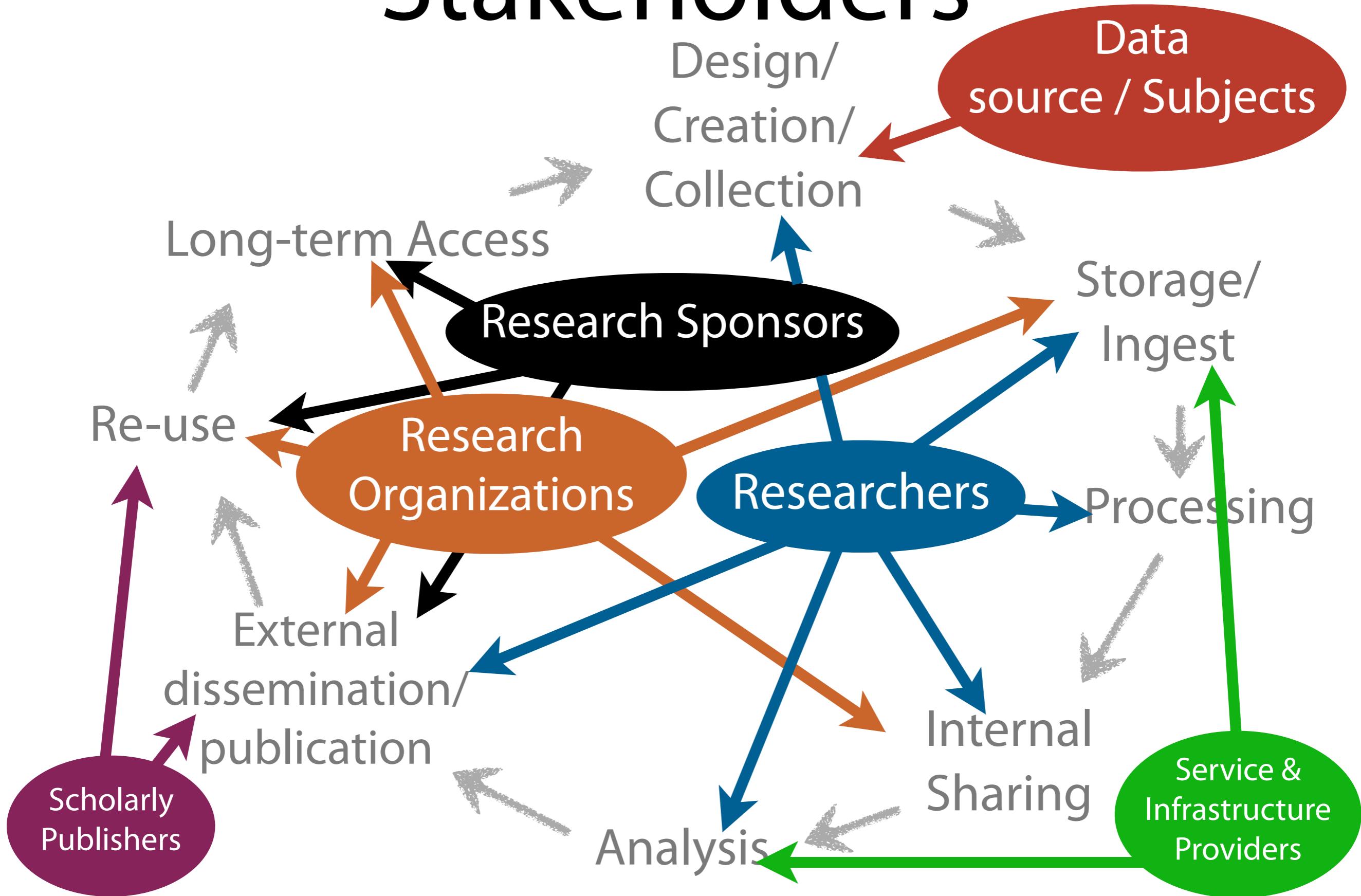
Stakeholders



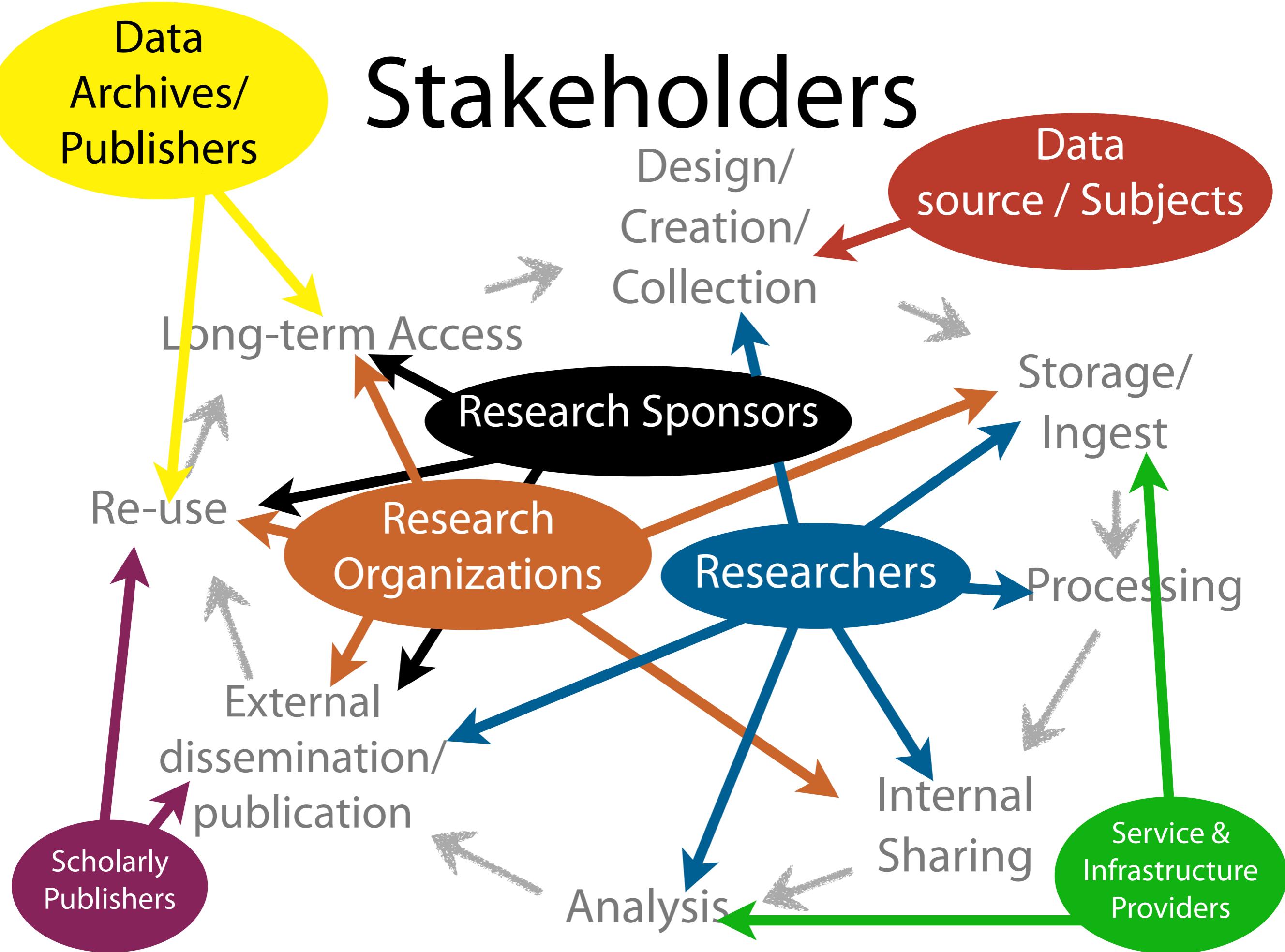
Stakeholders



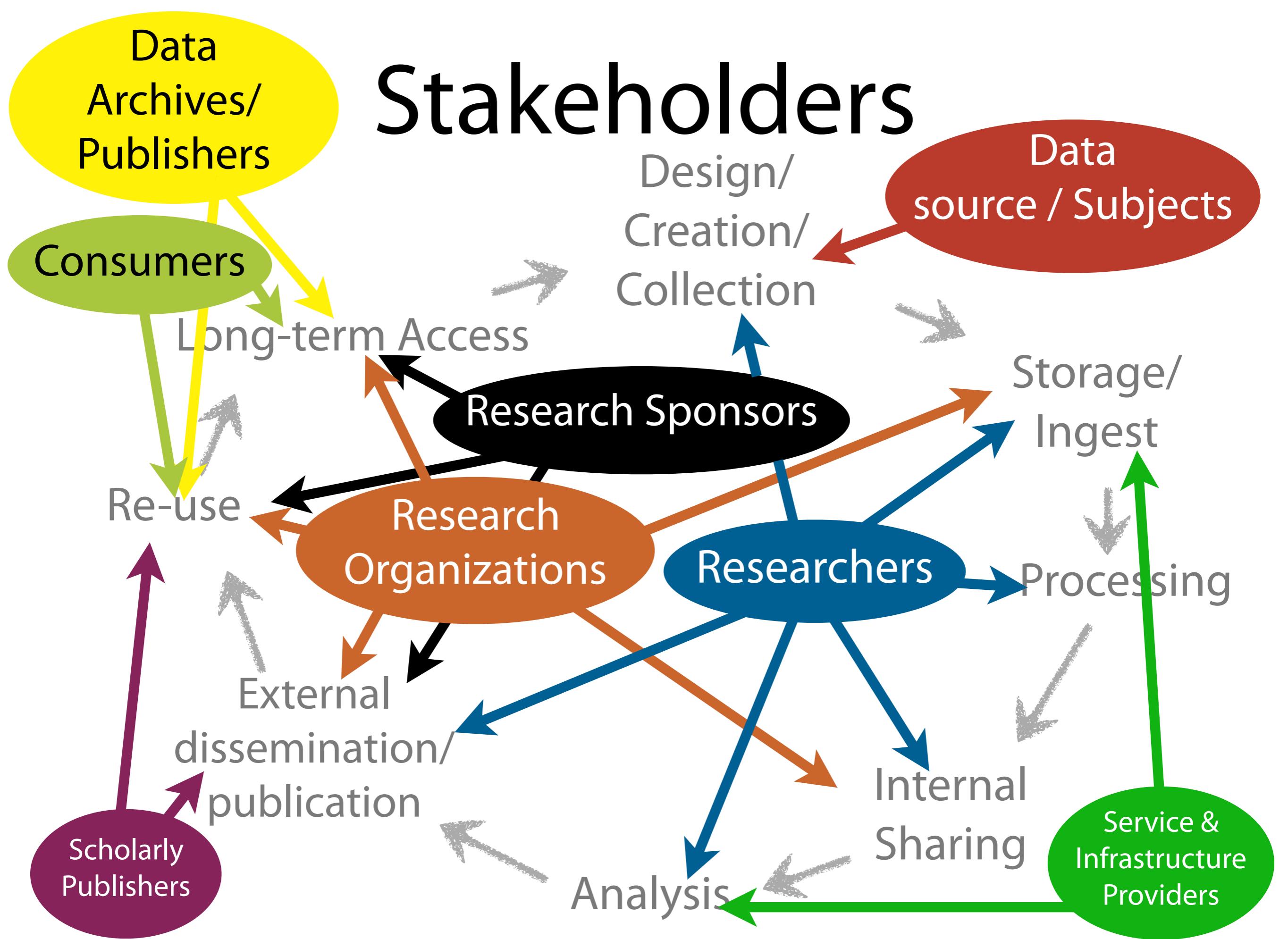
Stakeholders



Stakeholders



Stakeholders



Data Ecology

 **Jorge Emilio Rosés Labrada**
May 13 

Note to self: if you know that someone worked on the language you want to work on before you, take the time to send them an email. It always pays off 😊
Case in point: Someone just shared with me 456 short Sáliba recordings (each with one Sáliba word) that were collected in February of 1996 by Nancy Morse (SIL)!!!

Like · Comment · Share

Becky Smith Paterson, Zoe Tribur, Amos Teo and 21 others like this.

 **Patrick Hall** Neat. Have you noticed any sound changes since then?
May 13 at 11:02am · Like · 1

 **Jorge Emilio Rosés Labrada** I haven't had the chance to work with Sáliba speakers yet, Patrick. Perhaps this summer 😊
May 13 at 11:14am · Like · 2

 **Patrick Hall** Doh
May 13 at 11:18am · Like

 **Andrea L. Berez** Very true!
May 13 at 12:08pm · Like · 1

What is an Archive

- Organized preservation
- The attic
- A collection



TAPS

Target	Access
Preservation	Sustainability

TAPS

Target	Access
Mission Statement Submission Criteria Designated Communities Ongoing Relationship	
Preservation	Sustainability

TAPS

Target	Access
Mission Statement Submission Criteria Designated Communities Ongoing Relationship	Discoverability Reach
Preservation	Sustainability

TAPS

Target	Access
Mission Statement Submission Criteria Designated Communities Ongoing Relationship	Discoverability Reach
Preservation	Sustainability
Evidence of Long-Term Planning Preservation Strategies Integrity Authenticity	

TAPS

Target	Access
Mission Statement Submission Criteria Designated Communities Ongoing Relationship	Discoverability Reach
Preservation	Sustainability
Evidence of Long-Term Planning Preservation Strategies Integrity Authenticity	Adequate Infrastructure Financial Sustainability Disaster Preparedness Succession Plan

Why Archive

- Long tradition
- Resource discovery
- Ethical reasons
- Community access
- Narcissism
- Funding requirements

Discovery - OLAC

Discovery - OLAC



OLAC resources in and about the Oneida language

ISO 639-3: [one](#)

The combined catalog of all OLAC participants contains the following resources that are relevant to this language:

- [Language descriptions](#)
- [Other resources about the language](#)

Use faceted search to [explore resources for Oneida language](#).

Language descriptions

1. [ONLINE](#) *Glottolog 2.2 Resources for Oneida*. n.a. 2013. Max Planck Institute for Evolutionary Anthropology. [oai:glottolog.org:onei1249](#)
2. [ONLINE](#) *WALS Online Resources for Oneida*. n.a. 2008. Max Planck Institute for Evolutionary Anthropology. [oai:wals.info:ond](#)

Other resources about the language

1. [ONLINE](#) *A comparative study of Lake-Iroquoian Accent*. Michelson, Karen. 1988. Kluwer Academic Publishers. [oai:refdb.wals.info:568](#)
2. [ONLINE](#) *Oneida Verb Morphology*. Lounsbury, Floyd. 1953. Department of Anthropology, Yale University. [oai:refdb.wals.info:1986](#)
3. [ONLINE](#) *Two feminine genders in Oneida*. Abbott, Clifford. 1984. WALS Online RefDB. [oai:refdb.wals.info:3073](#)
4. [ONLINE](#) *The use of demonstratives in Oneida narratives*. Diessel, Holger. 1999. WALS Online RefDB. [oai:refdb.wals.info:3542](#)
5. [ONLINE](#) *Tense and Aspect Systems*. Dahl, Östen. 1985. Blackwell. [oai:refdb.wals.info:4272](#)
6. [ONLINE](#) *Evidential Diachrony in Northern Iroquoian*. Mithun, Marianne. 1986. Evidentiality: The Linguistic Coding of Epistemology. [oai:refdb.wals.info:4662](#)
7. [ONLINE](#) *An Oneida Dictionary*. Christjohn, Amos; Hinton, Maria. 1996. Wisconsin Department of Public Instruction. [oai:refdb.wals.info:5295](#)
8. *Testing procedures for estimating transfer of information among Iroquois dialects and languages*. Hickerson, Harold; Hickerson, Nancy P.; Turner, Glen D. 1962. International Journal of American Linguistics. [oai:sil.org:4833](#)
9. [ONLINE](#) *Oneida: a language of Canada*. n.a. 2013. SIL International. [oai:ethnologue.com:one](#)
10. [ONLINE](#) *The Languages of Native North America*. Mithun, Marianne. 1999. Cambridge: Cambridge University Press. [oai:rosettaproject.org:rosettaproject_one_detail-1](#)
11. *Oneida verb morphology*. Lounsbury, Floyd G. 1953. Yale University publications in anthropology ; no. 48. [oai:gial.edu:20234](#)
12. [ONLINE](#) *LINGUIST List Resources for Oneida*. Anthony Aristar, Director of Linguist List (editor); Helen Aristar-Dry, Director of Linguist List (editor). 2014-05-22. The LINGUIST List ([www.linguistlist.org](#)). [oai:linguistlist.org:lang_one](#)

Other search terms: dialect, vernacular, grammar, syntax, morphology, phonology, orthography

Discovery - OLAC

UWDC
University of Wisconsin Digital Collections

VISIT A COLLECTION SEARCH THE COLLECTIONS
Select a collection search... 

HOME COLLECTIONS NEWS LEARNING RESOURCES ABOUT US CONTACT US

Home > Collections > The Wisconsin Oneida Language Preservation Project > Collection Home

The Wisconsin Oneida Language Preservation Project

Search the collection  About the collection
Guided search Copyright

The Wisconsin Oneida Language Preservation Project presents original stories and curriculum, with accompanying audio, that teaches the Wisconsin Oneida language. Materials presented here include Kindergarten Curriculum, Songs, and Stories in both English and written Oneida language. Audio files accompany these texts and provide additional instruction, allowing language learners to read and simultaneously listen to native Oneida speakers present the same material.

Materials in the Collection

Songs
K-6 Curriculum
Kindergarten Curriculum
First Grade Curriculum
Second Grade Curriculum
Third Grade Curriculum
Fourth Grade Curriculum
Fifth Grade Curriculum
Sixth Grade Curriculum
WPA Stories

To access or cite this collection:
<http://digital.library.wisc.edu/1711.dl/Oneida>



SOVEREIGN OF WISCONSIN
ONEIDA NATION

Related materials

- The Oneida Nation of Wisconsin

Discovery - OLAC



OLAC resources in and about the Kurti language

ISO 639-3: [ktm](#)

The combined catalog of all OLAC participants contains the following resources that are relevant to this language:

- [Primary texts](#)
- [Lexical resources](#)
- [Language descriptions](#)
- [Other resources about the language](#)

Other known names and dialect names: Kuruti, Kuruti-Pare, Ndrugul

Use faceted search to [explore resources for Kurti language](#).

Primary texts

1. [ONLINE](#) *Genesis 1-15*. The Long Now Foundation. 1999. Kurti Bible Translation Group. oai:rosettaproject.org:rosettaproject_ktm_gen-1

Lexical resources

1. [ONLINE](#) *RB1-005*. Blust, Robert (depositor). n.d. Kaipuleohone. oai:scholarspace.manoa.hawaii.edu:10125/4683

Language descriptions

1. [ONLINE](#) *Glottolog 2.2 Resources for Kurti*. n.a. 2013. Max Planck Institute for Evolutionary Anthropology. oai:glottolog.org:kurt1250
2. [ONLINE](#) *Kurti Organised Phonology Data*. n.a. 2011. SIL Language and Culture Archives. oai:sil.org:42270

Other resources about the language

1. [ONLINE](#) *Kurti: a language of Papua New Guinea*. n.a. 2013. SIL International. oai:ethnologue.com:ktm

Other known names and dialect names: Kuruti, Kuruti-Pare, Ndrugul

Other search terms: dialect, vernacular, discourse, stories, conversation, dialogue, documentation, lexicon, dictionary, vocabulary, wordlist, phrase book, grammar, syntax, morphology, phonology, orthography

Discovery - OLAC

The screenshot shows the top navigation bar of the SIL International website. It features the SIL logo with the number '80' and the years '1934-2014'. The navigation menu includes links for HOME, SHOPPING CART (0 ITEMS), CONTACT US, and SIGN IN. There is also a search bar with the placeholder 'Search SIL International...' and a magnifying glass icon. Below the navigation bar is a purple horizontal bar containing links for Language Development, Language & Culture, Resources, Training, Worldwide, and About SIL.

Kurti Organised Phonology Data

[Facebook Like](#) 0 [Twitter Tweet](#) 0 [G+1](#) 0

Date: 2011
Extent: 5 pages
Description: Manuscript work, see item for date information
Publication Status: Draft (posted 'as is' without peer review)
Country: Papua New Guinea
Subject Languages: Kurti [ktm]
Content Language: English [eng]
Field: Linguistics
Work Type: Manuscript
Work Type: phonological description
Subject: Phonology
Nature of Work: Text
Available Online: <http://www.sil.org/pacific/png/abstract.asp?id=928474542270>
Entry Number: 42270

Language Development Dict & Lexicography Lang Assessment Lang Programs Mgt. Lang Technology Literacy & Education Media Development Translation	Language & Culture Anthropology Arts & Ethnomusicology L & C Documentation L & C Learning Linguistics Sign Languages Sociolinguistics	Resources L & C Archives Publications Software & Fonts	Training About SIL Training For Specialists Programs	Discover SIL Discover SIL Administration Consultants Our History Annual Updates FAQs Financial Information News	SIL International Publications About SIL Publications Return Policy Standing Orders FAQs Contact SIL Publications
		RELATED SITES Ethnologue ScriptSource ISO 639-3	WORLDWIDE On Location Why Languages Matter Endangered Languages	GET INVOLVED	

Discovery - OLAC

Case Study

★ Why I did it.

- I needed to test two hypotheses

- I chose one specific type of language artifact

Case Study

★ Why I did it.

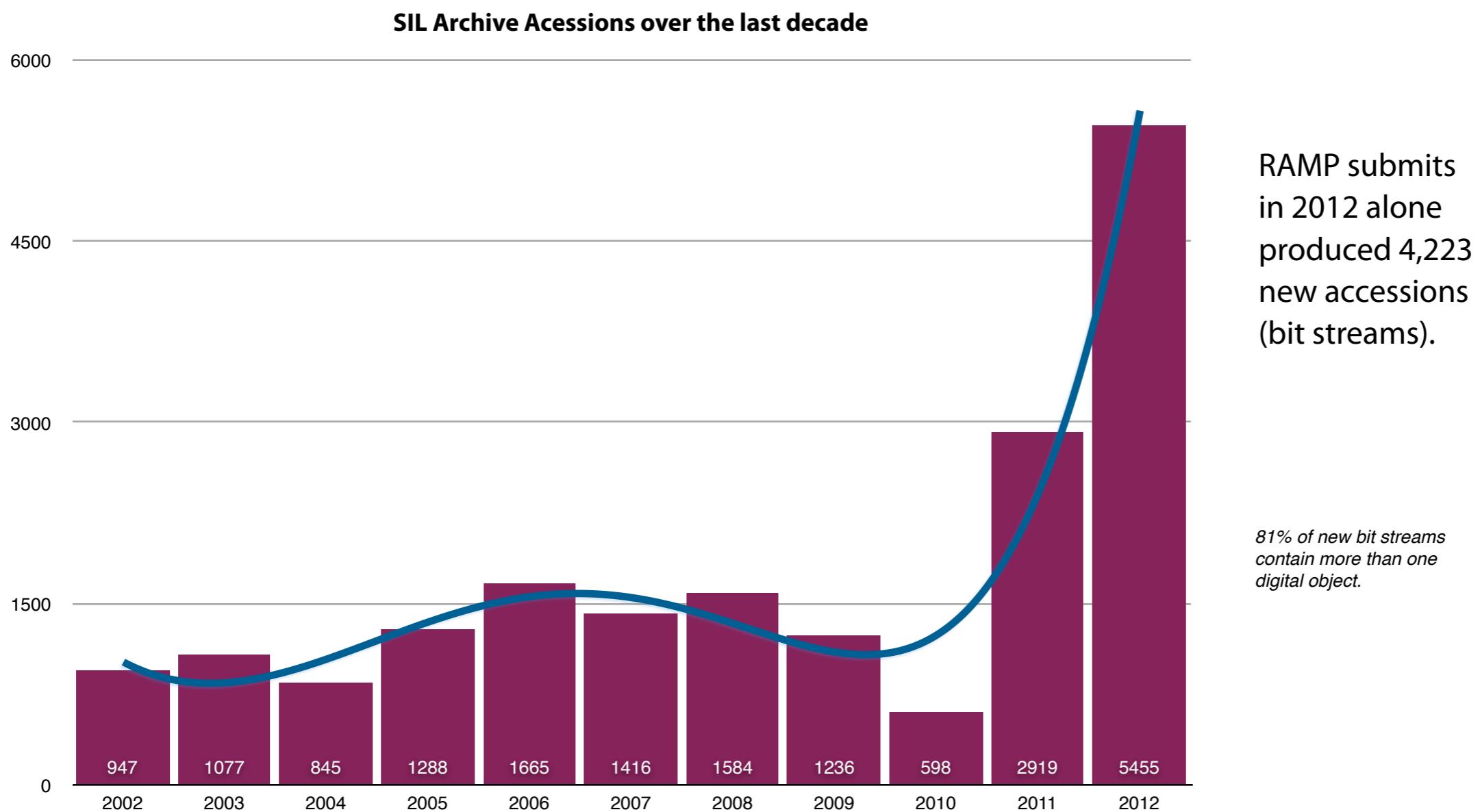
- I needed to test two hypotheses

- I chose one specific type of language artifact

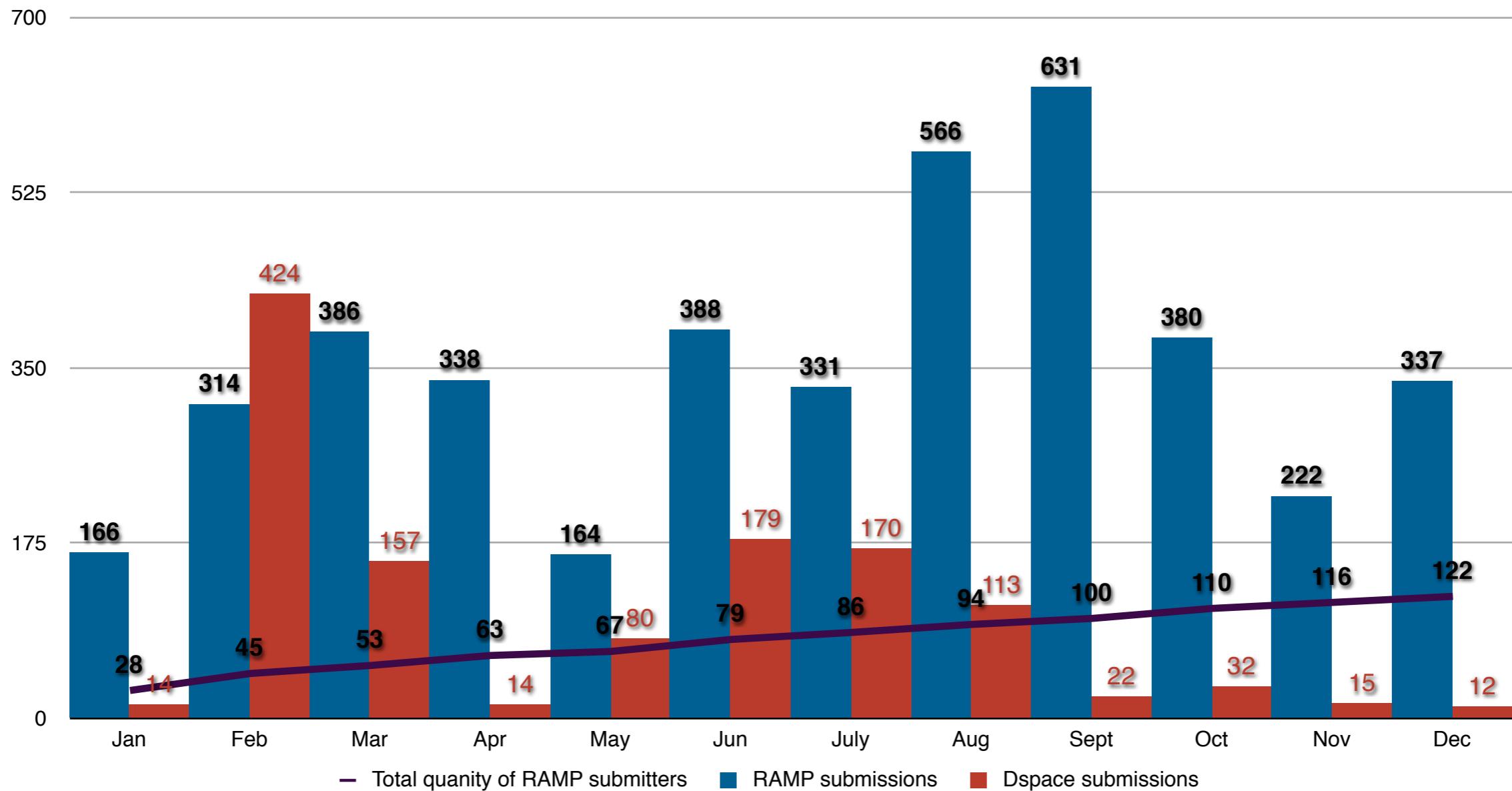
- I. Language artifacts are not being archived
- II. Ratios between SIL archiving habits and non-SIL archiving habits are similar in portion to each other

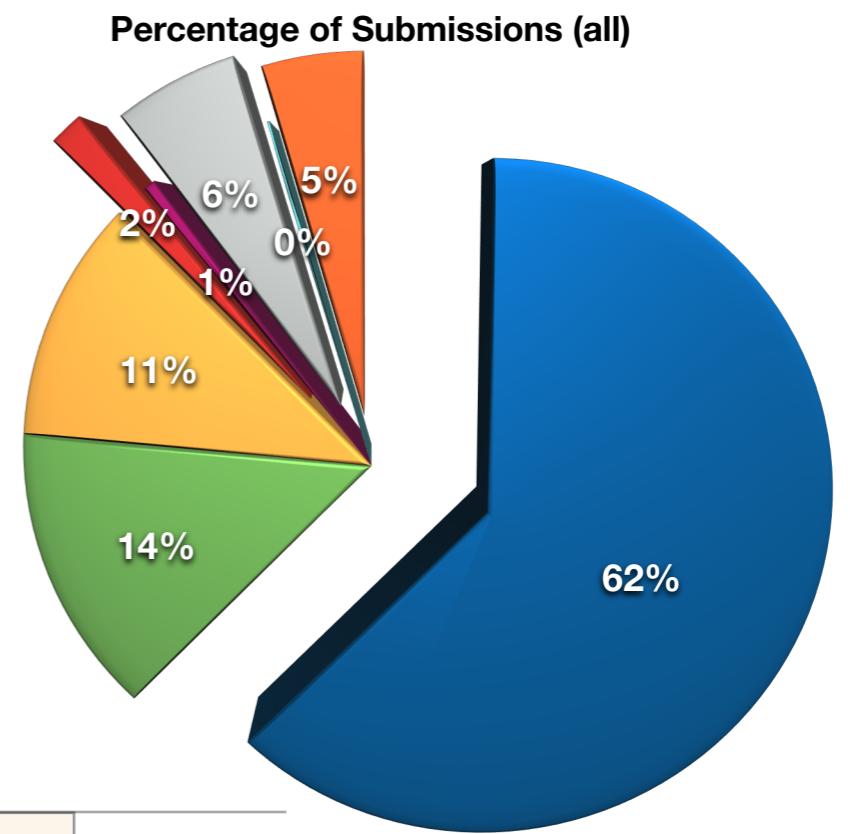
Accession Rates

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

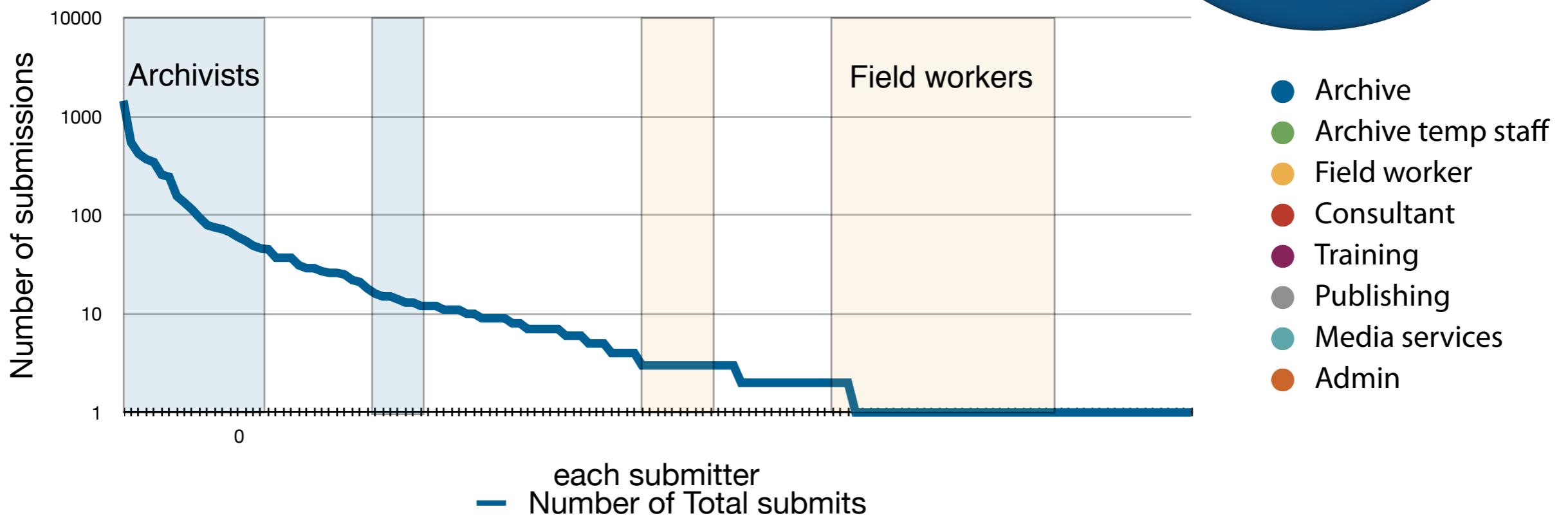


RAMP contributed an increased capacity (200%) to accession materials, but was it equally successful in reaching its intended audience?





Submitters and number of submissions



Adoption among the targeted user group is around 2%.

Previous Claims



Previous Claims

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



Previous Claims

- SIL has nearly 80 years of history working with minority language communities.
- About 1 million relevant non-digital objects are estimated to exists in SIL networks.

Previous Claims

- 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.

Case Study

★ Why I did it.

- I needed to test two hypotheses

- I chose one specific type of language artifact

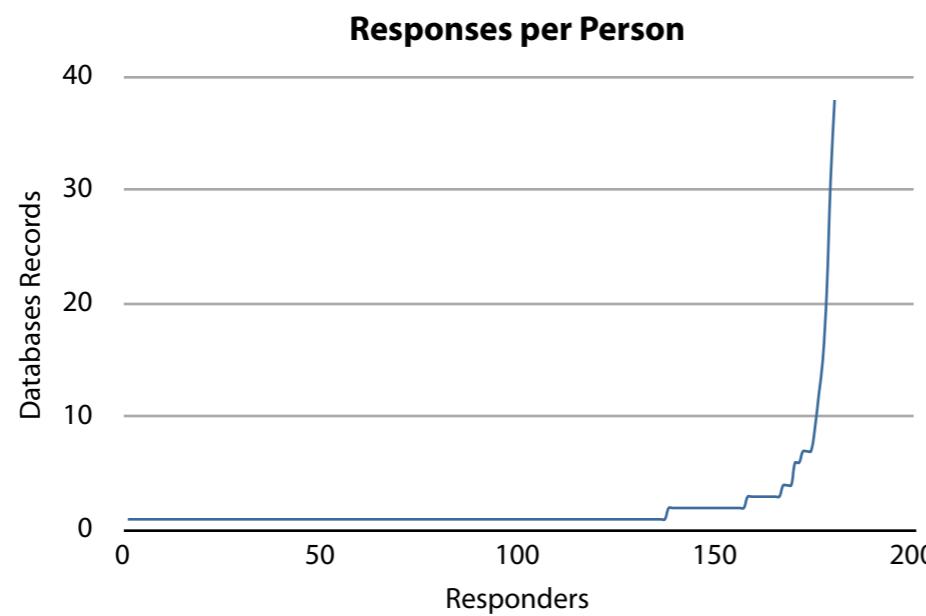
- I. Language artifacts are not being archived
- II. Ratios between SIL archiving habits and non-SIL archiving habits are similar in portion to each other

Responses

About 311 languages are represented in the questionnaire results

The questionnaire has been open for responses since

6 November 2013

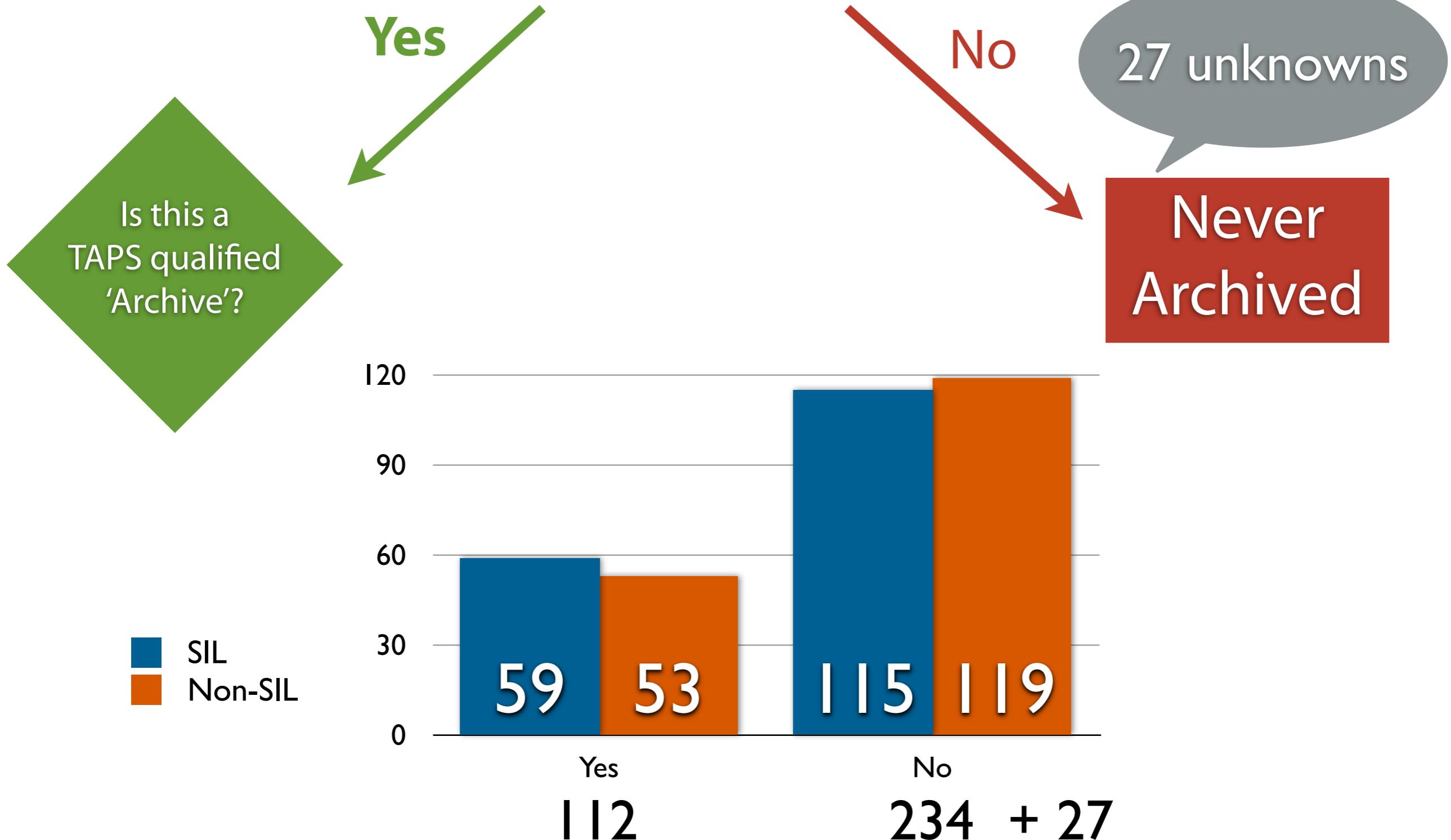


About 179 Respondents

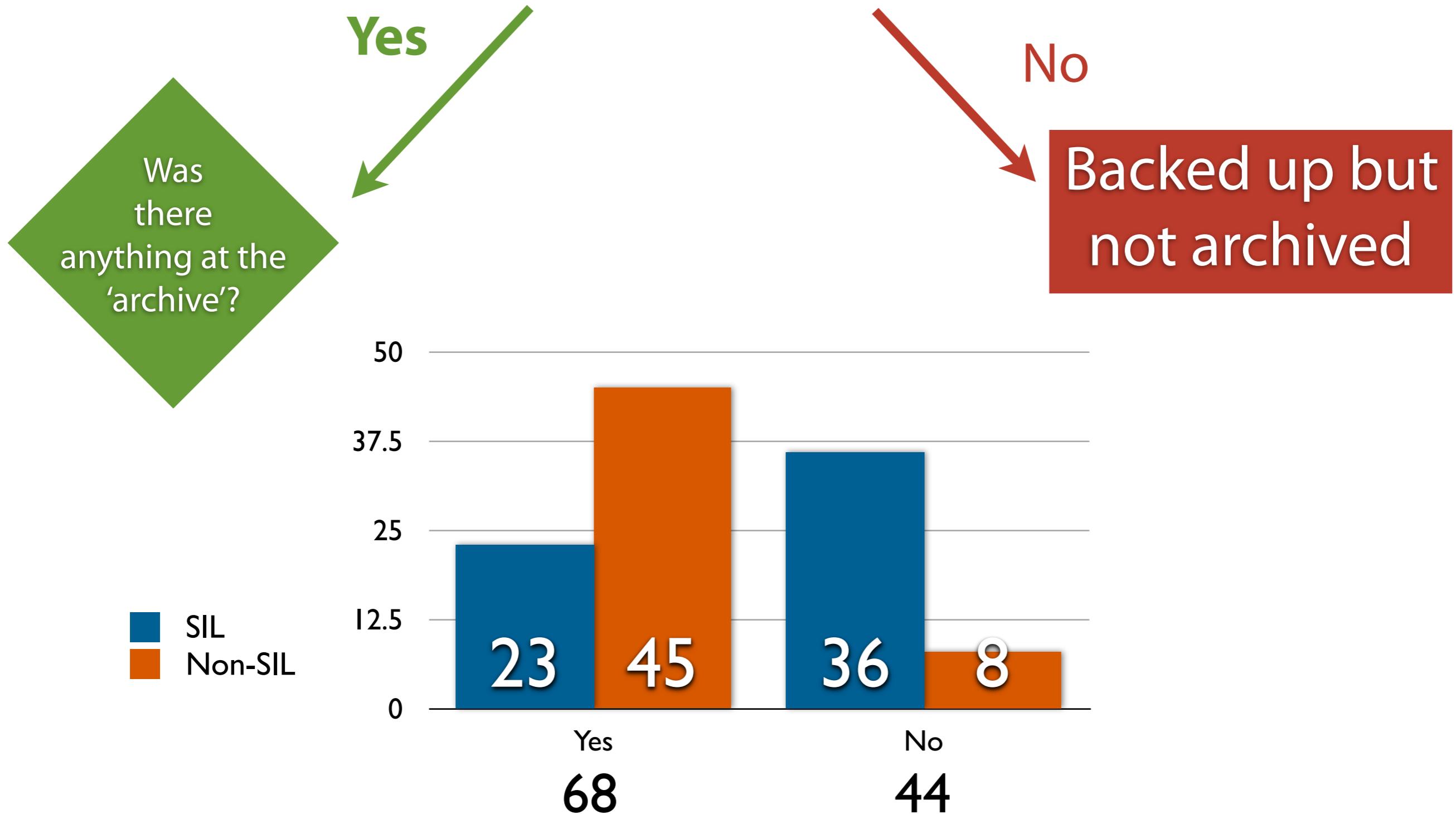
About 373 lexical datasets are represented

Participants	137	19	9	3	2	3	1	1	1	1	1	
Database records contributed	1	2	3	4	6	7	9	12	15	21	31	38

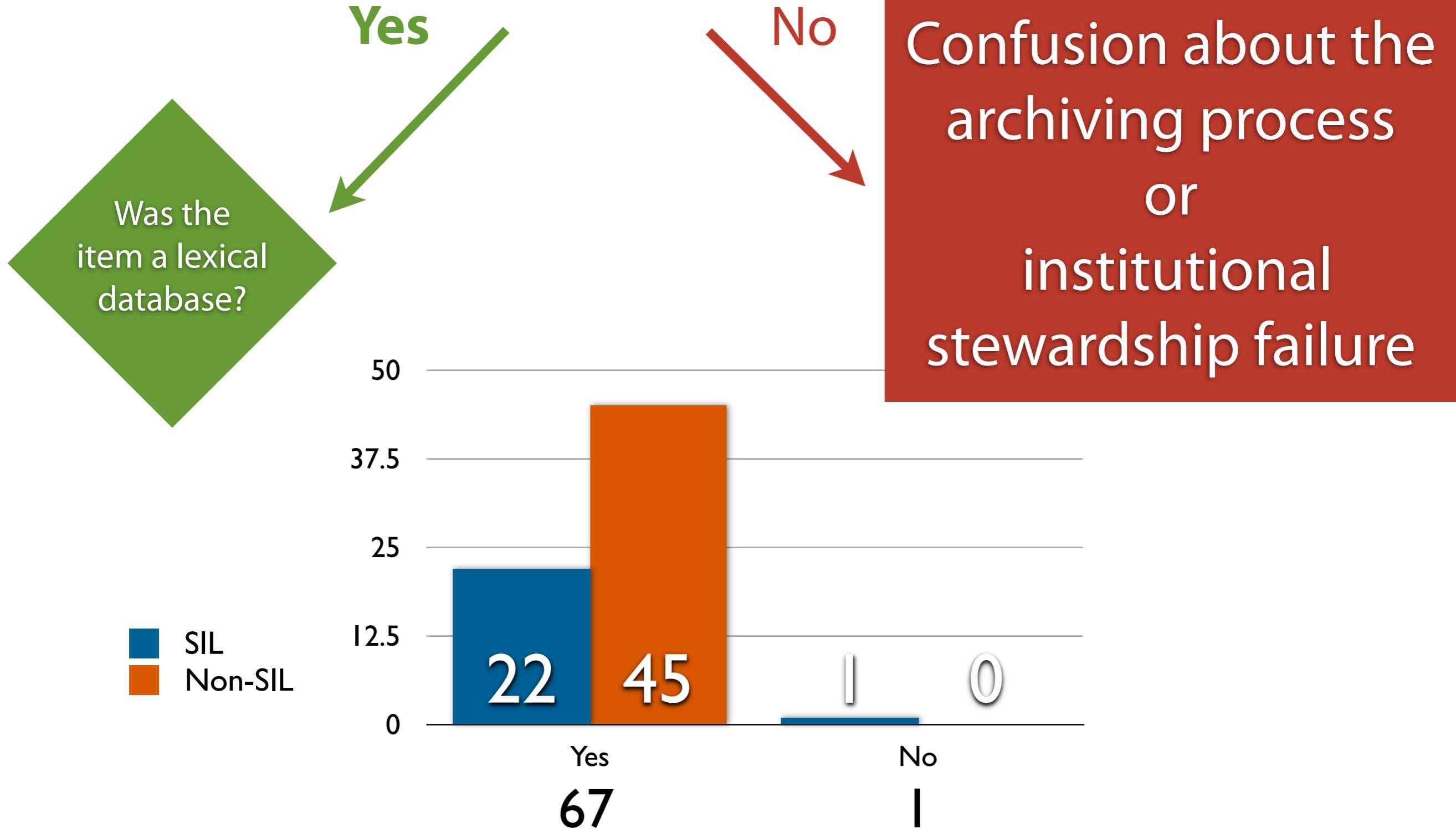
Did the respondent say they archived a lexical data set?



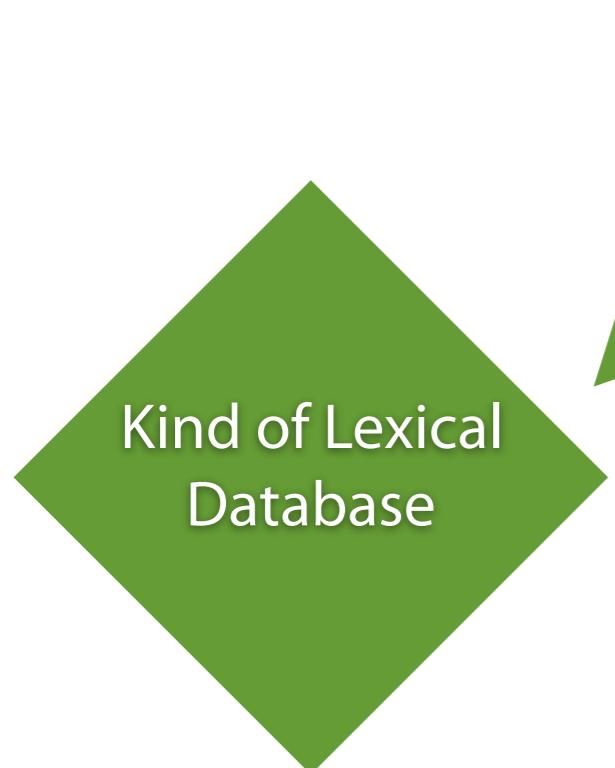
Is this a TAPS qualified 'Archive'?



Was there anything at the 'archive'?

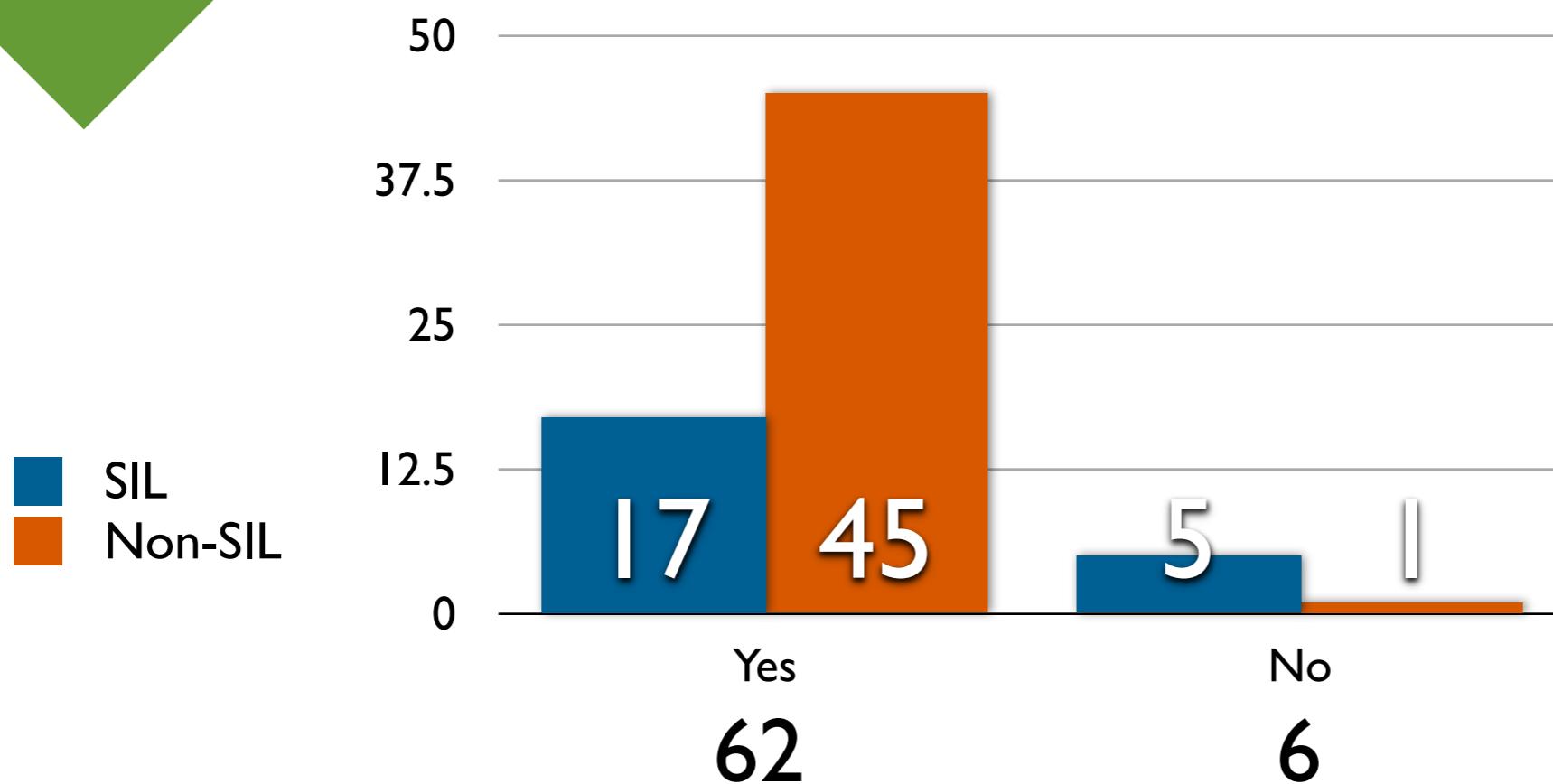


Was the item a lexical database?

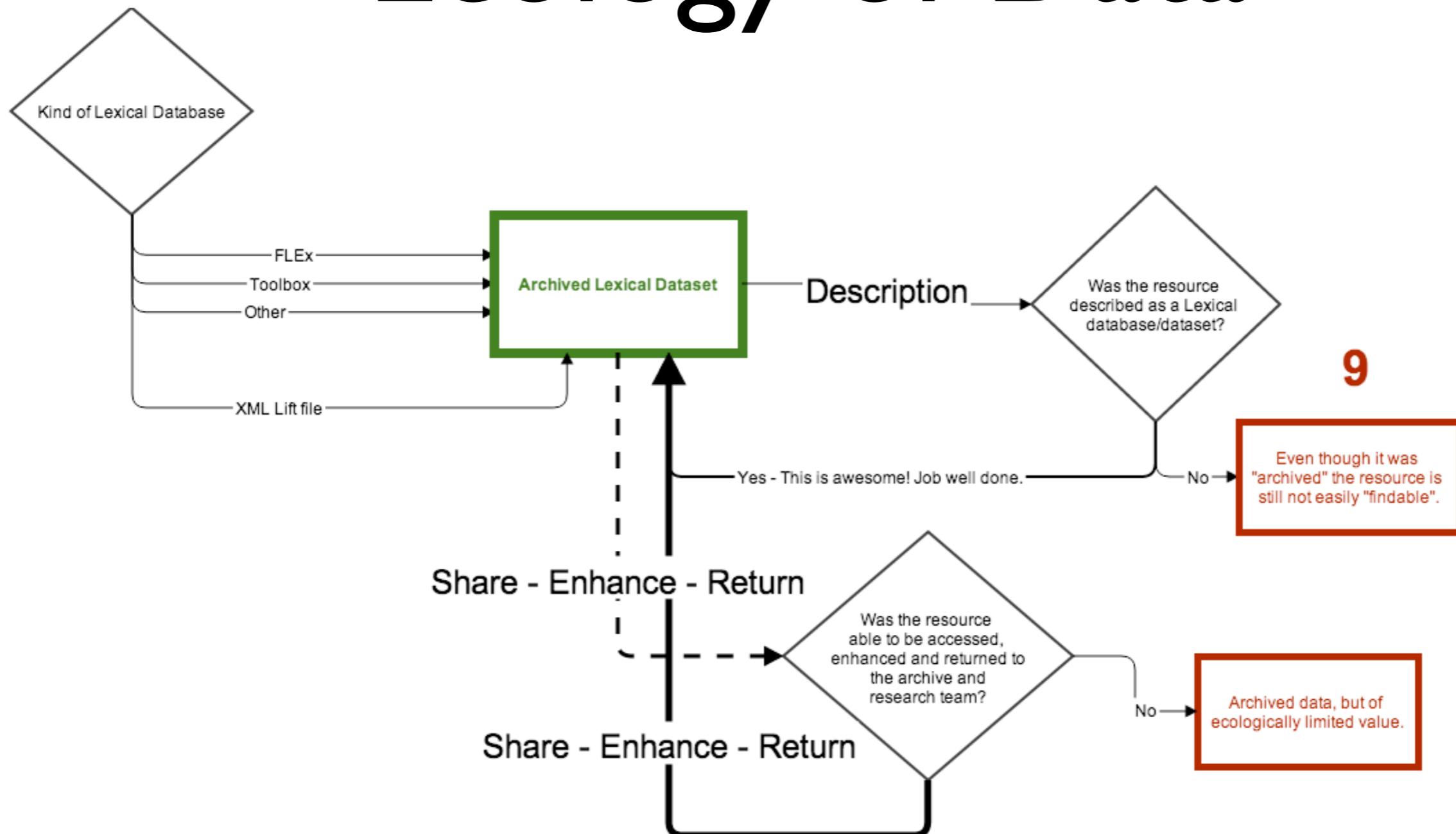


No

Confusion on what
to archive
Preservation vs.
Presentation

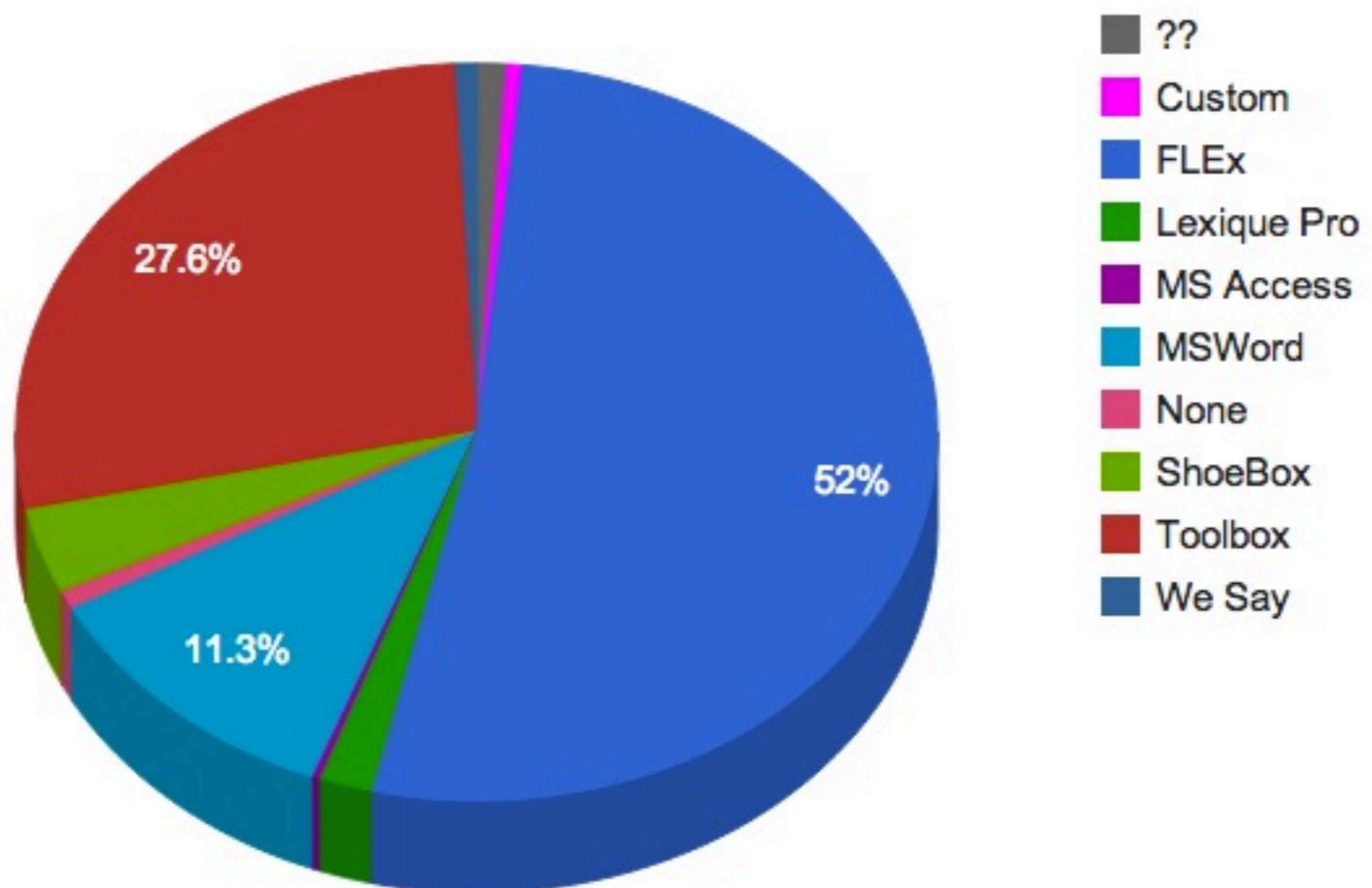


Ecology of Data



Kinds of software

Market Share based on Responses



Coverage

Coverage



Information Ecology

CITOGENESIS STEP #1:
THROUGH A CONVOLUTED PROCESS,
A USER'S BRAIN GENERATES FACTS.
THESE ARE TYPED INTO WIKIPEDIA.

THE "SCROLL LOCK" KEY WAS
DESIGNED BY FUTURE
ENERGY SECRETARY STEVEN
CHU IN A COLLEGE PROJECT.



A RUSHED WRITER CHECKS WIKIPEDIA
FOR A SUMMARY OF THEIR SUBJECT.

US ENERGY SECRETARY STEVEN CHU,
(NOBEL PRIZEWINNER AND CREATOR OF
THE UBIQUITOUS "SCROLL LOCK" KEY)
TESTIFIED BEFORE CONGRESS TODAY...



STEP #2

SURPRISED READERS CHECK WIKIPEDIA,
SEE THE CLAIM, AND FLAG IT FOR REVIEW.
A PASSING EDITOR FINDS THE
PIECE AND ADDS IT AS A CITATION.

GOOGLE IS YOUR
FRIEND, PEOPLE.

<REF>{{CITE WEB|URL=

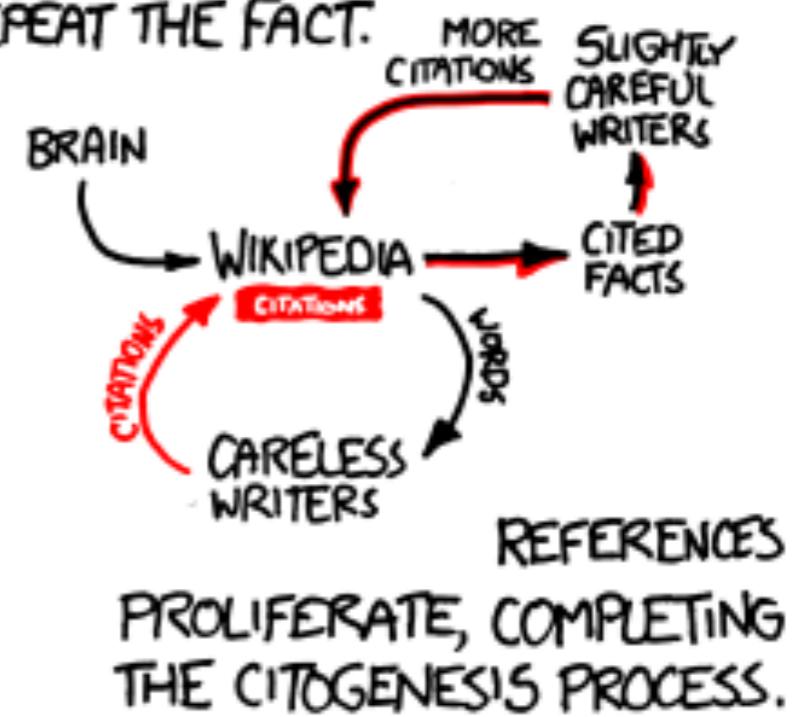


STEP #3

Information Ecology

STEP #4

NOW THAT OTHER WRITERS
HAVE A REAL SOURCE, THEY
REPEAT THE FACT.



*The other 7 people tend to be in
publishing roles.*

The other 7 people tend to be in publishing roles.

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

The other 7 people tend to be in publishing roles.

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



The other 7 people tend to be in publishing roles.

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



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).

Archivists



Not Directly Archiving

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 other 7 people tend to be in publishing roles.

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

141 Total submitters

The other 7 people tend to be in publishing roles.

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



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).

141 Total submitters

- 122 RAMP users

The other 7 people tend to be in publishing roles.

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

Archivists
↓



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).

Archivists

141 Total submitters

- 122 RAMP users
- ` 19 non-RAMP users

The other 7 people tend to be in publishing roles.

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



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).

Archivists



141 Total submitters
- 122 RAMP users
` 19 non-RAMP users

12 of 19 only make only 1 submission
The other 7 people tend to be in publishing roles.

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

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).