

Tutorial on LingSync

*A Free Tool for Creating and Maintaining a Shared Database
For Communities, Linguists and Language Learners*



Thursday, May 8, 2014
9:00 am - 12:00 pm

Joel Dunham (UBC)
Elise McClay (McGill)
Hisako Noguchi (Concordia)

Plan of the Tutorial

9:00-9:10	Introduce the LingSync team
9:10-9:40	<i>Why LingSync?</i> (presentation)
9:40-9:50	Learn X/ Georgian Together video demo
9:50-10:10	Question period 1
10:10-10:20	<i>coffee break</i>
10:20-10:35	<i>LingSync Practical Tutorial</i> (presentation)
10:35-10:45	Question period 2
10:45-11:00	Participatory collaborative data entry
11:00-11:30	Question period 3

Introductions

- Elise McClay
- Hisako Noguchi
- Gretchen McCulloch
- Louisa Bielig
- Gina Cook

Joel Dunham

- Ph.D. candidate at UBC Linguistics
- Fieldwork: Blackfoot & Okanagan (& Lillooet)
- Research: aspectual semantics, noun incorporation
- self-taught developer and aspiring computational linguist
- Dissertation title: *The Online Linguistic Database: Software for linguistic fieldwork*
- Recent contributor to LingSync

Why LingSync?

- What is LingSync?
- Why would I want to use it?

Outline

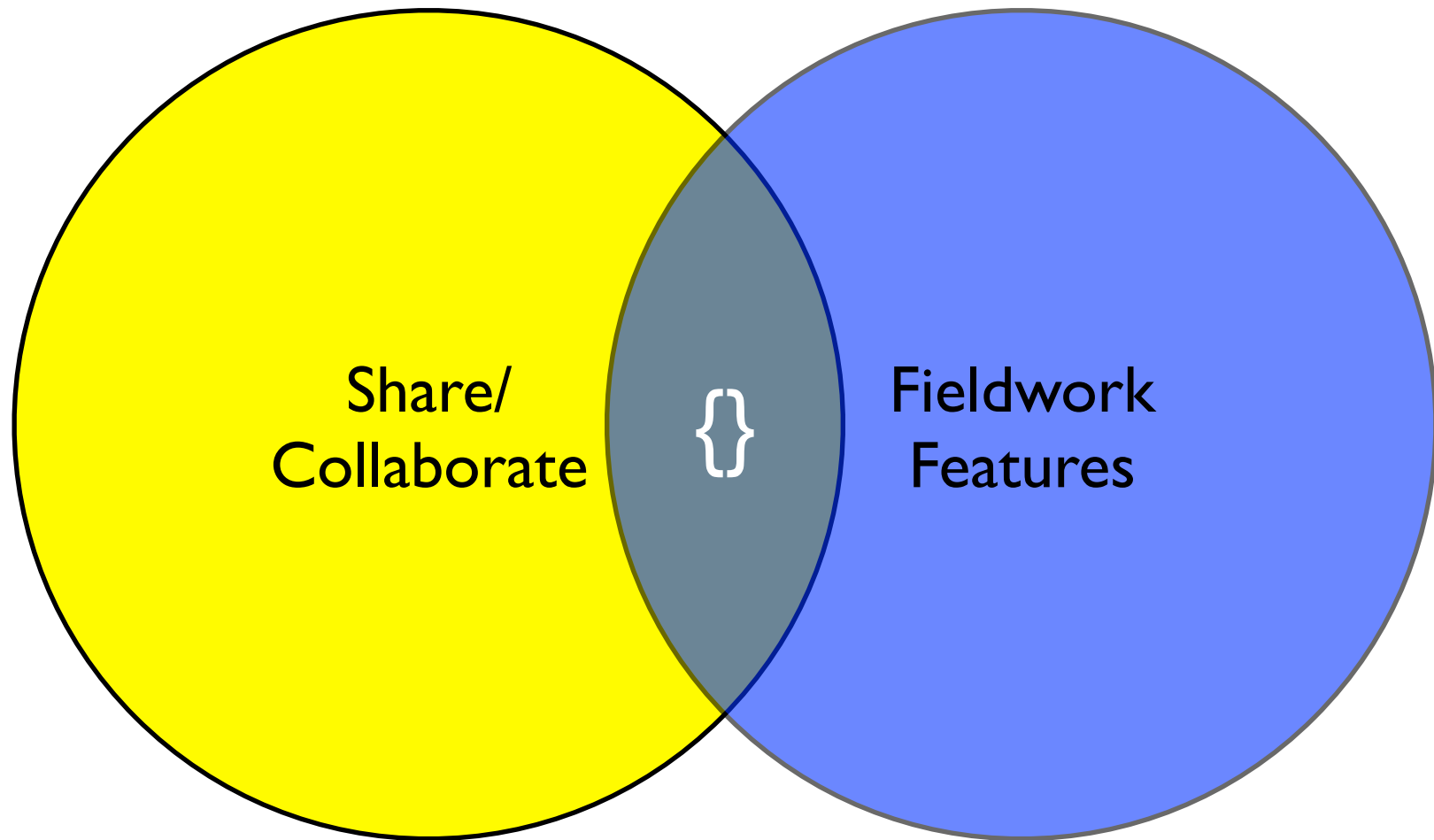
- What is linguistic fieldwork?
- Collaborative approaches to linguistic fieldwork are valuable.
- There is a dearth of collaborative fieldwork software.
- LingSync facilitates collaborative fieldwork and does a whole lot more.

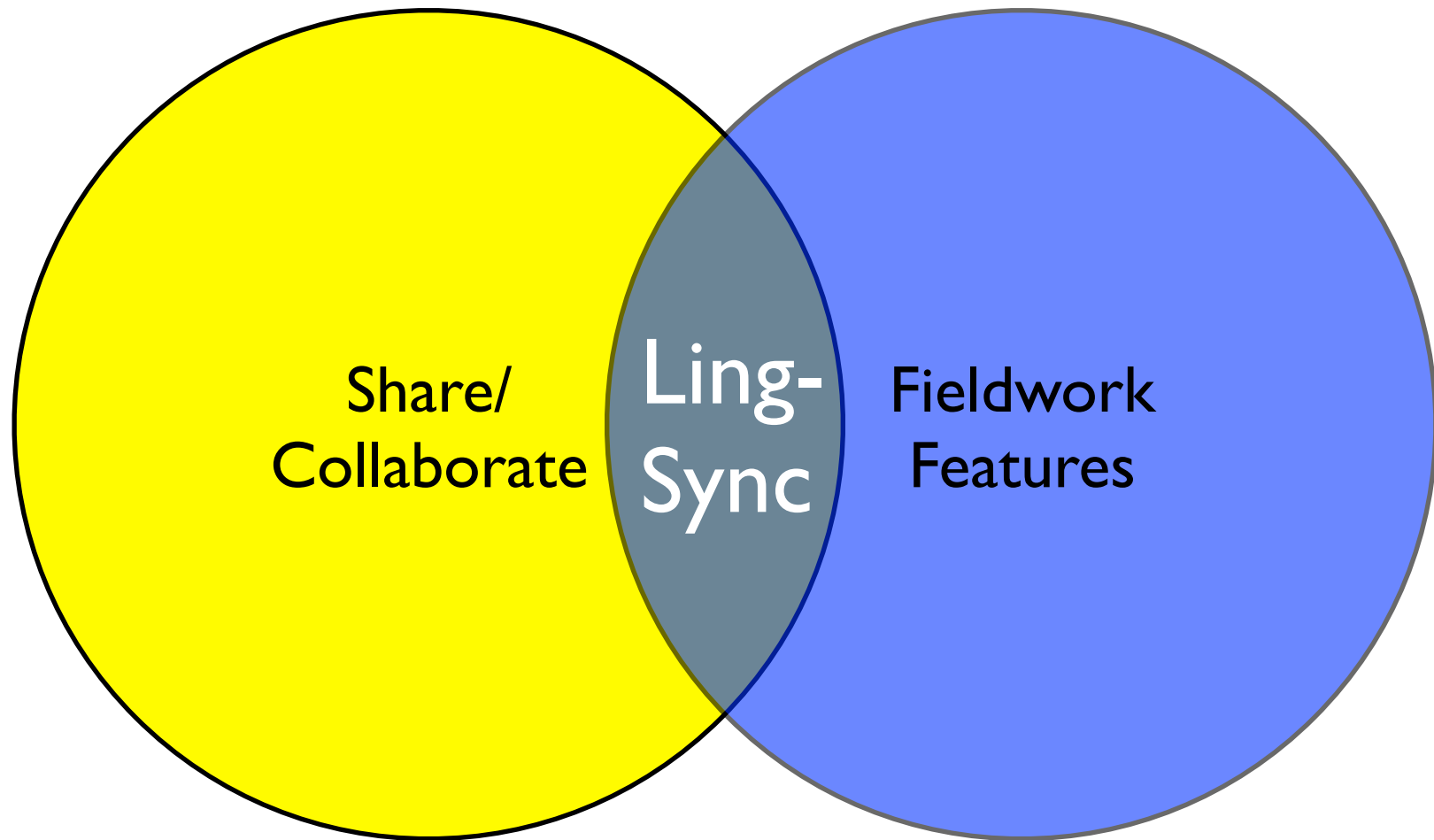


Share/
Collaborate



Fieldwork
Features



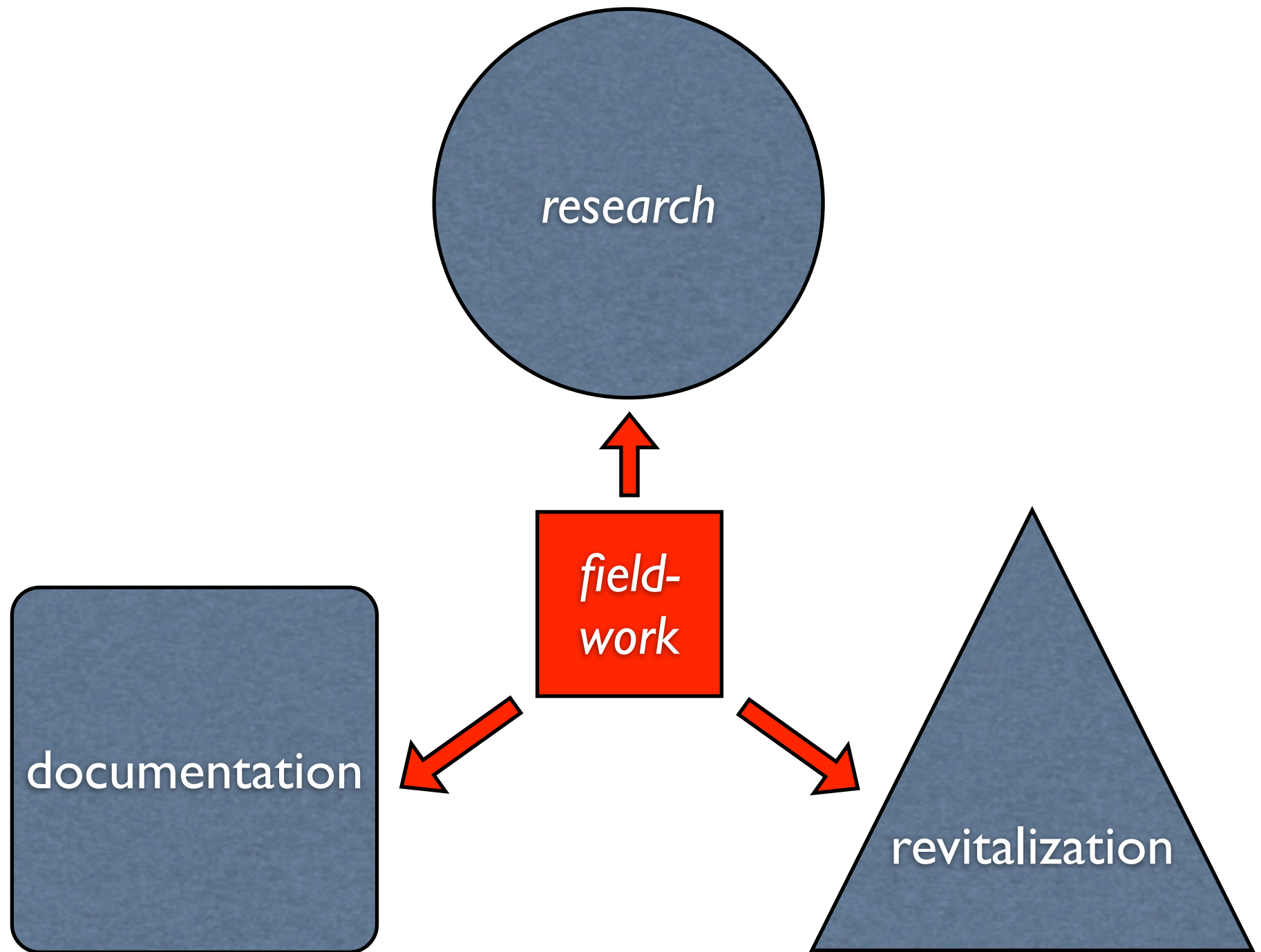


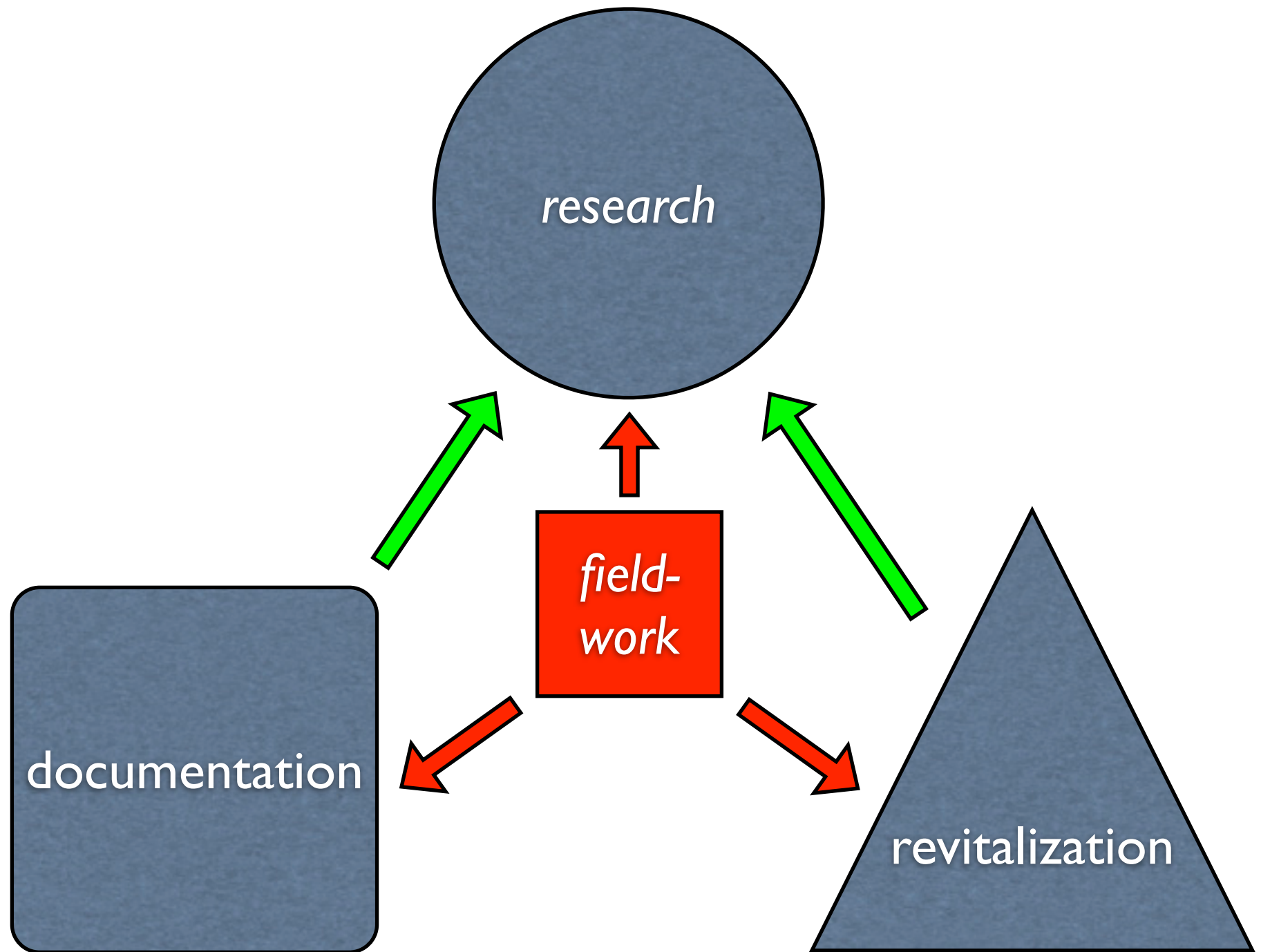
Linguistic Fieldwork

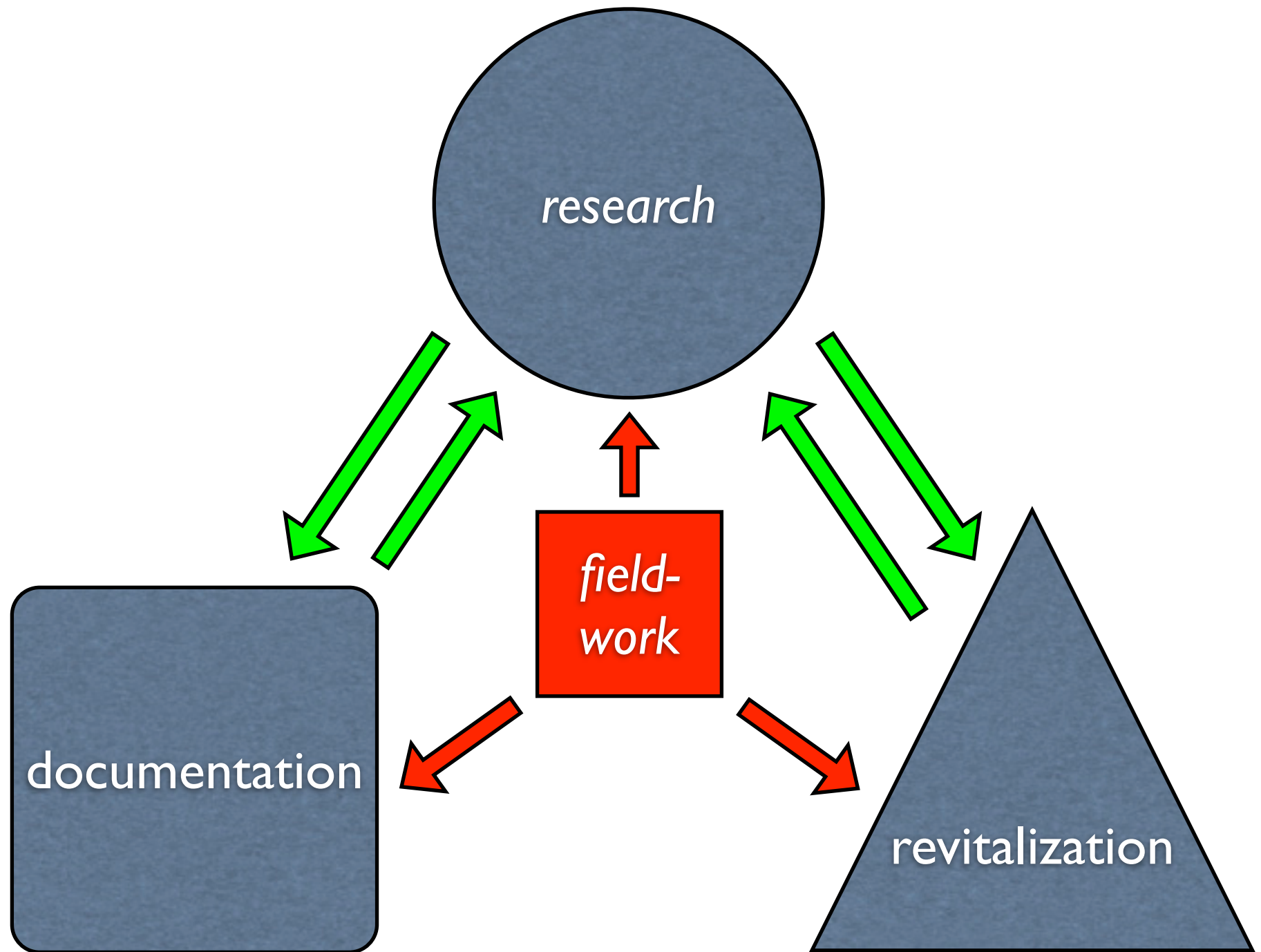
working with speakers to generate artifacts
that encode knowledge of a particular
language

Fieldwork Trichotomy

1. Documentation
2. Revitalization
3. Research







Fieldwork-based Research

- highly endangered, under-studied, and under-documented languages
- field methods courses: in their ~5 years of impact, generate a LOT of transcriptions, audio, and low-level analyses
- language-learning groups (proficiency required for high-level research), text collection

Research benefiting revitalization

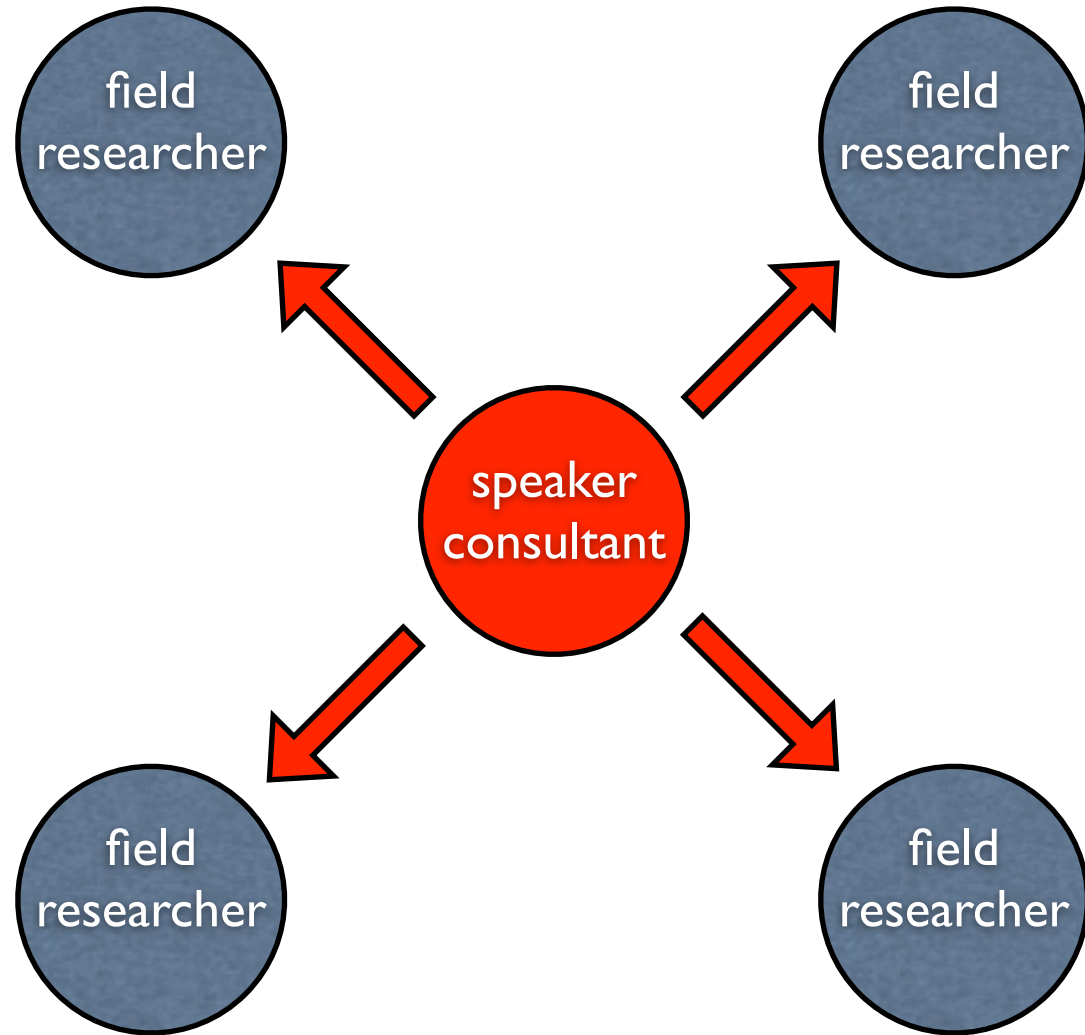
First of all, let us look at how *wa7* affects the different event types we introduced in 17.3. Examine the example in (11), together with its English translations:

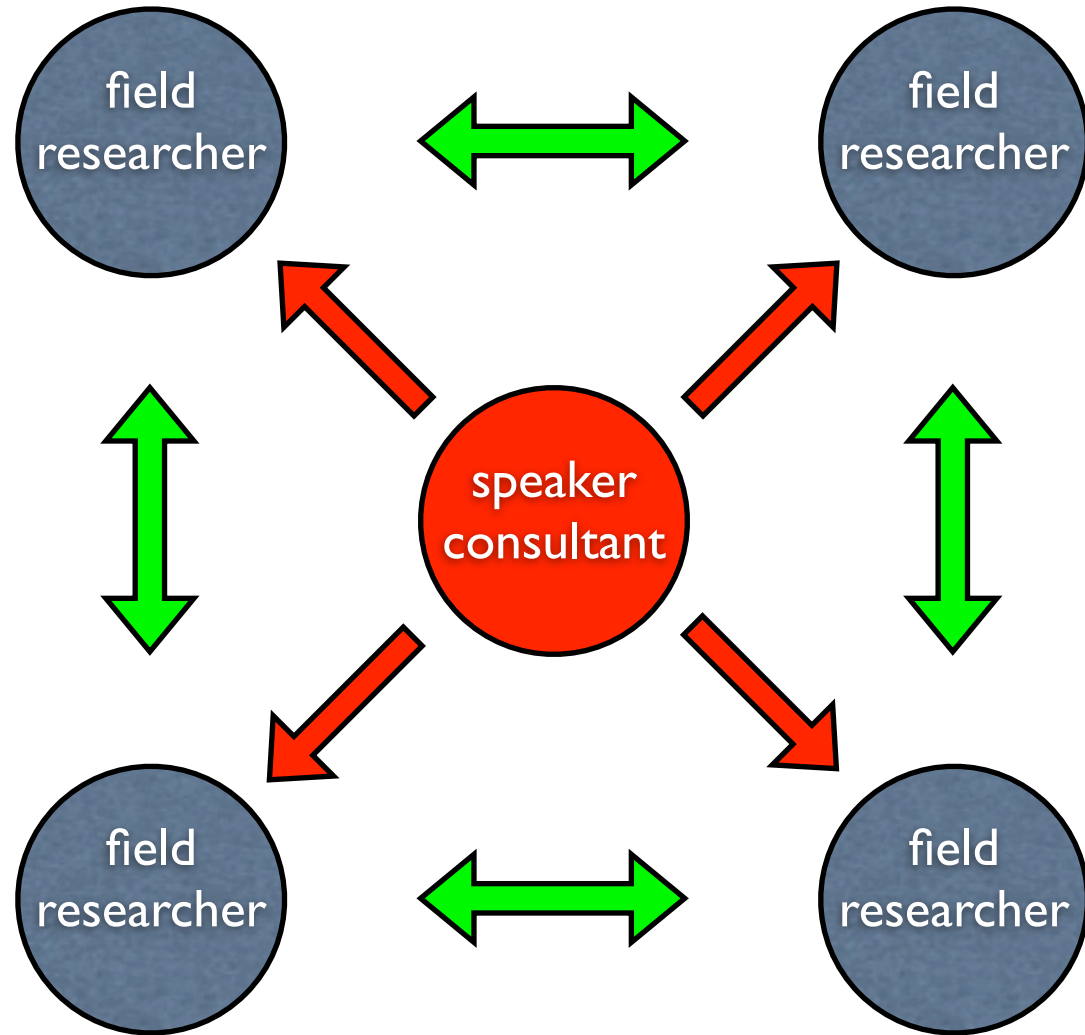
11. *Wá7lhkan pun.*

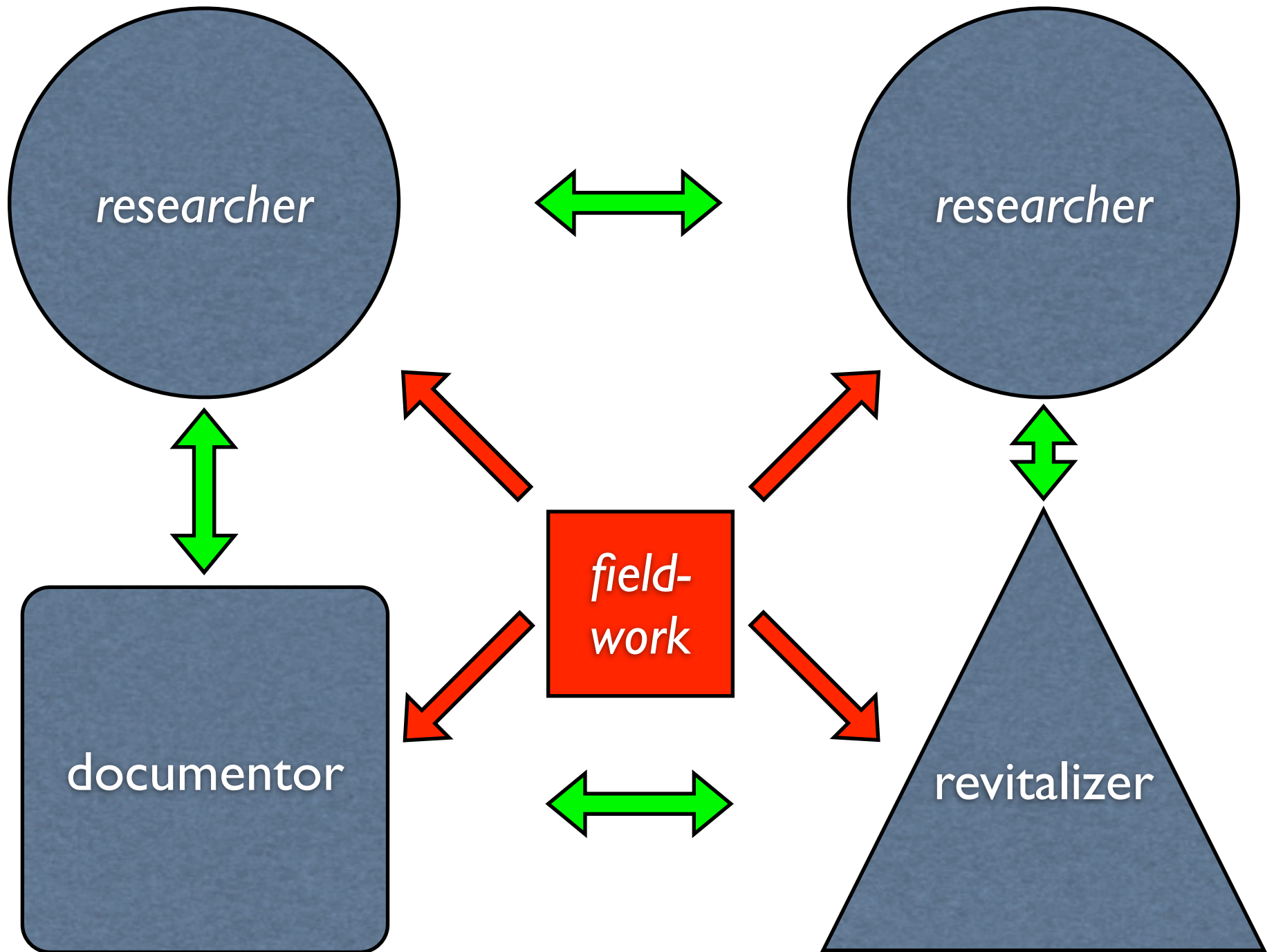
* “I am/was finding it.”

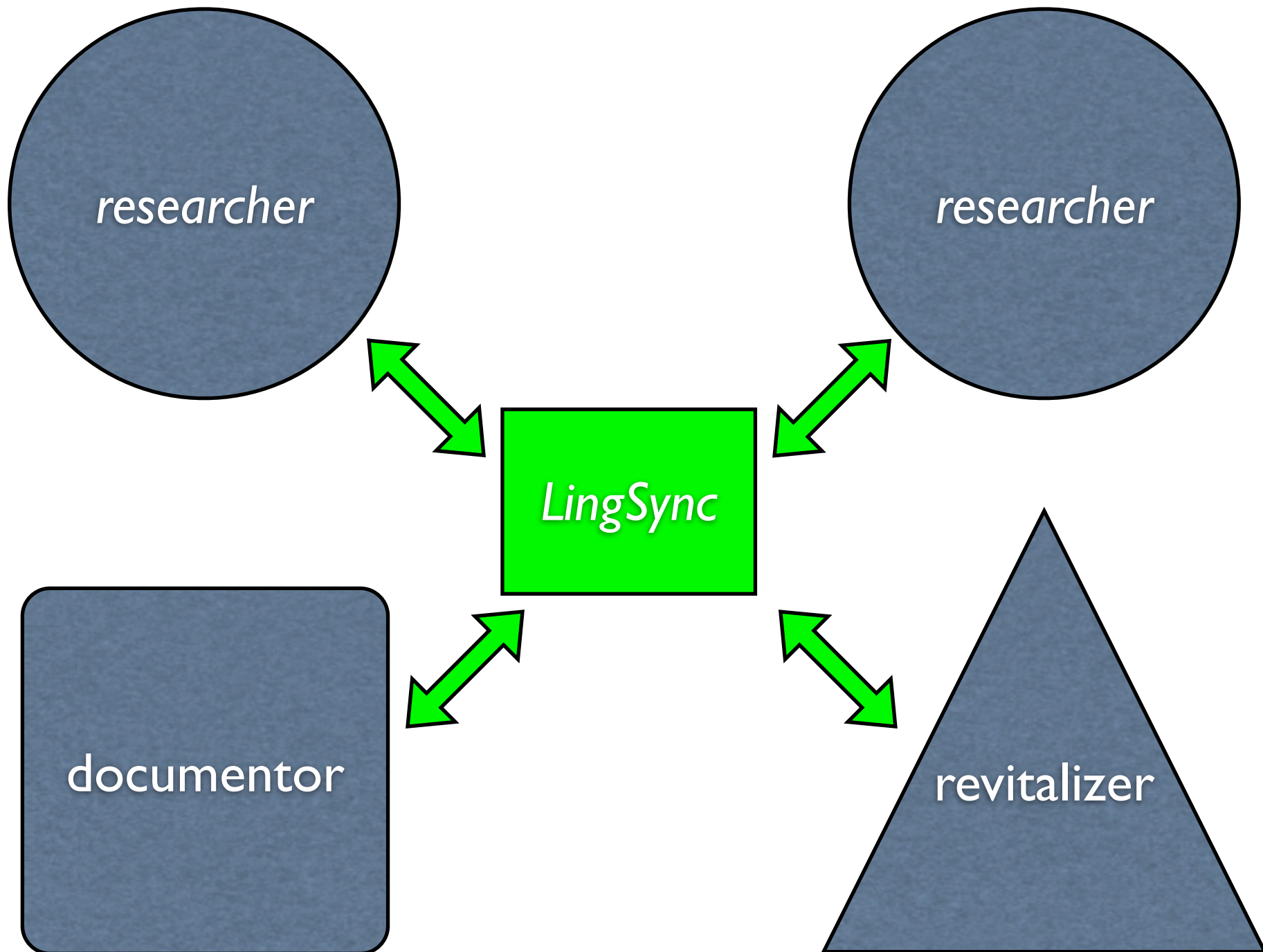
✓ “I find it (habitually).”

(Davis 2006, ch. 18, *with modification*)





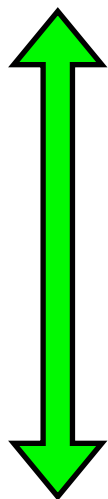




What is LingSync?

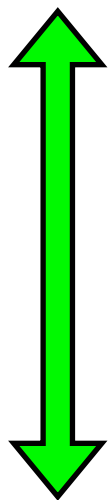
*A Free Tool for Creating and Maintaining a Shared
Database For Communities, Linguists and
Language Learners*

LingSync
corpus



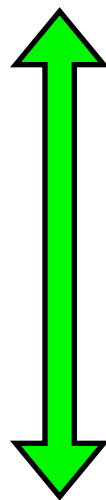
field-
worker

*LingSync
corpus*



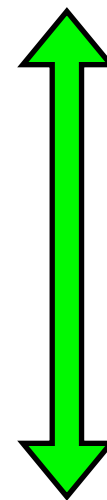
field-
worker

*LingSync
corpus*

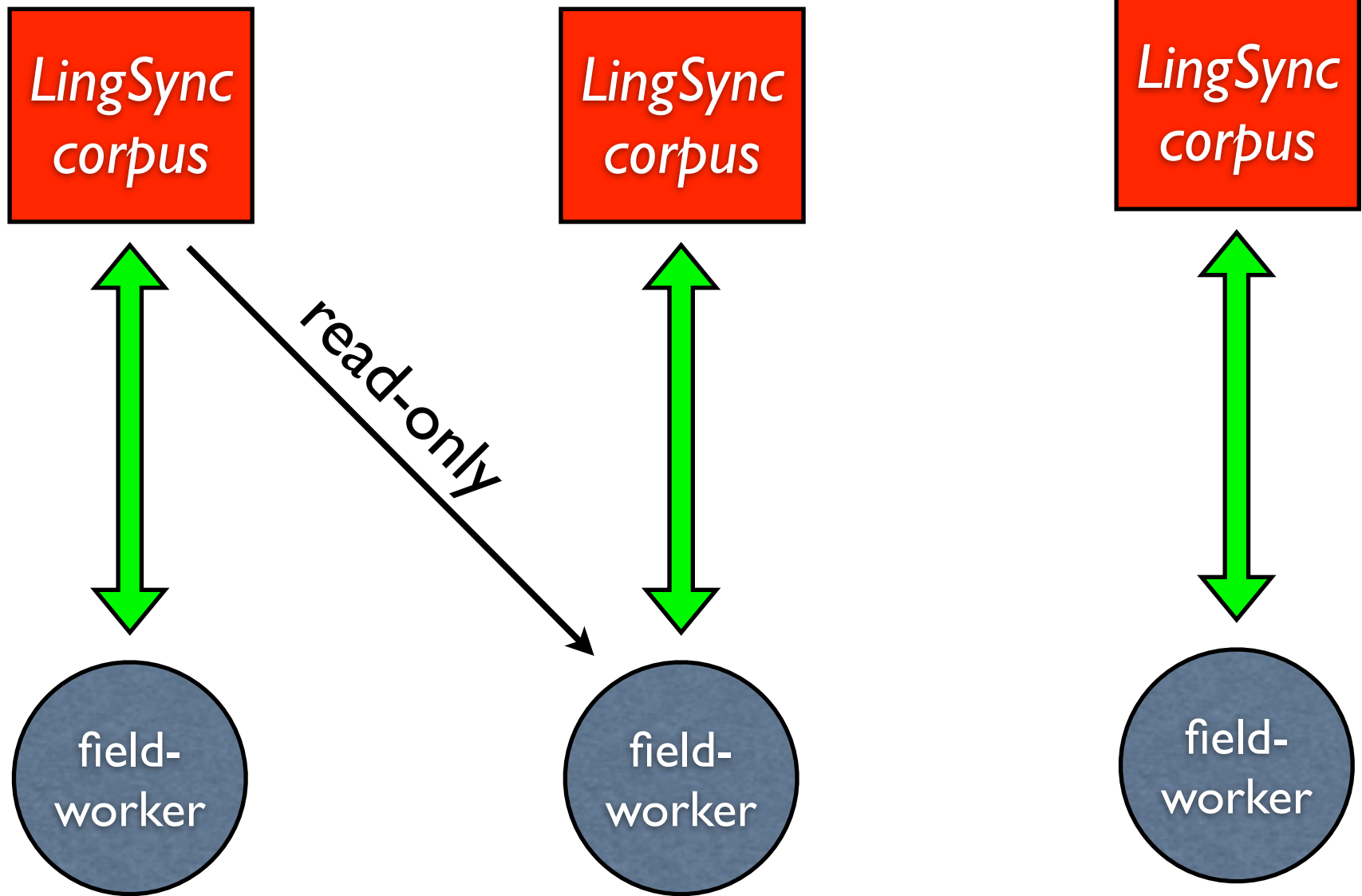


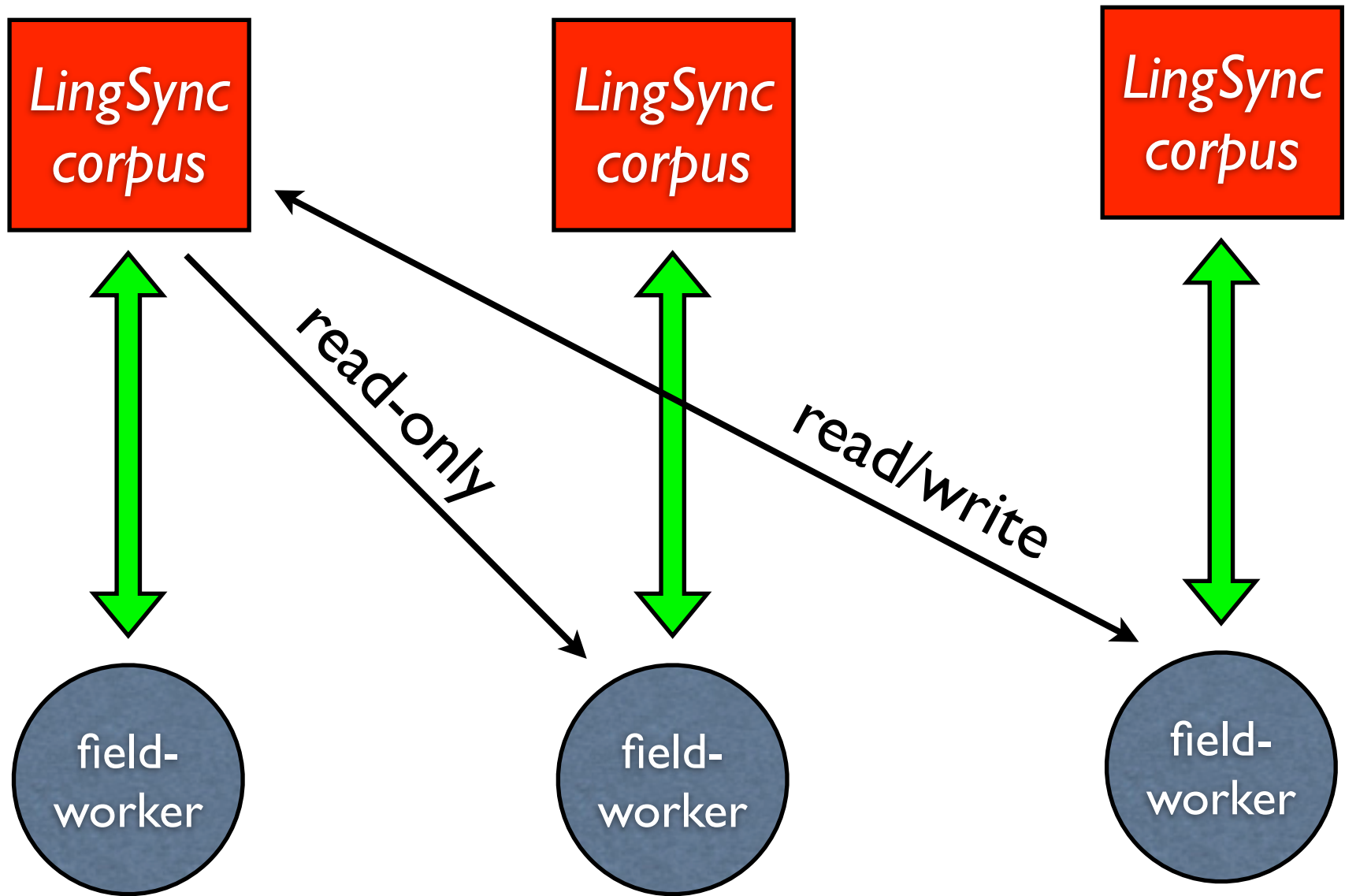
field-
worker

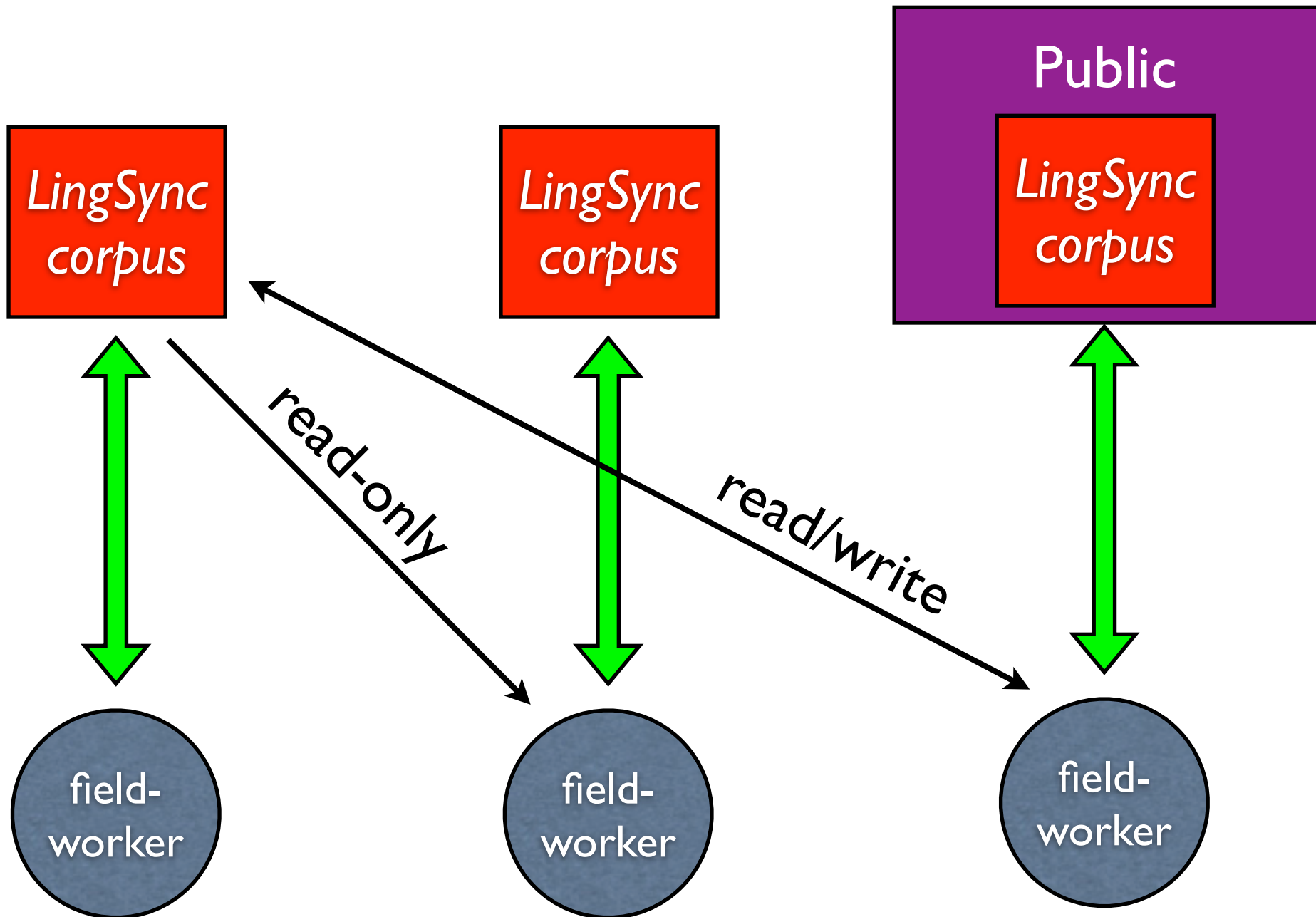
*LingSync
corpus*

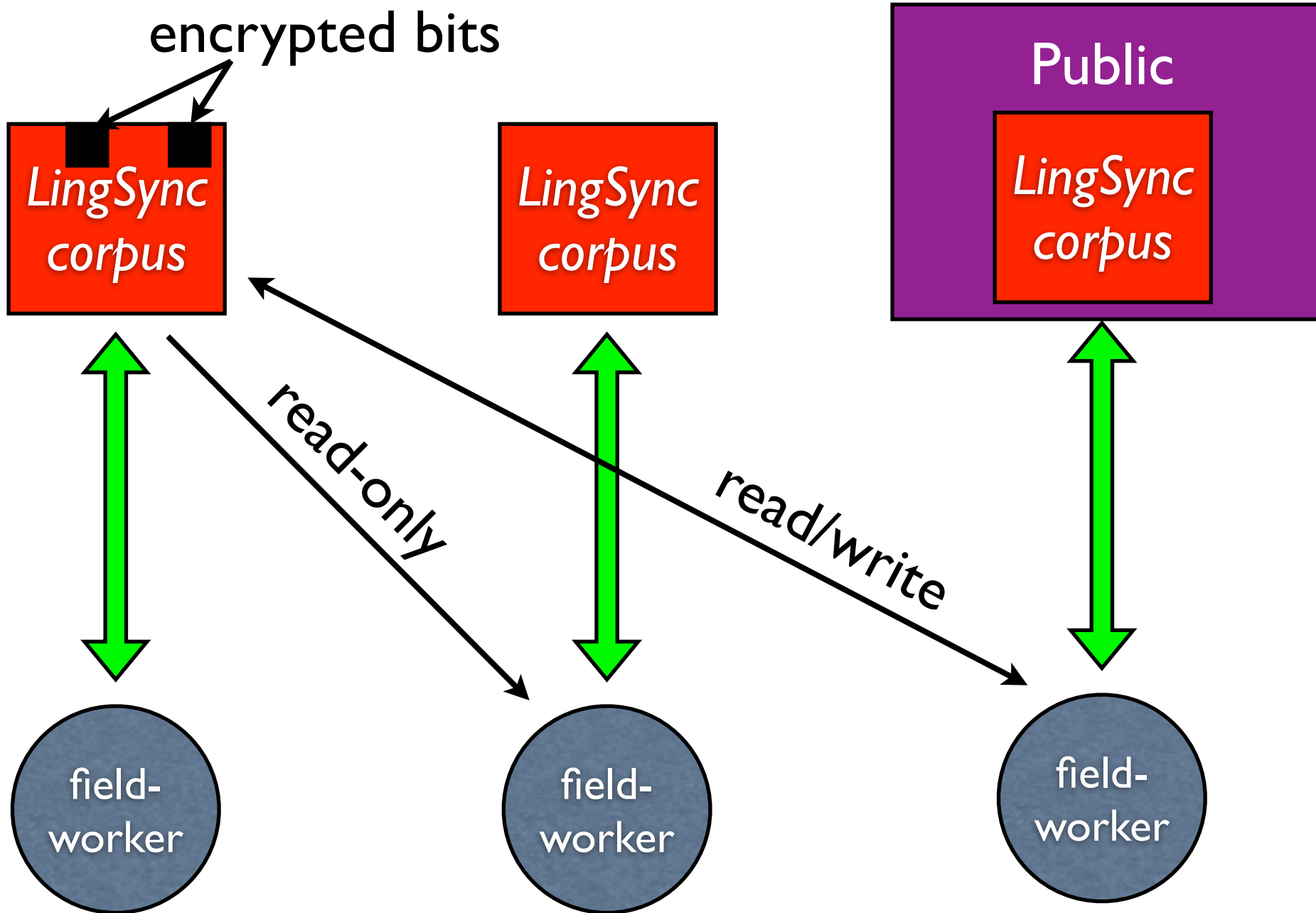


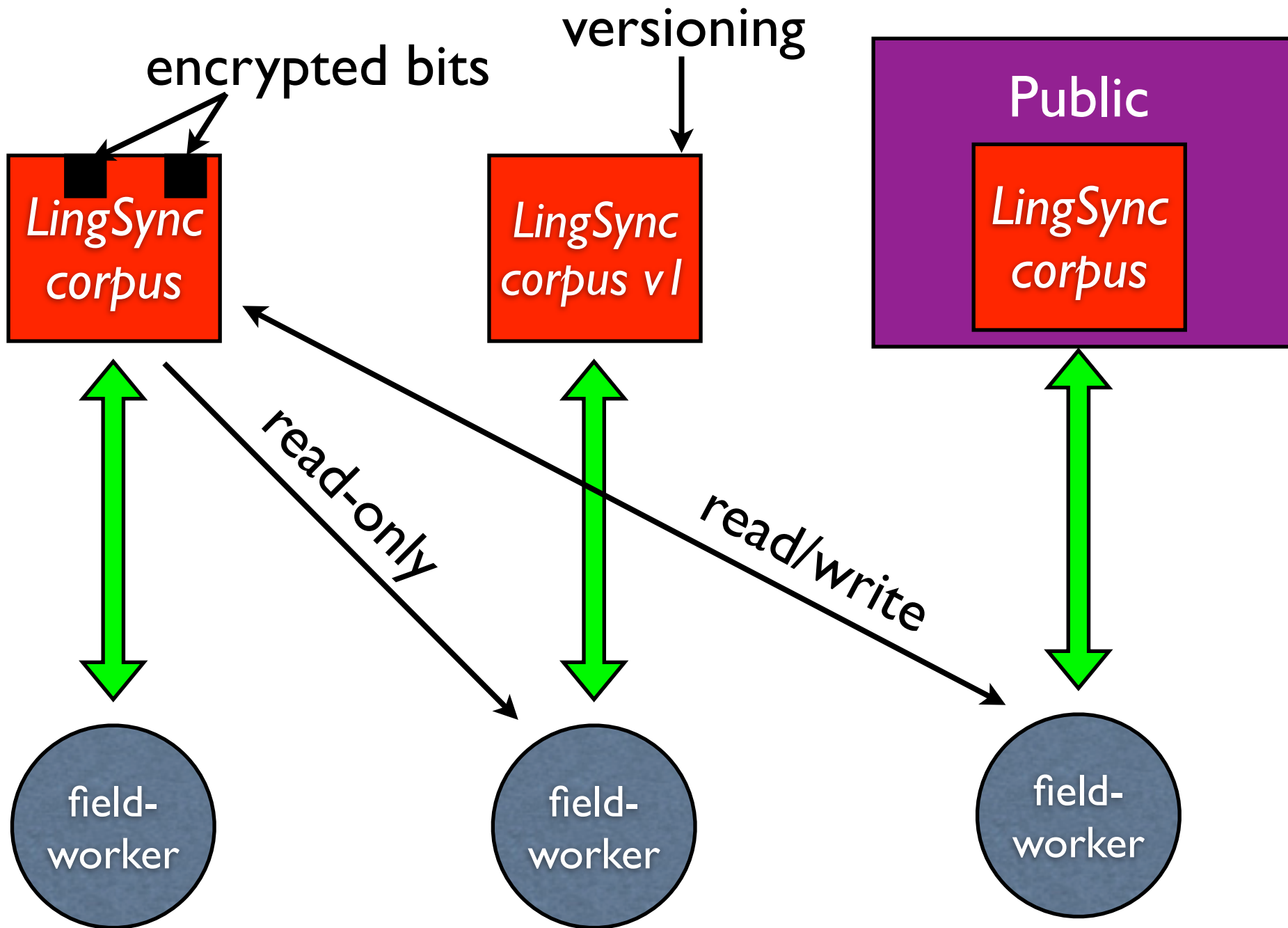
field-
worker

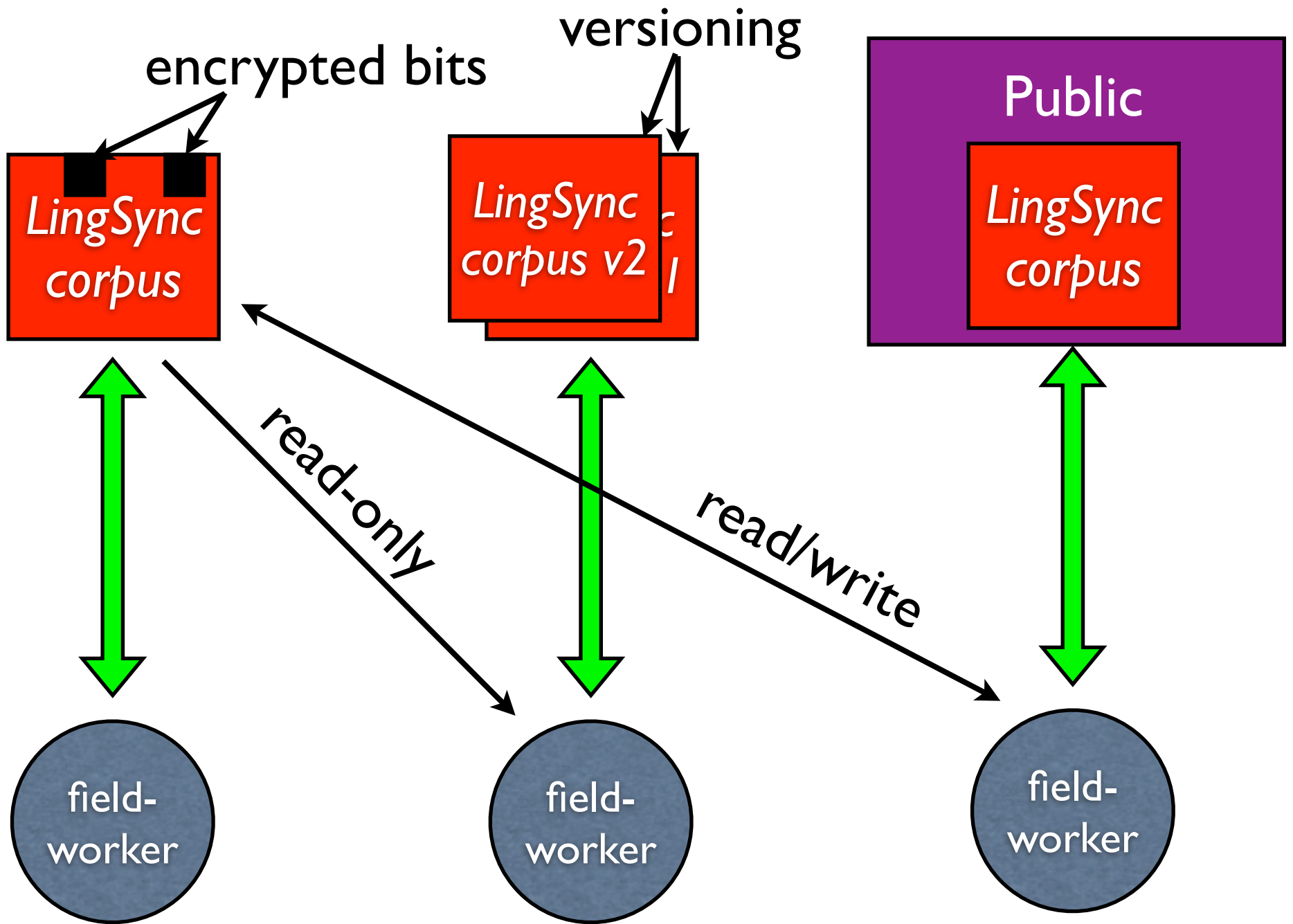


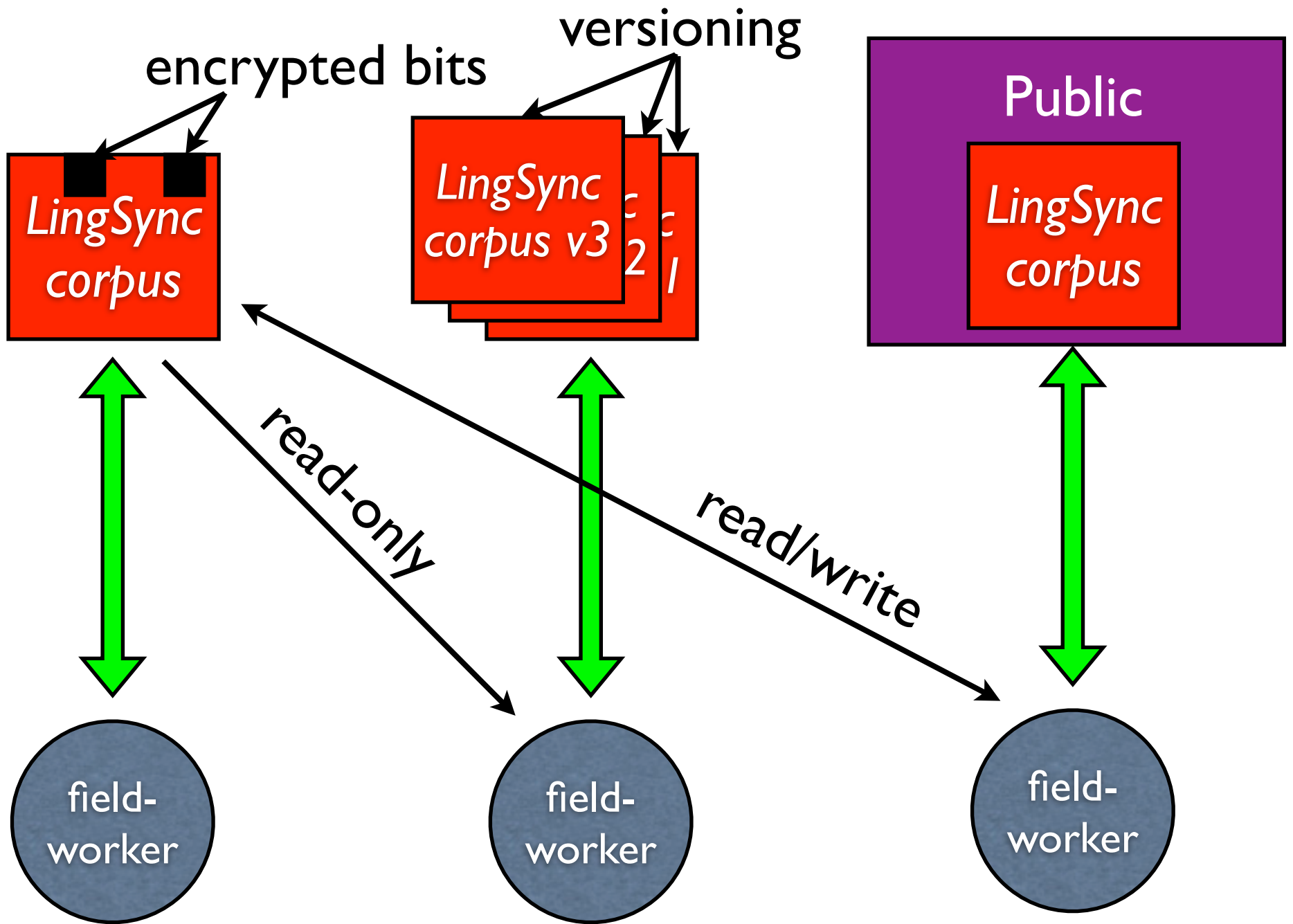


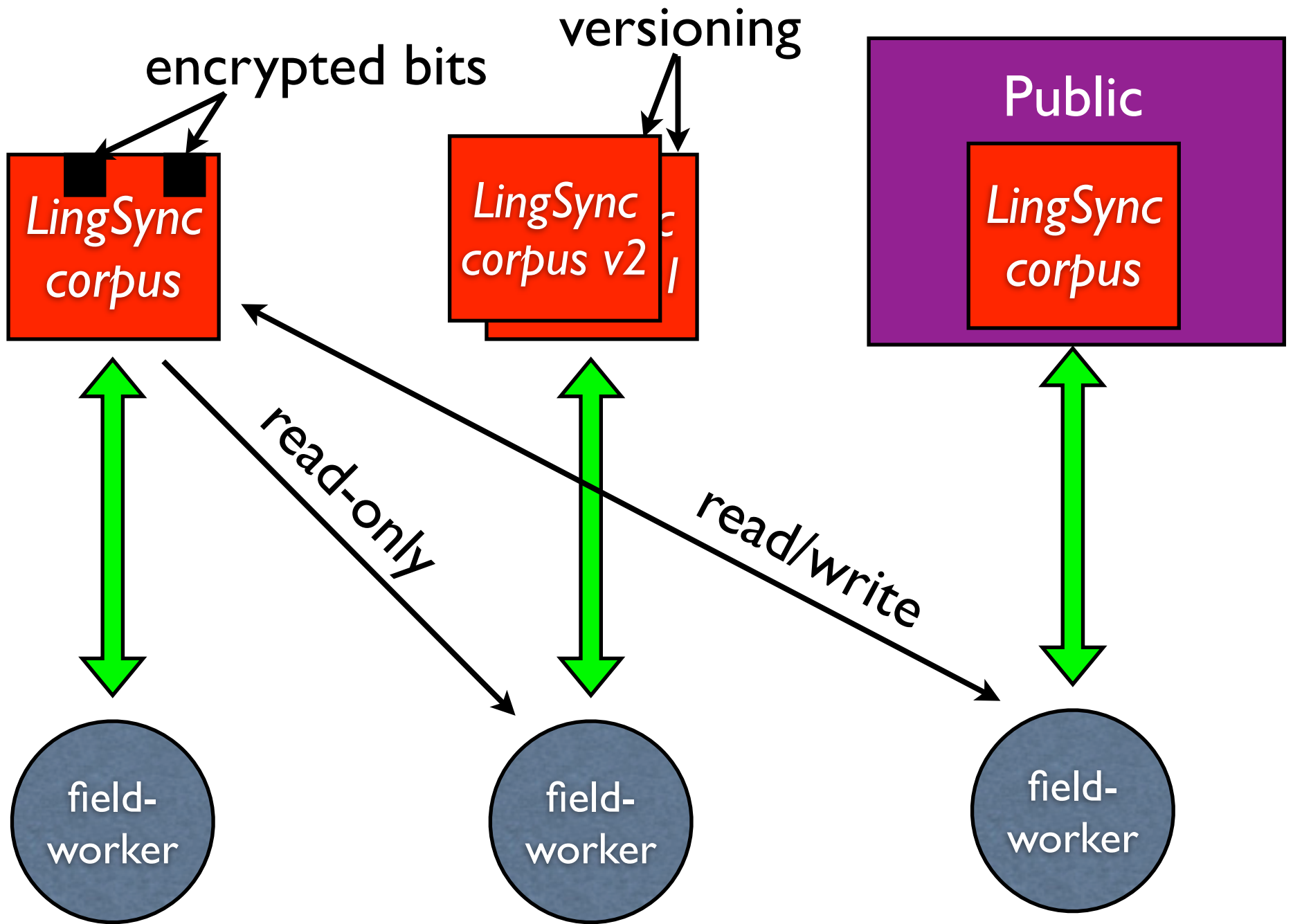


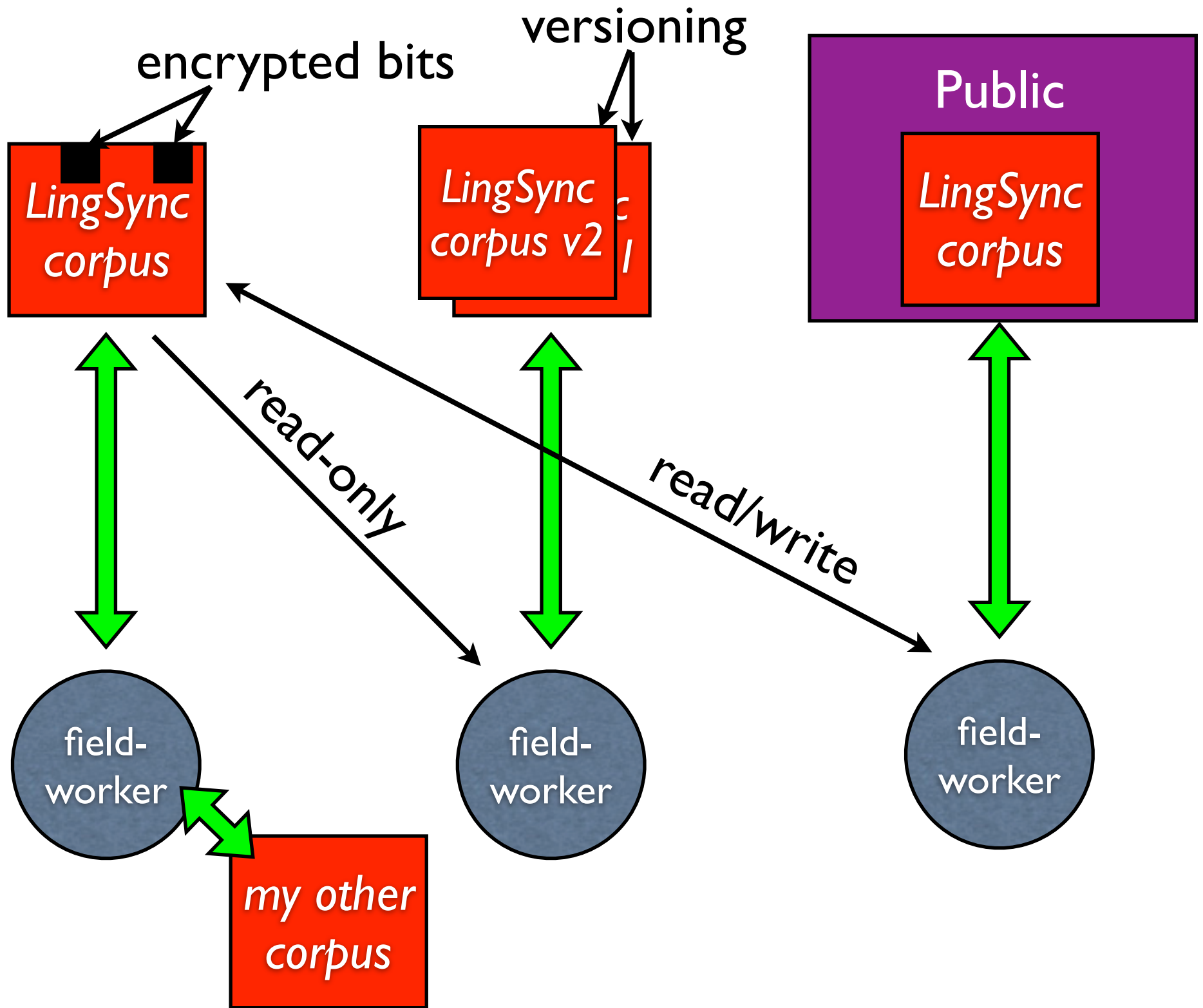






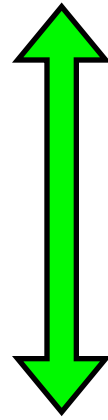







*LingSync corpus
web service*

*LingSync corpus
web service*




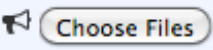

LingSync
Spreadsheet
interface

LingSync Spreadsheet

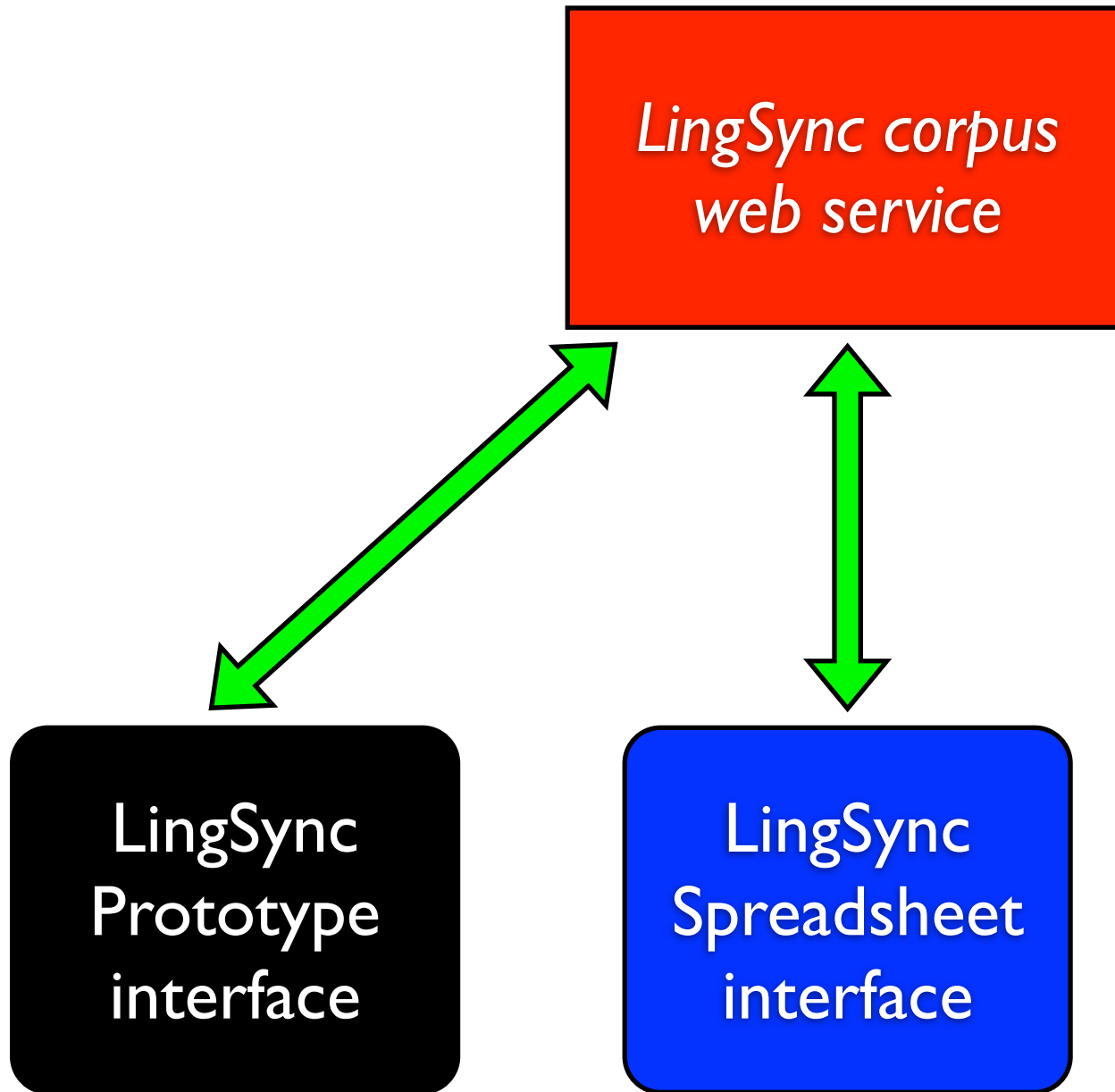
[SETTINGS](#) [DATAENTRY](#) [SEARCH](#) [SESSION](#) [FAQ](#) [Semi-auto Glosser](#) [Change this s](#) [Blackfoot](#)

Utterance	Translation
Morphemes	Judgement
Gloss	Tags
Comments	






 No file chosen 

nítsapsikaapino'toaa nítsapsikaapino'toaa ?	I poked him in his black eye * --
annááhk John iihpómmatsii ómi imitáá ann-wa-hka John ohpomm-atw-yii om-yi imitáá DEM-PROX.SG-INVS John buy-TRANS-DIR DEM-OBV.SG dog	John bought that dog. -- --



LingSync Prototype

 poke  Blackfoot - Change this session...  Notifications

Unicode Drag and Drop

Data Entry Area (1-5 datum)









utterance ? nĩtsapsikaapino'toaa

morphemes ? nĩtsapsikaapino'toaa


gloss ? ?

translation ? I poked him in his black eye

validationStatus ? Checked

Checked

 jrwdunham 1 minutes ago

Context I provided to Consultant: Sam is one of those people with two different coloured eyes. One eye is blue and the other is black. BB knows that I poked Sam in the eye. She asks me 'which eye did you poke Sam in?'. I try the present form as a response.

Search Result





Title:

1

poke

Description:

This is the result of searching for : poke In Blackfoot on Mo
GMT-0700 (PDT)

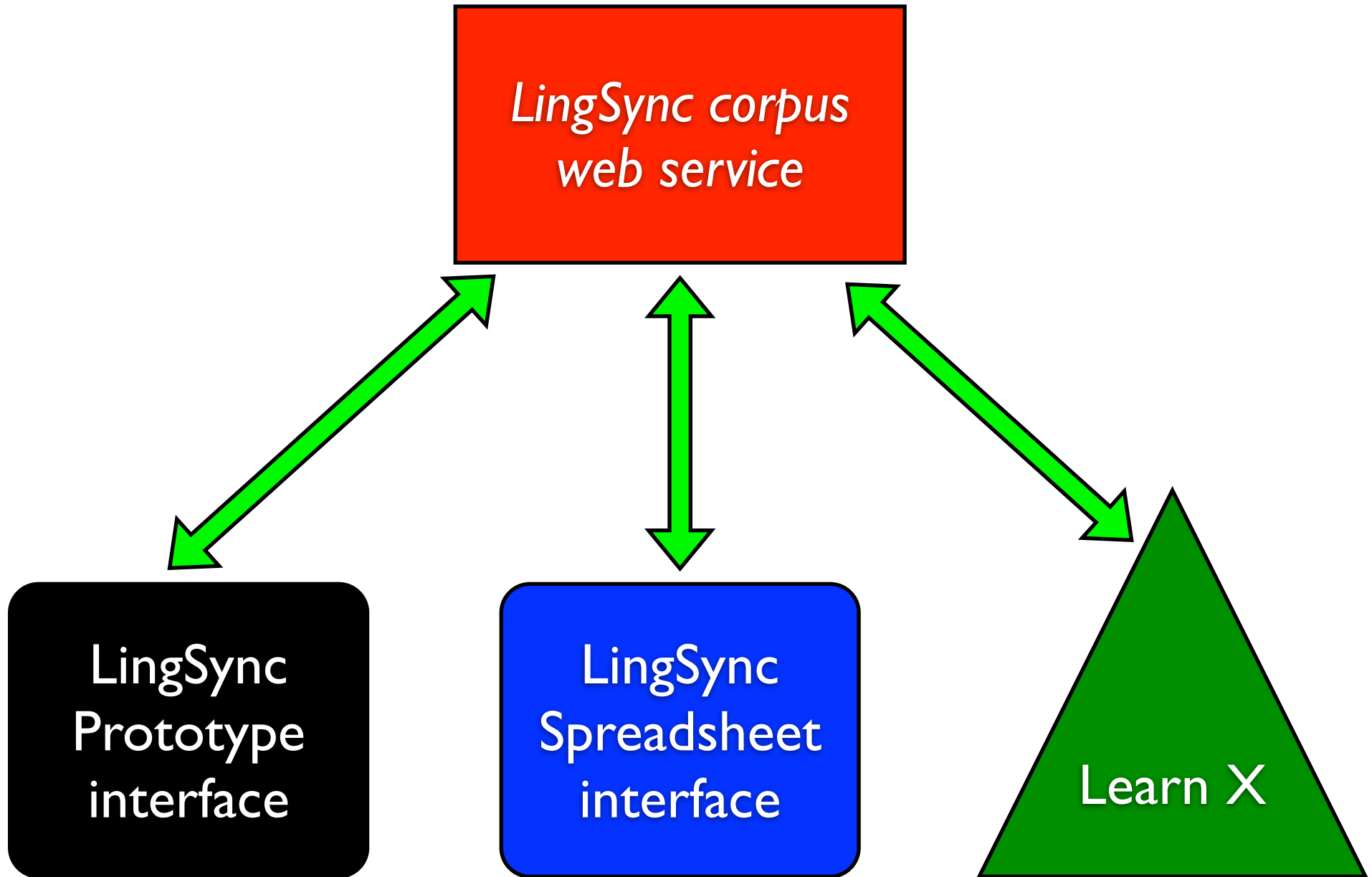
    LaTeX CSV

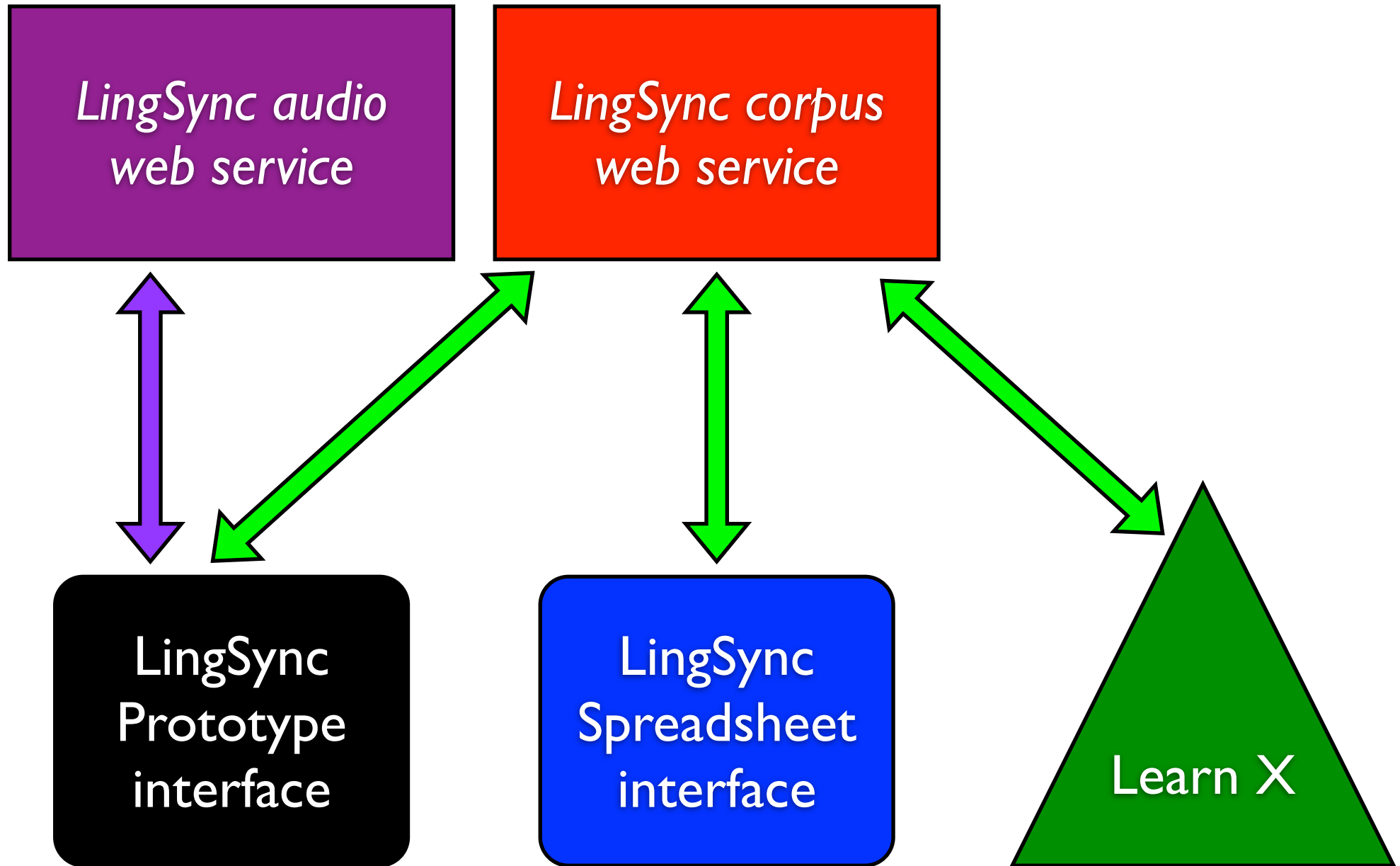
* nĩtsapsikaapino'toaa
nĩtsapsikaapino'toaa
?
I **poked** him in his black eye'

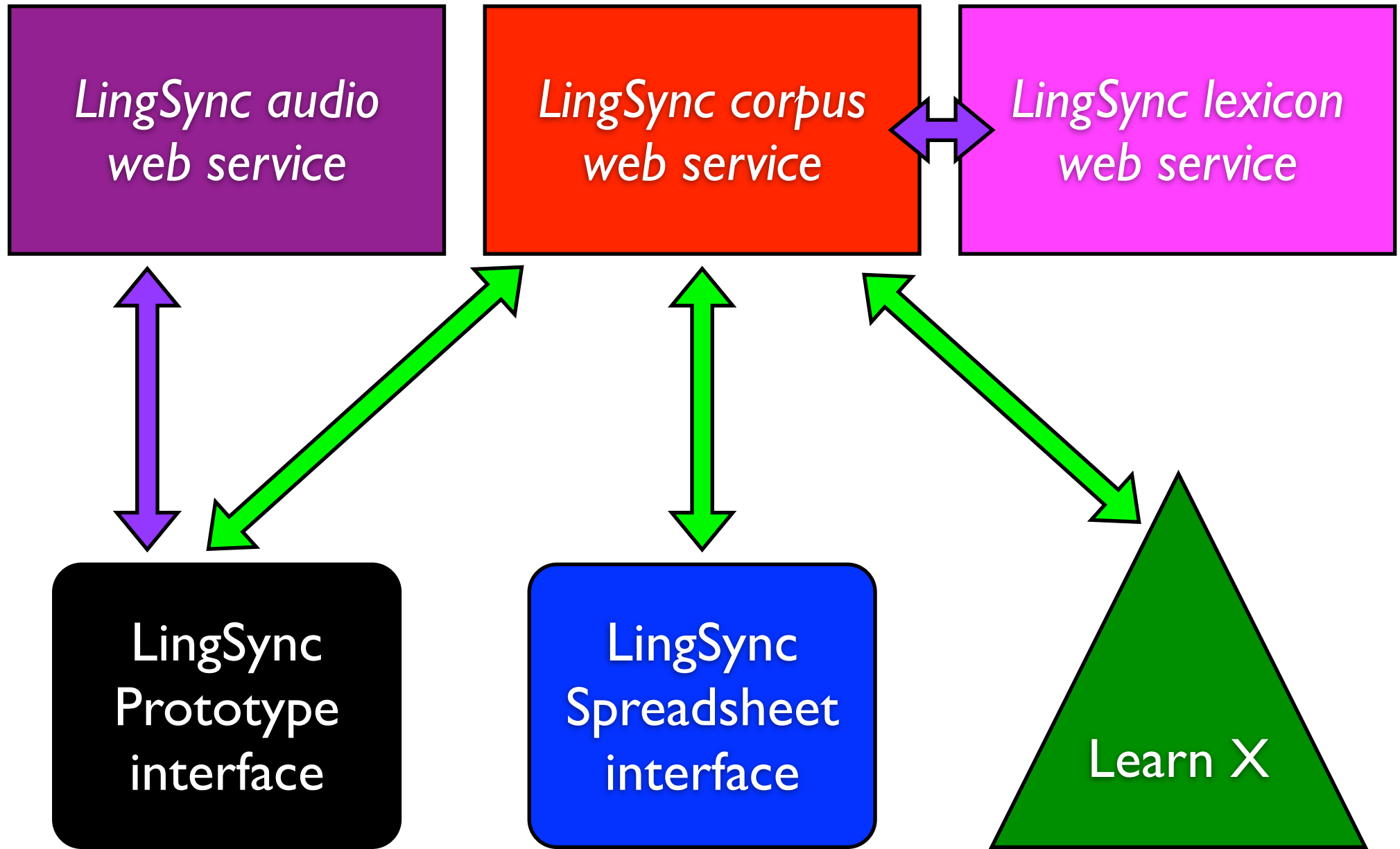
Show 10 | 50 | 100 per page / 1 of 1 pages shown (All)

Default Datalist - Empty 0

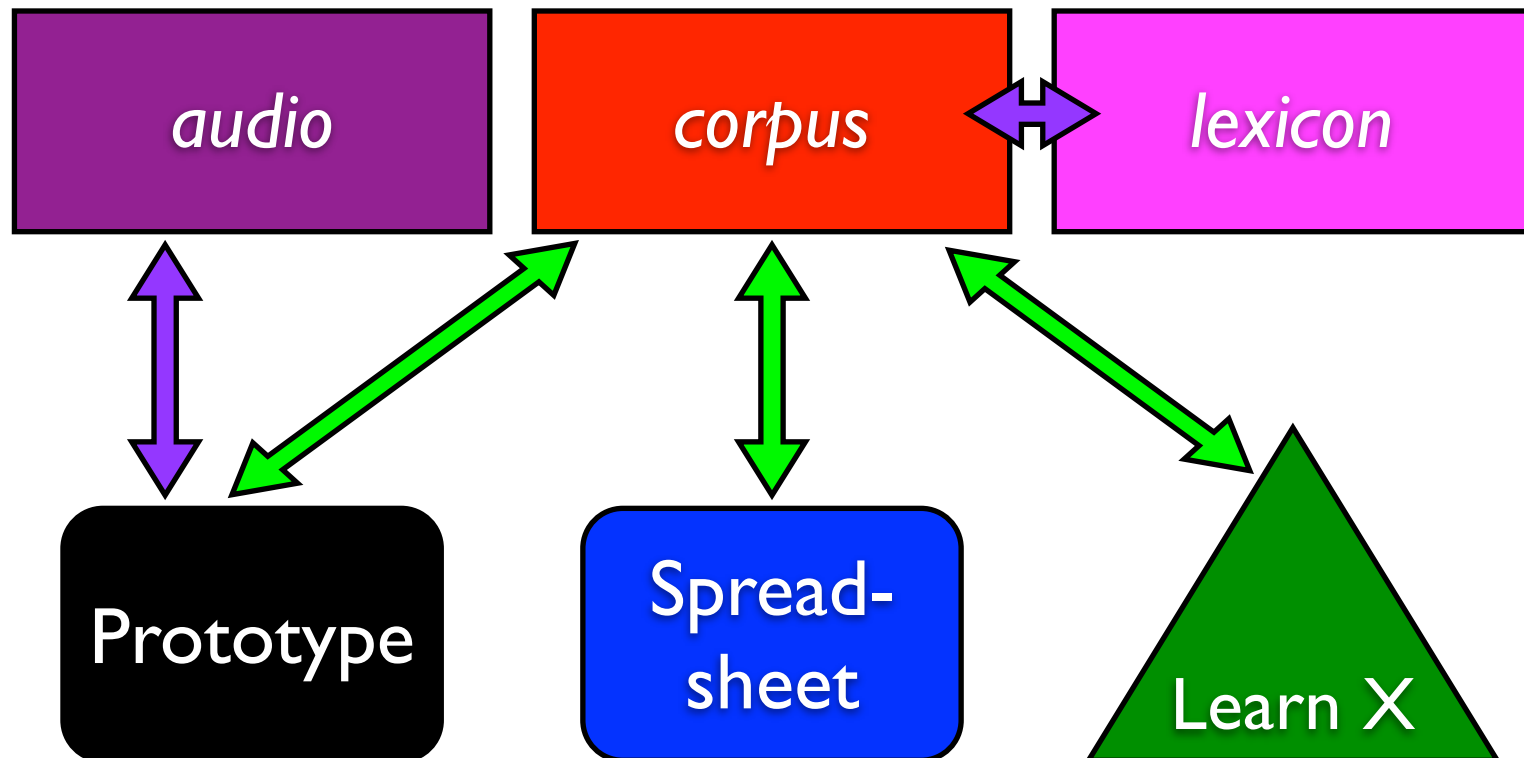
Show 10 | 50 | 100 per page / 1 of 1 pages shown (All)





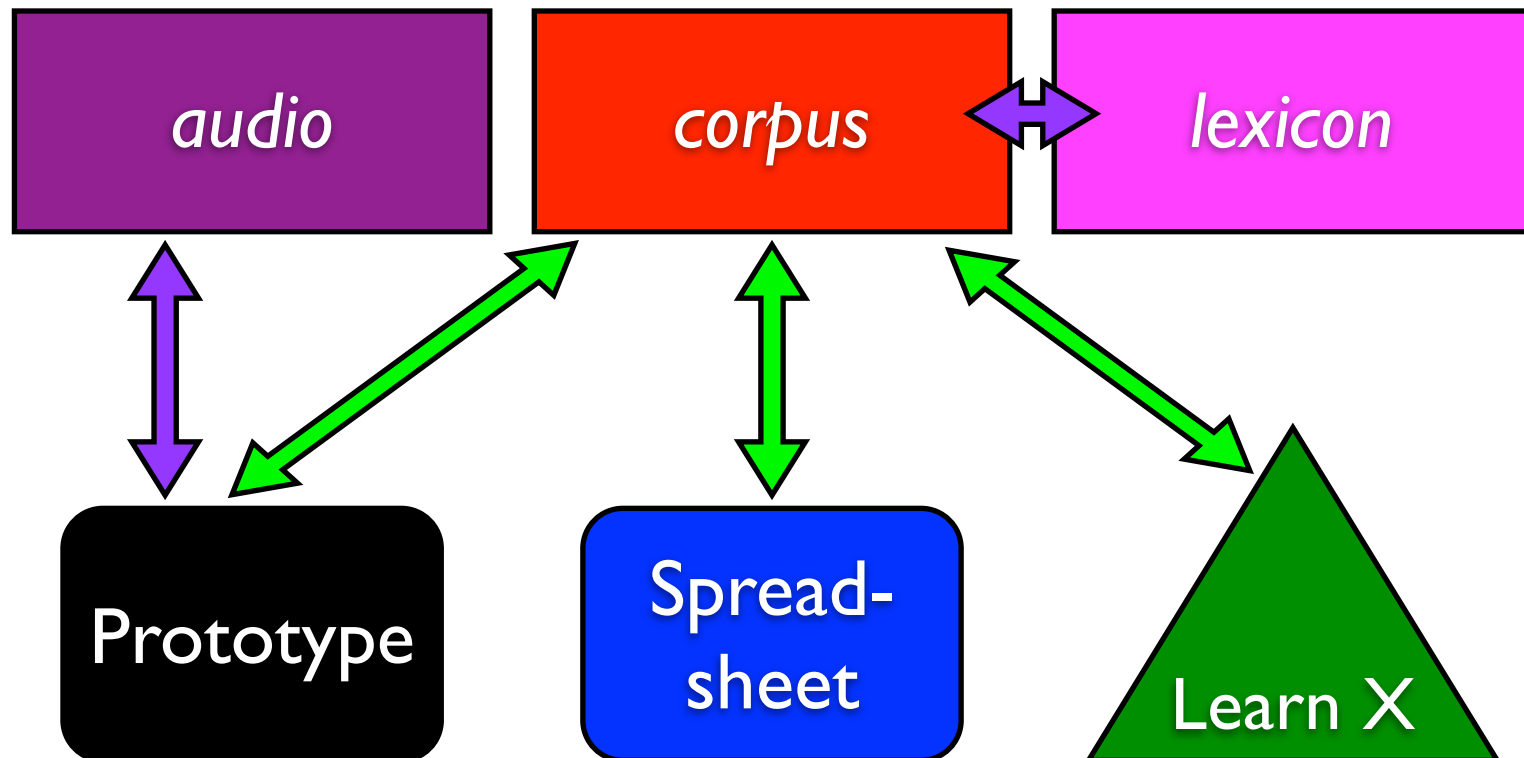


LingSync Architecture



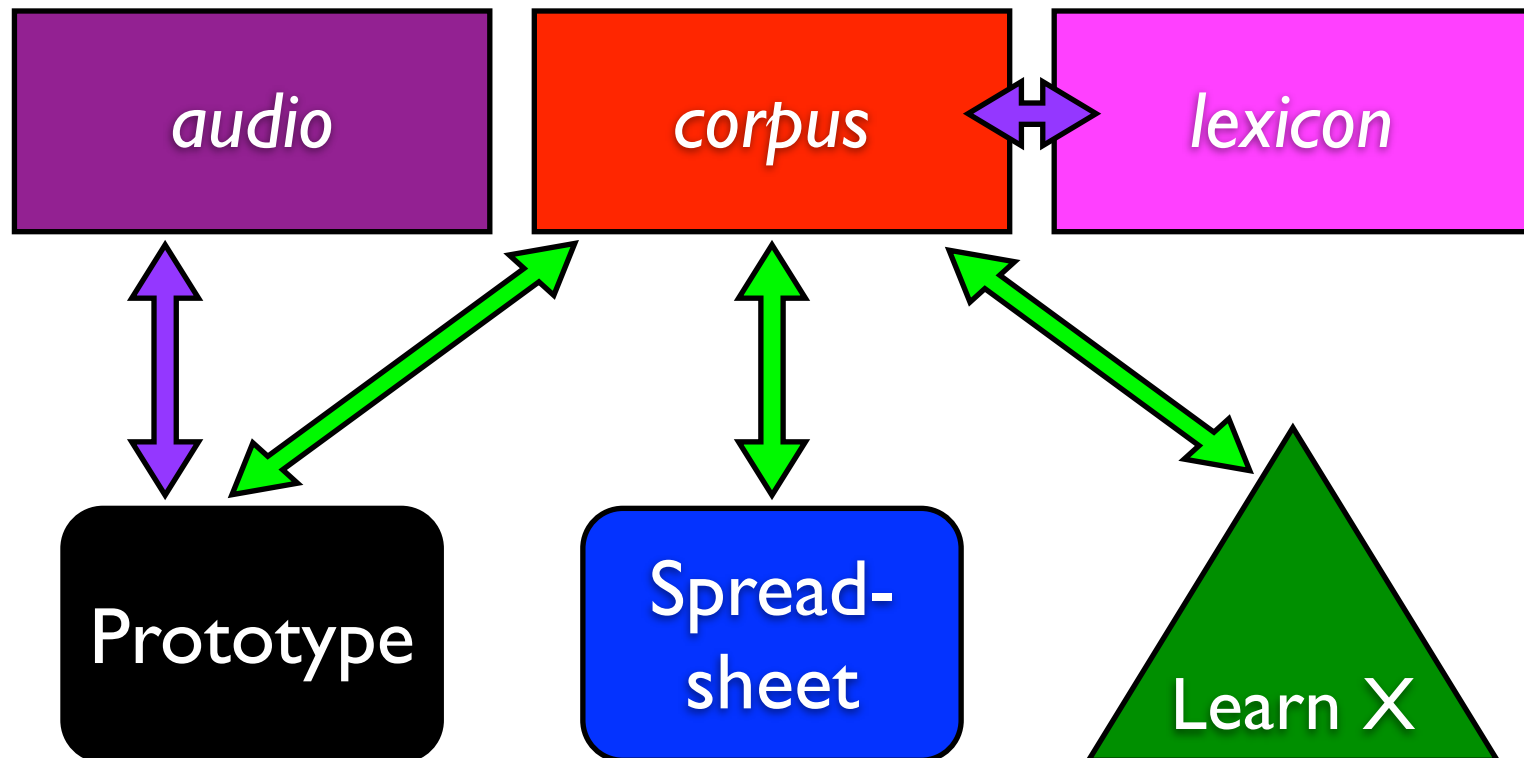
LingSync Architecture

- many platforms



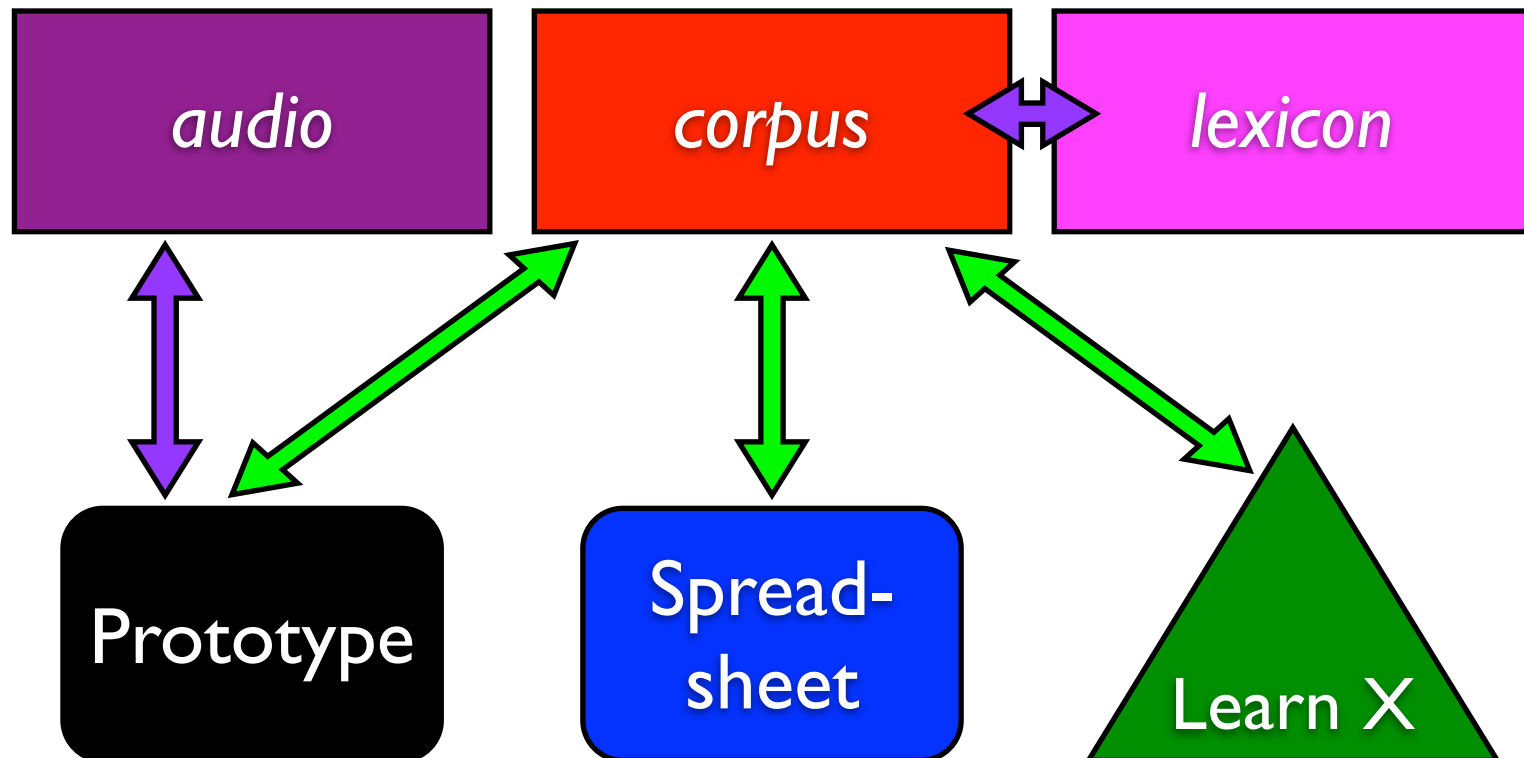
LingSync Architecture

- many platforms
- online/offline



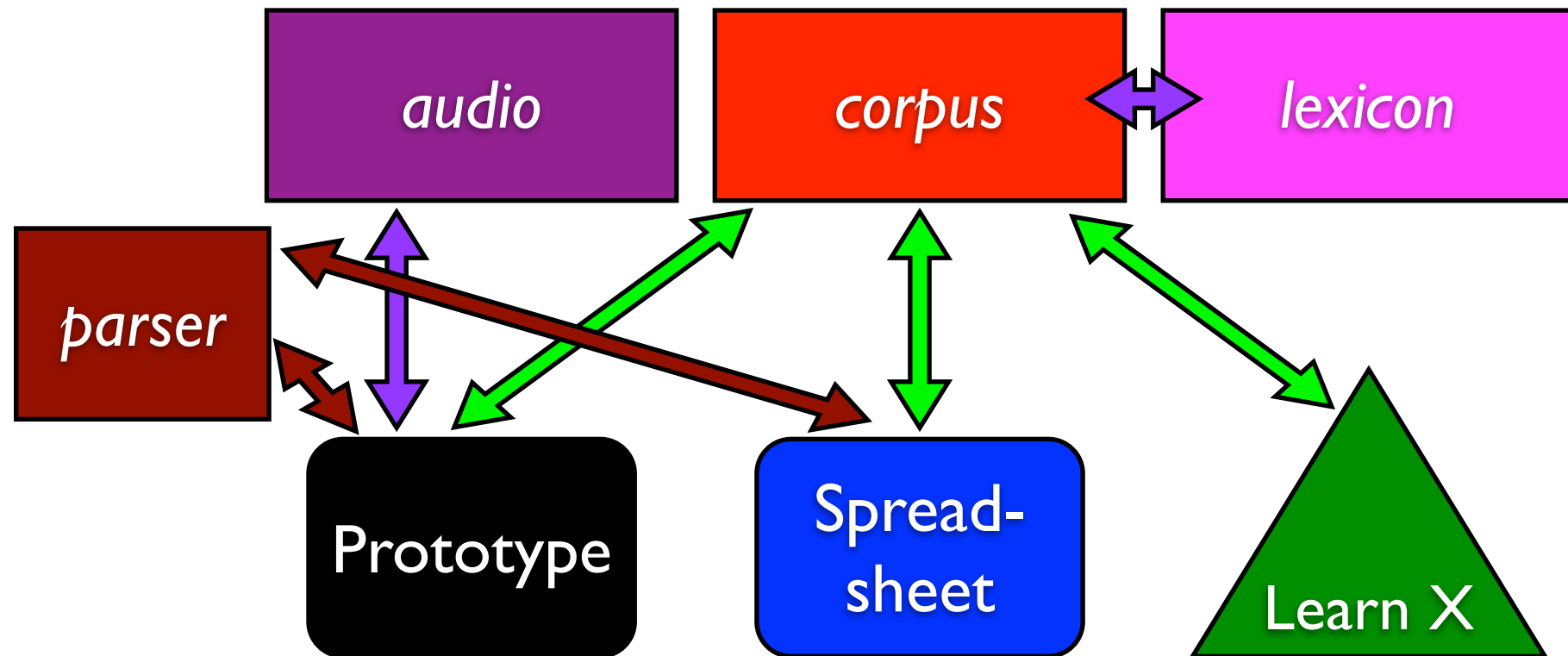
LingSync Architecture

- many platforms
- many purposes
- online/offline



LingSync Architecture

- many platforms
- many purposes
- online/offline
- extensible

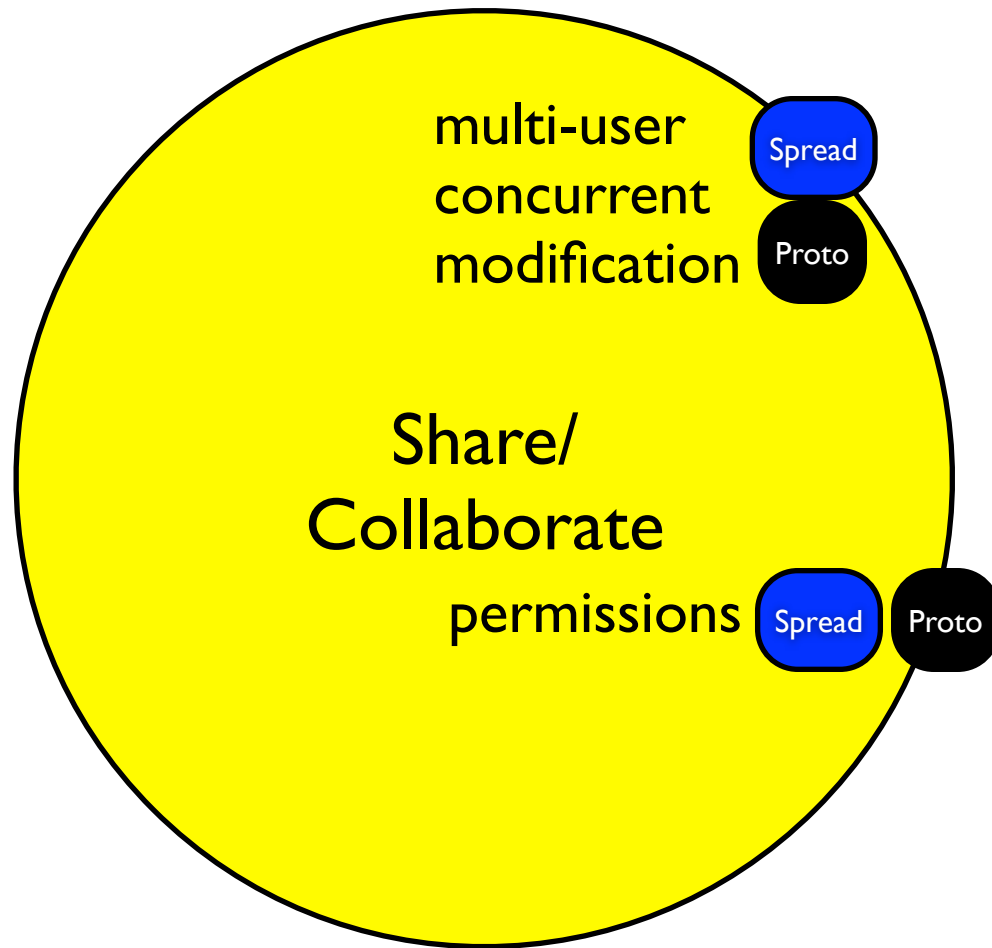


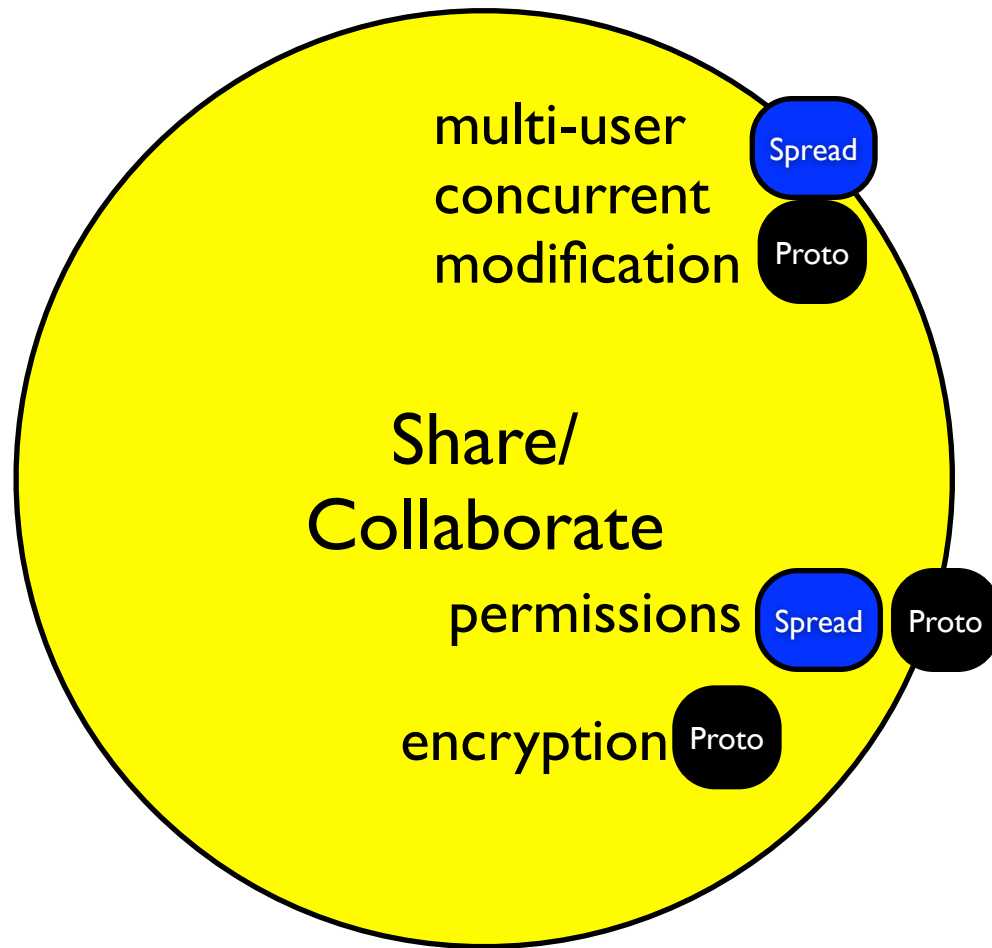
multi-user
concurrent
modification

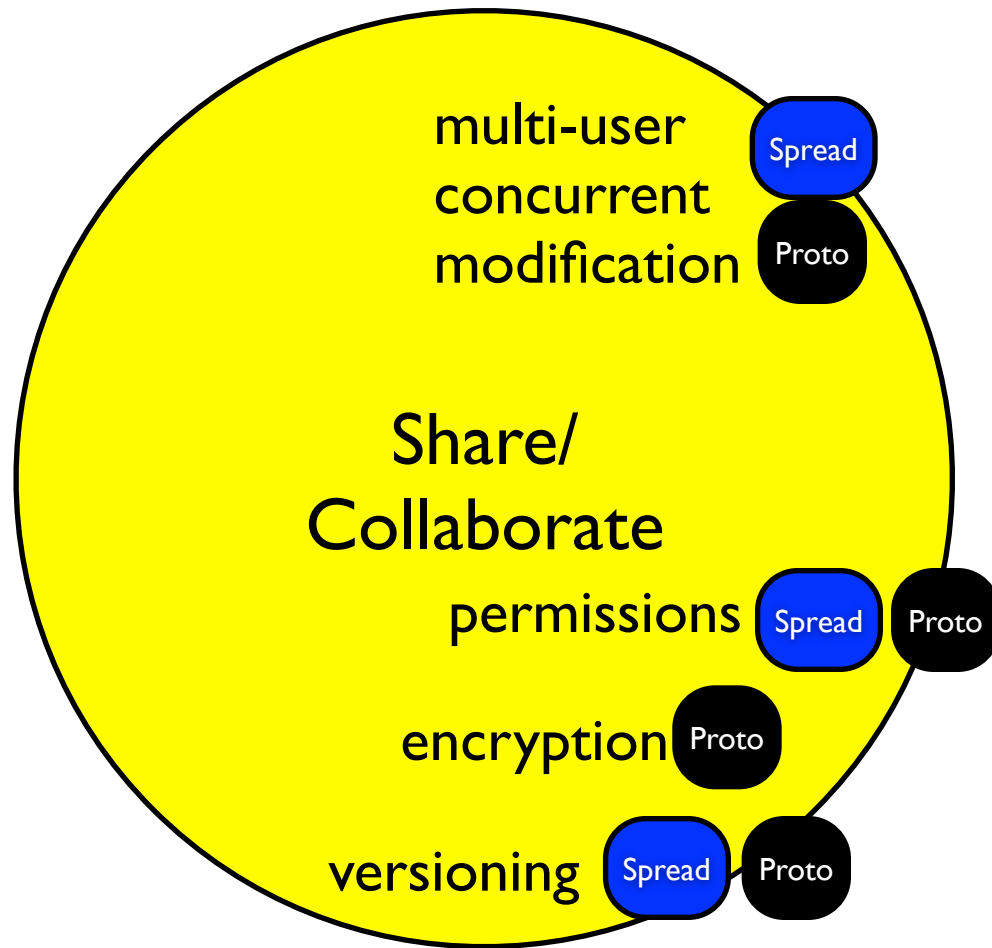
Spread

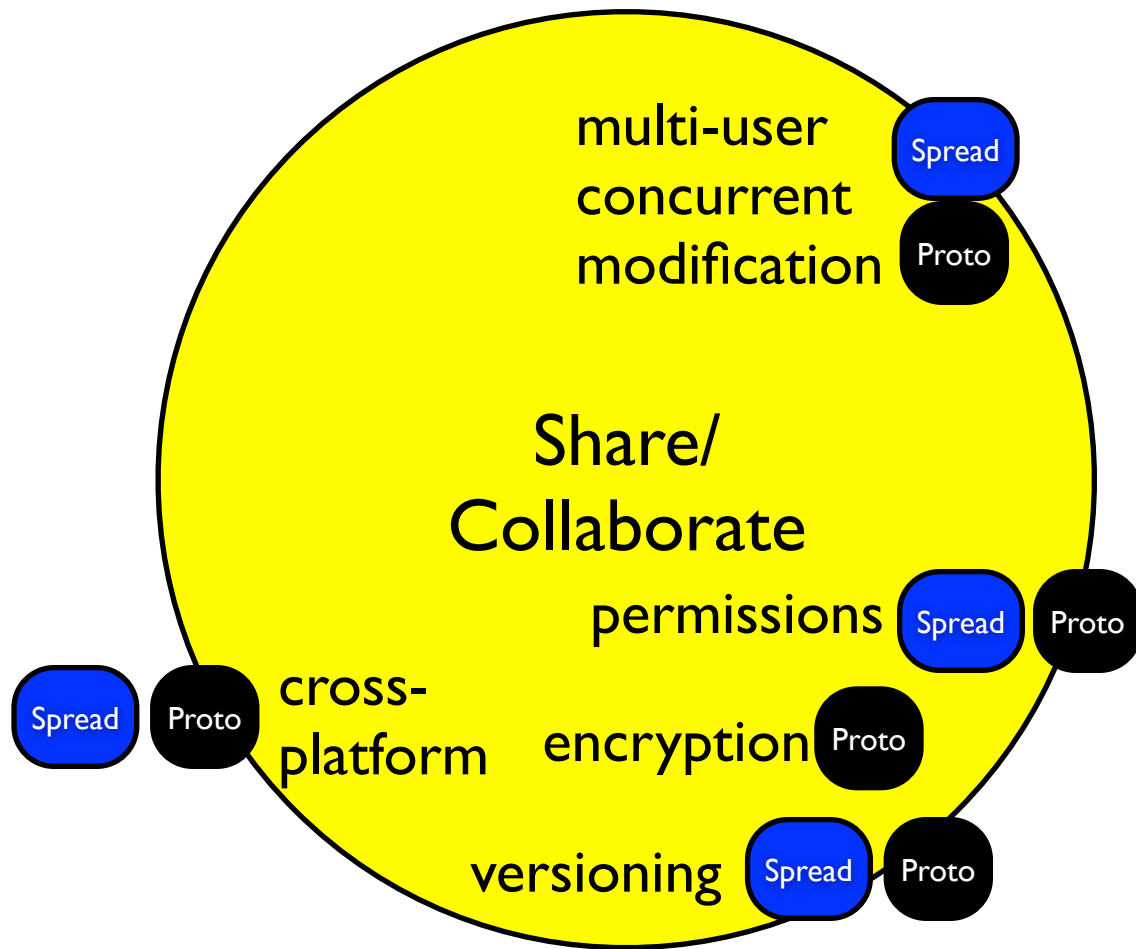
Proto

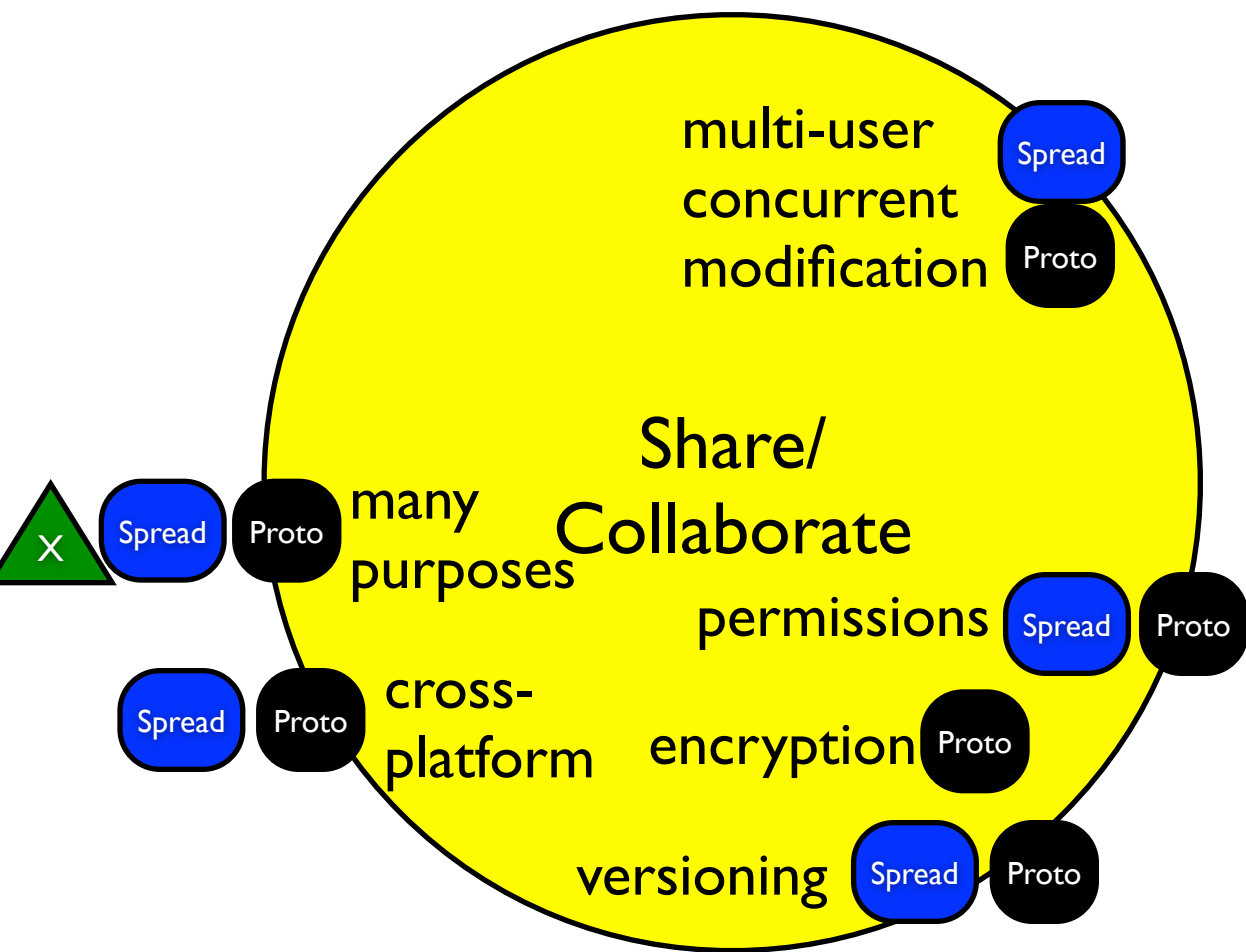
Share/
Collaborate

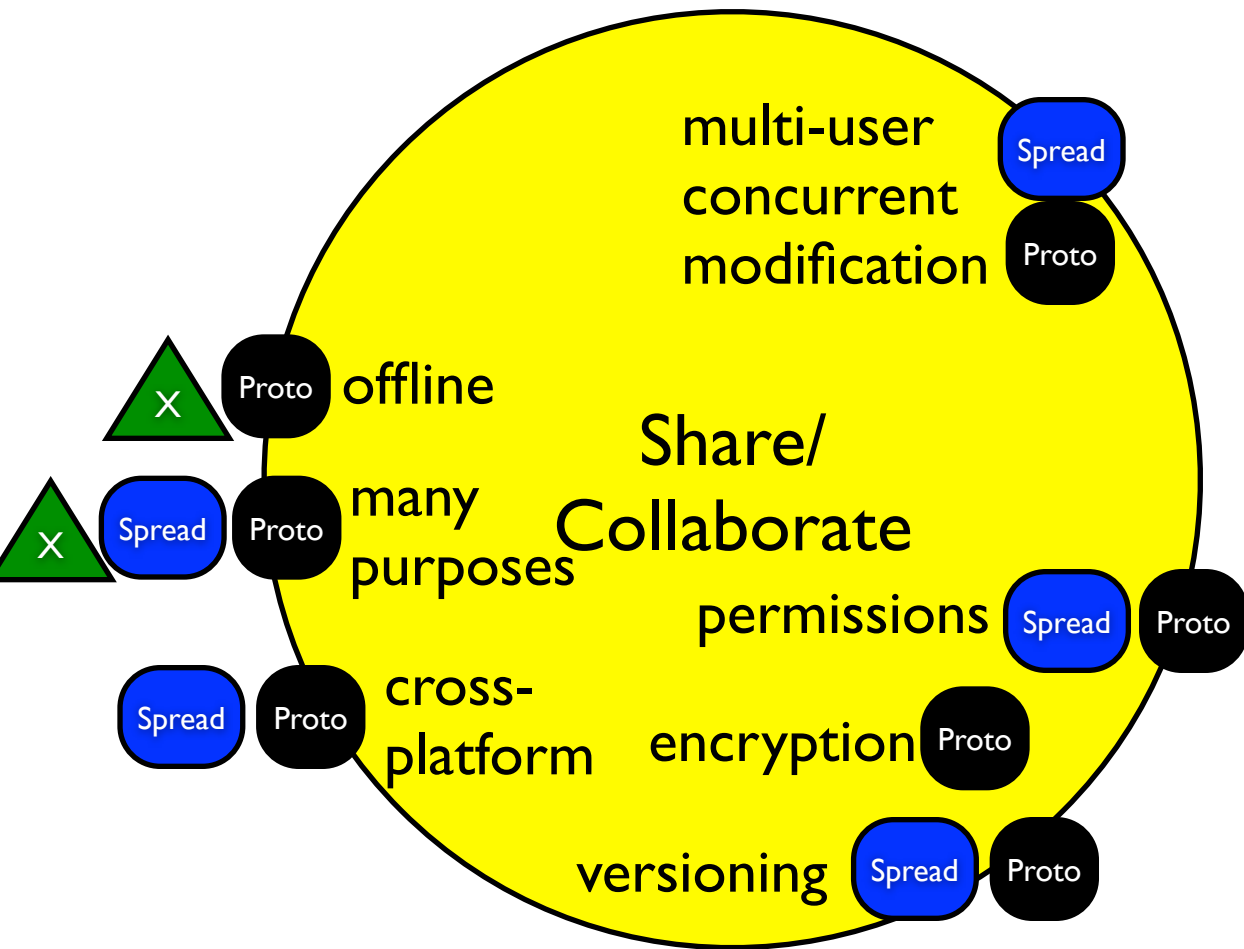


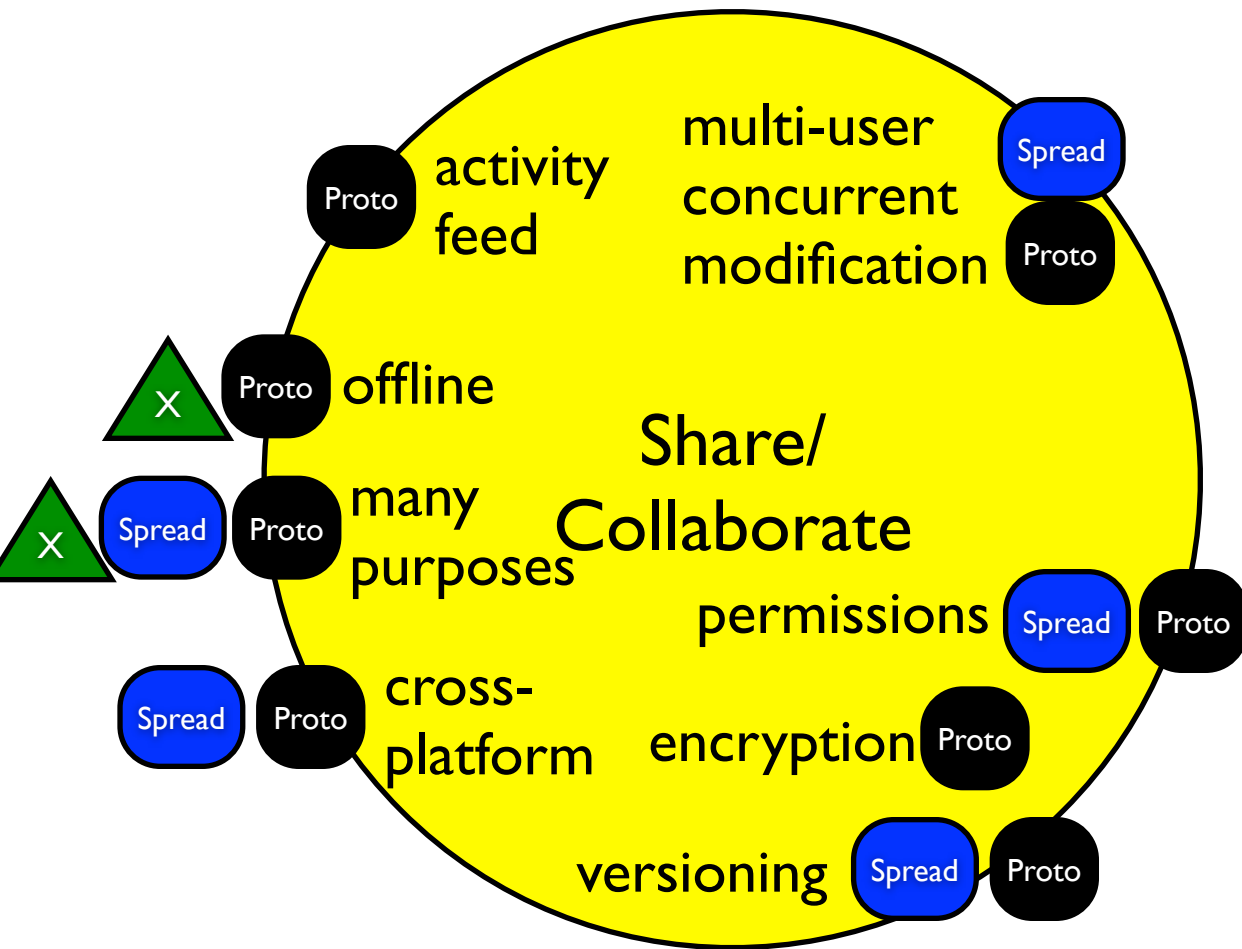














Fieldwork
Features

Audio/Video

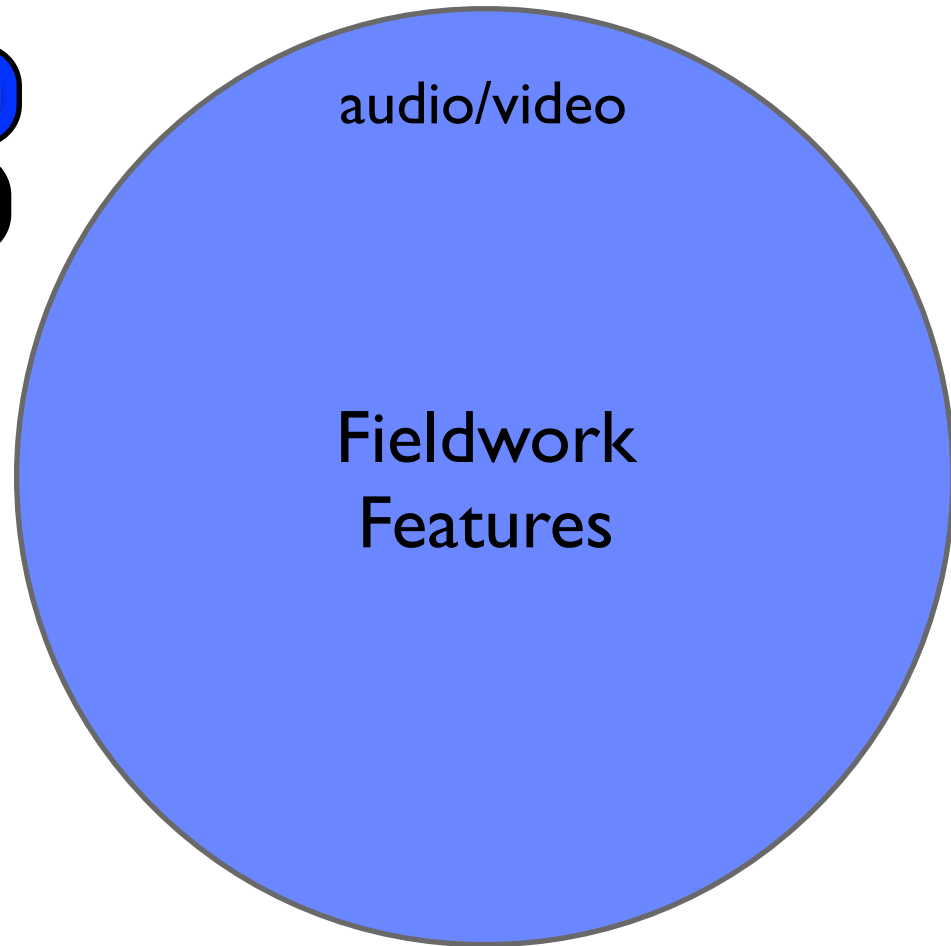
- Play audio/video right next to the utterance

Spread

Proto

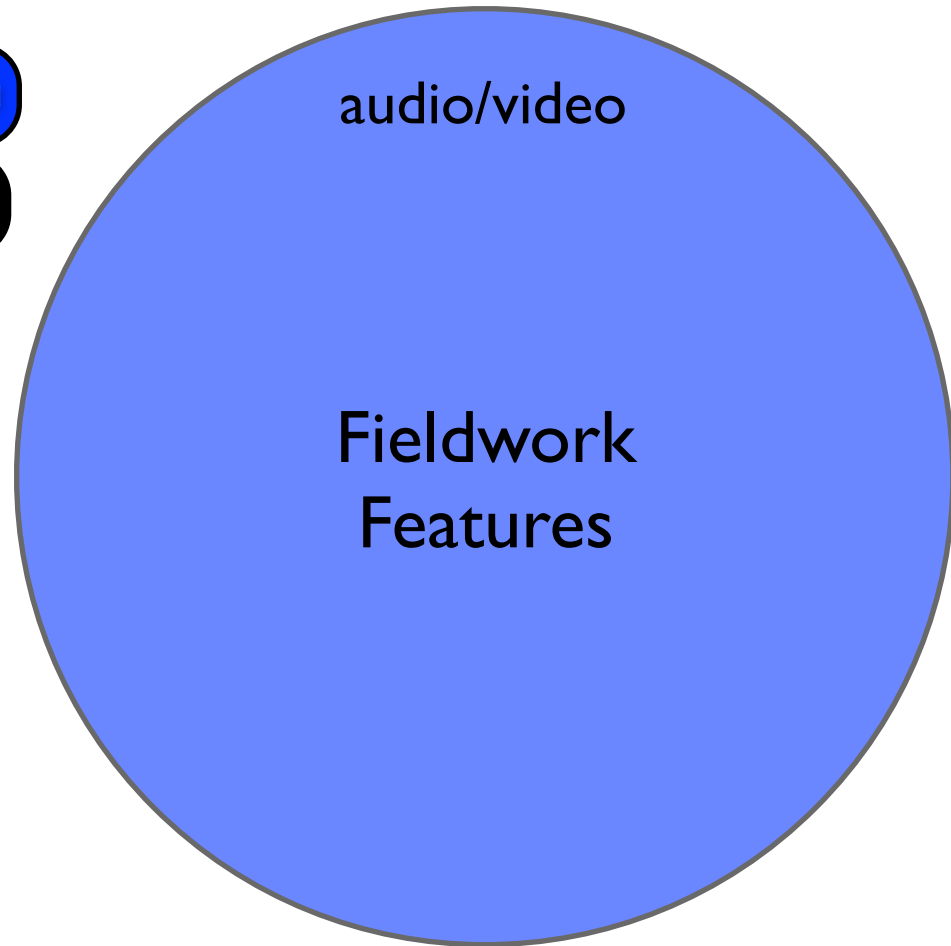
- Record audio directly into the app

Spread



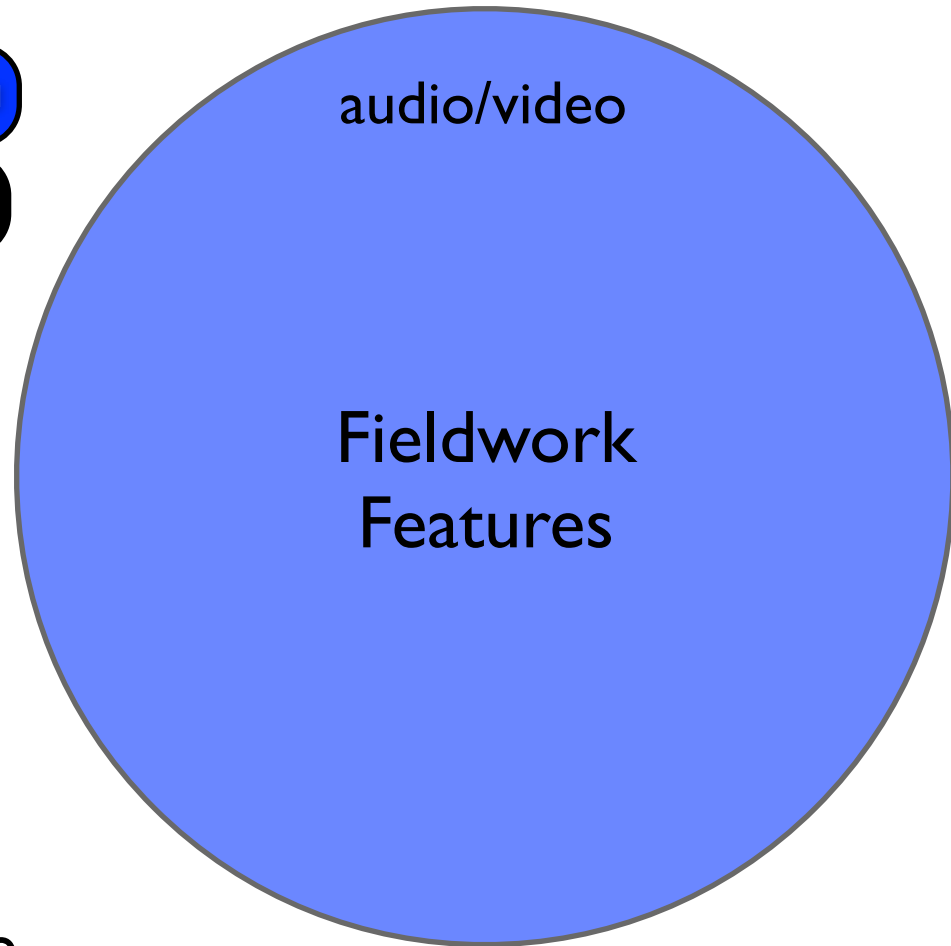
Audio/Video

- Play audio/video right next to the utterance Spread
Proto
- Record audio Spread
directly into the app
- Drag-and-drop audio/video files into your data Proto



Audio/Video

- Play audio/video right next to the utterance Spread
Proto
- Record audio Spread
directly into the app
- Drag-and-drop audio/video files into your data Proto
- Link to audio/video on the web, e.g., YouTube Proto

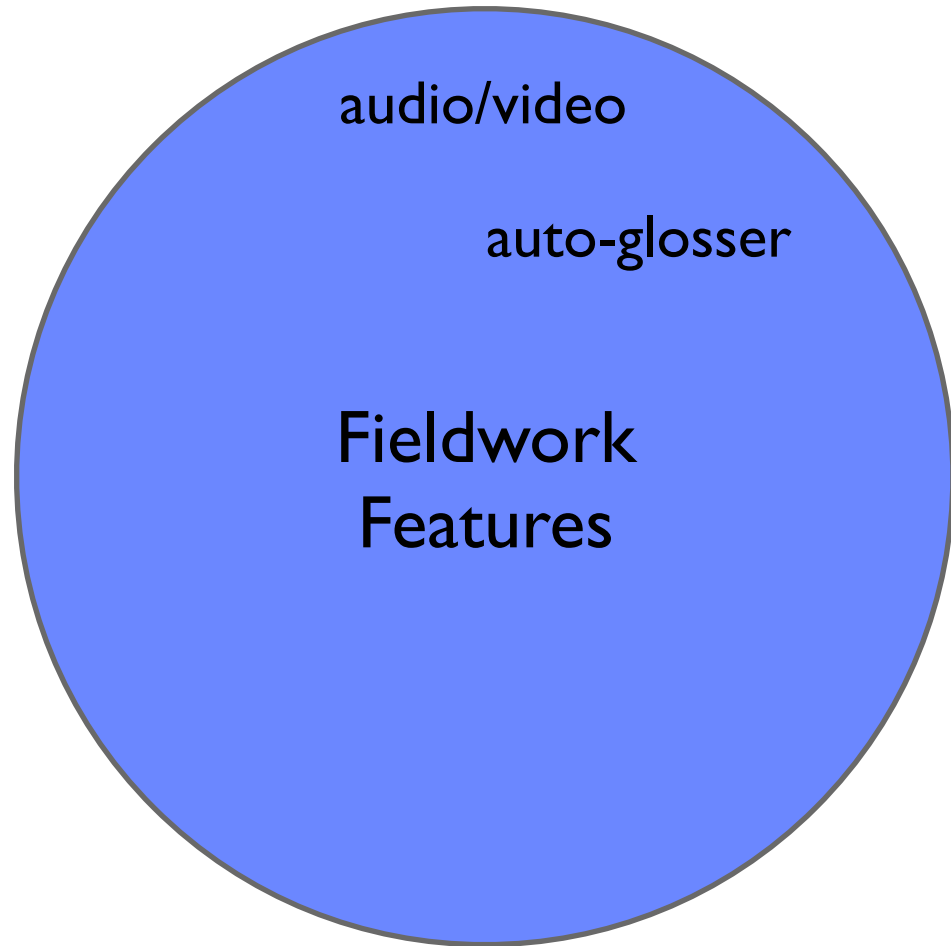


Semi-automatic Glosser

Spread

Proto

- Predicts morpheme breaks and glosses

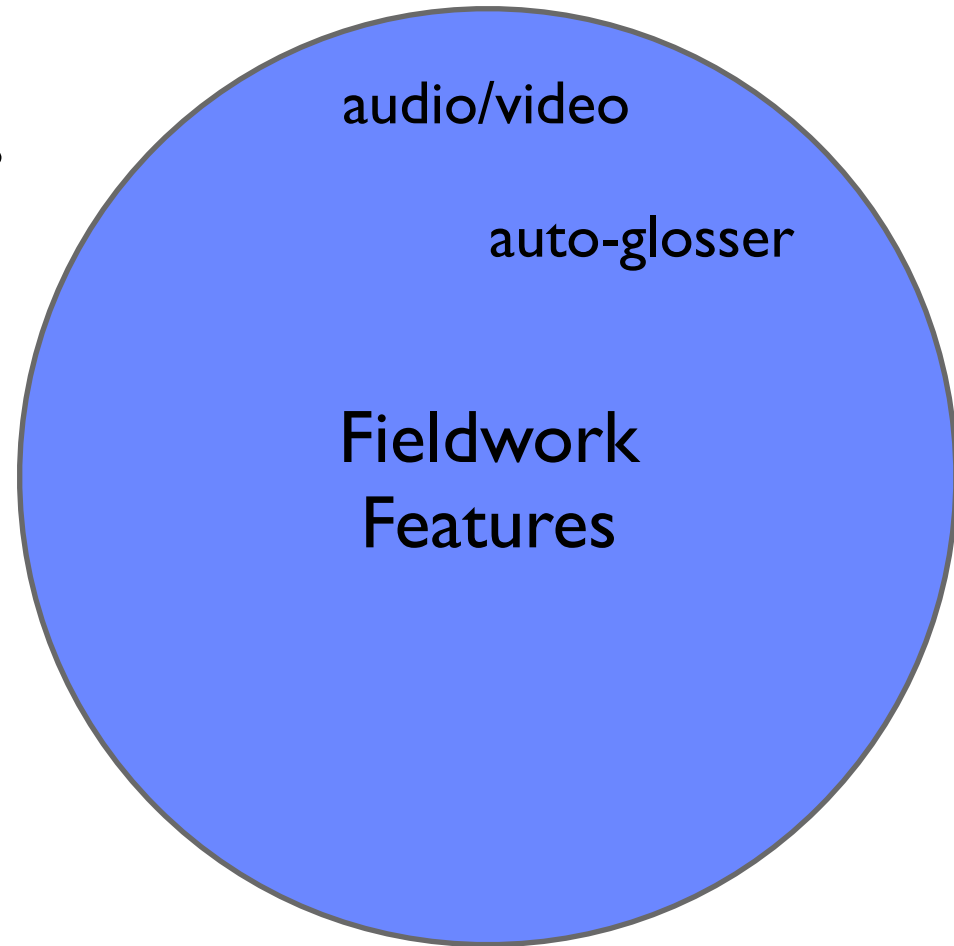


Semi-automatic Glosser

Spread

Proto

- Predicts morpheme breaks and glosses
- Starts learning from your data right away

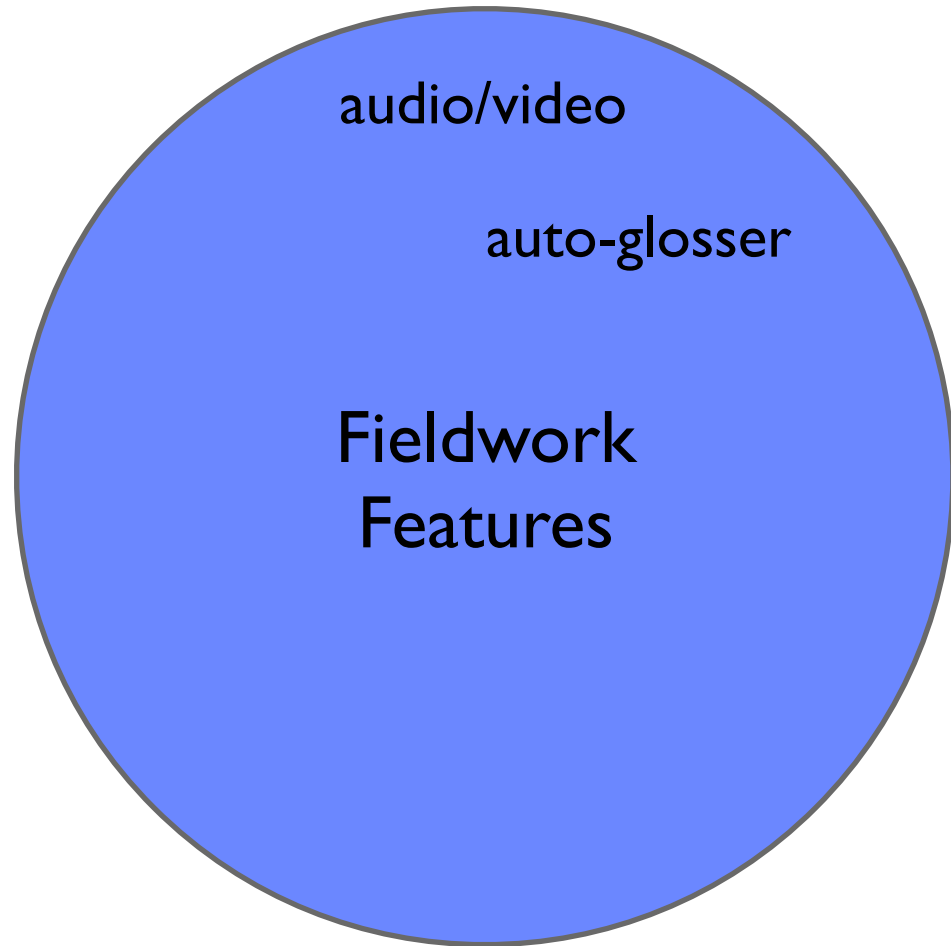


Semi-automatic Glosser

Spread

Proto

- Predicts morpheme breaks and glosses
- Starts learning from your data right away
- Context-sensitive

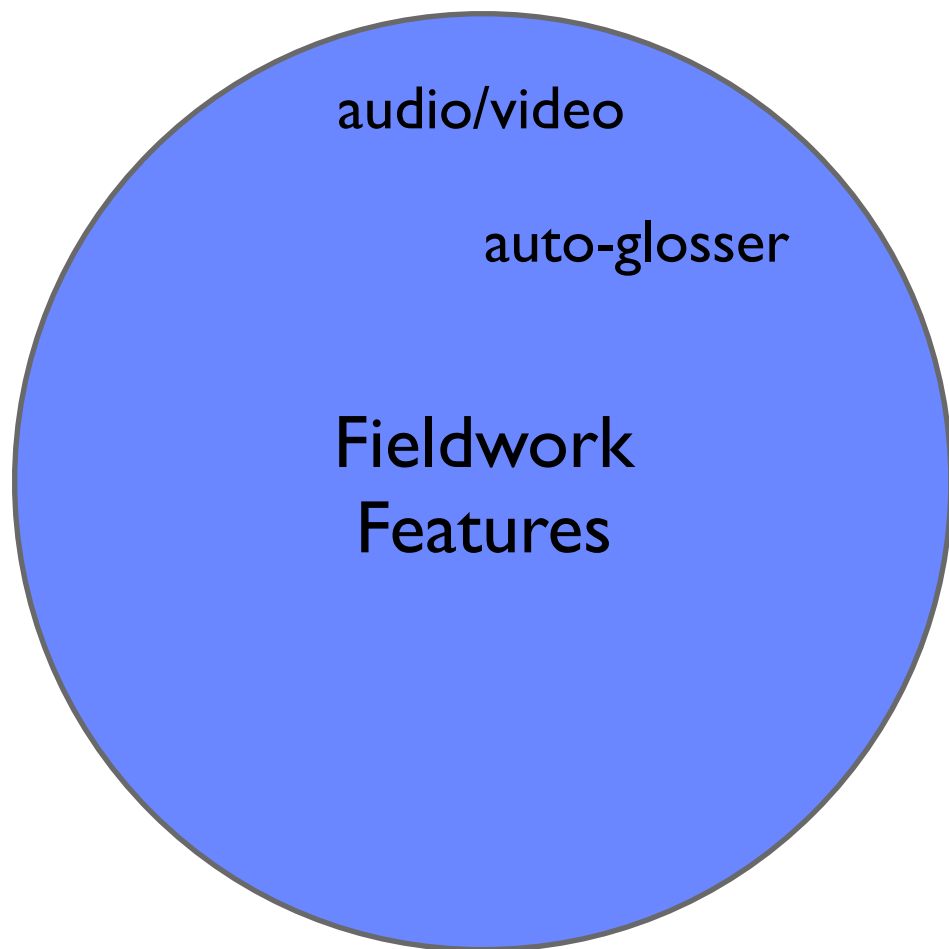


Semi-automatic Glosser

Spread

Proto

- Predicts morpheme breaks and glosses
- Starts learning from your data right away
- Context-sensitive
- Theory non-specific: you define your glosses

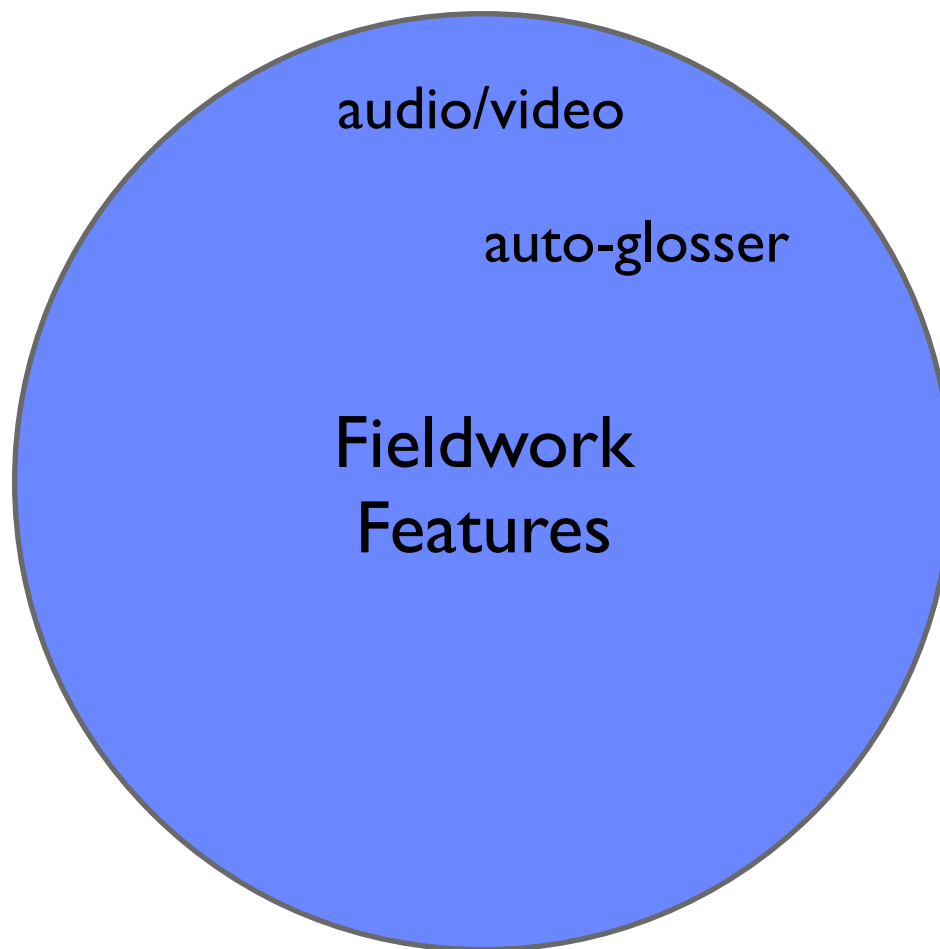


Semi-automatic Glosser

Spread

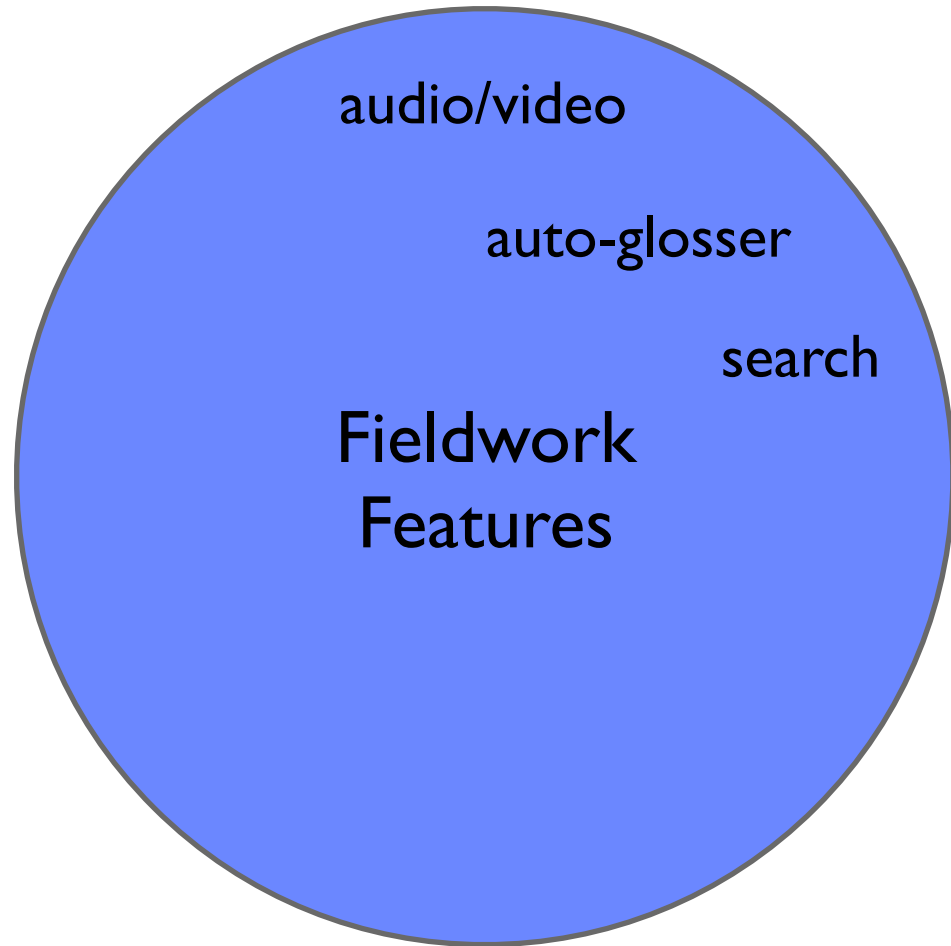
Proto

- Predicts morpheme breaks and glosses
- Starts learning from your data right away
- Context-sensitive
- Theory non-specific: you define your glosses
- Uses Leipzig Glossing Rules



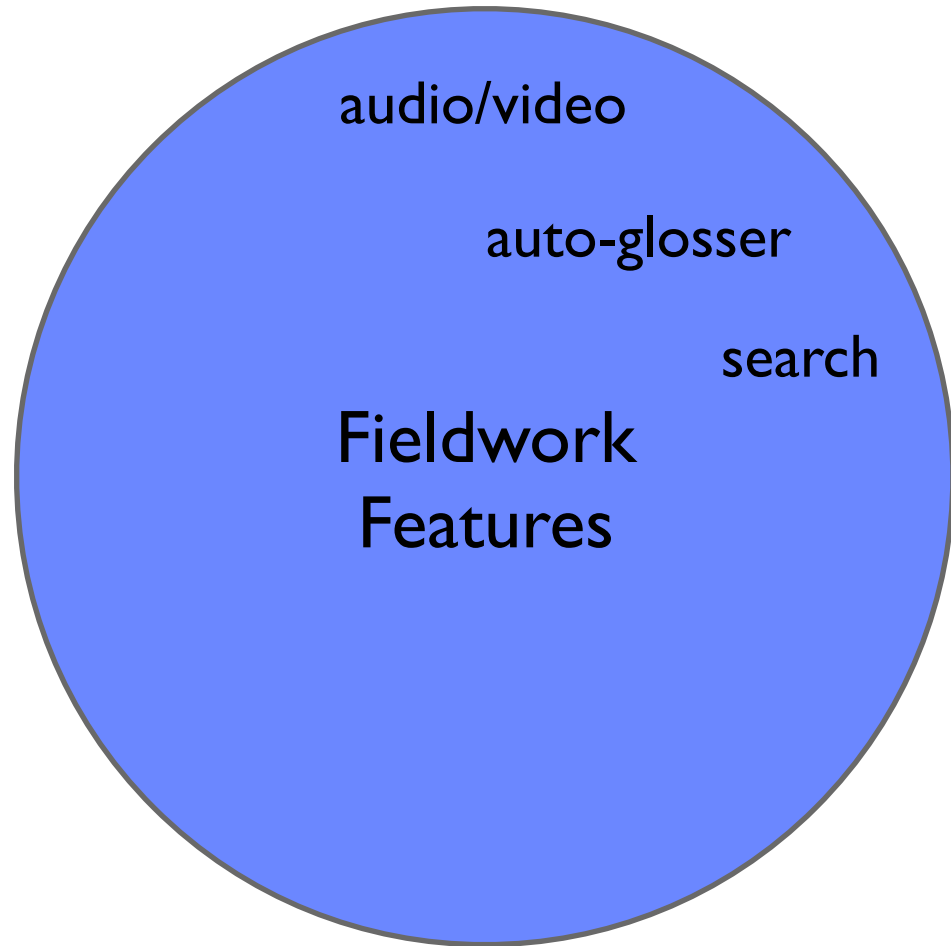
Search

- Basic search Spread



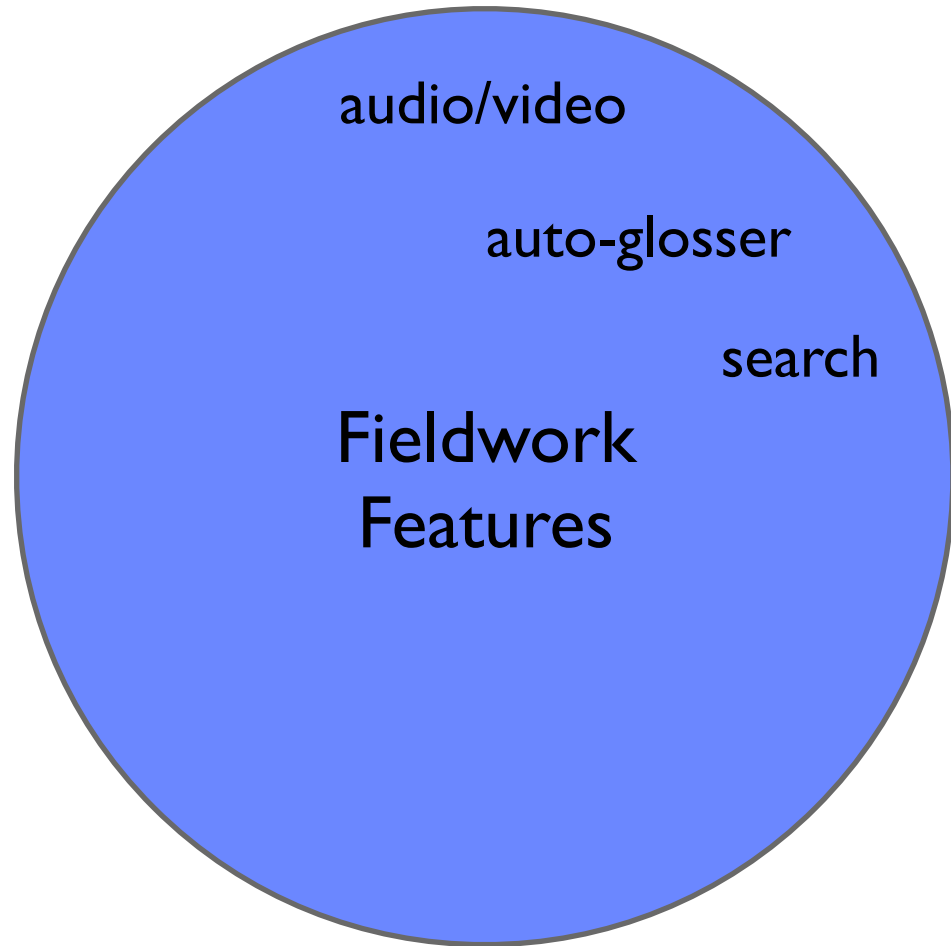
Search

- Basic search Spread
- Advanced search Proto



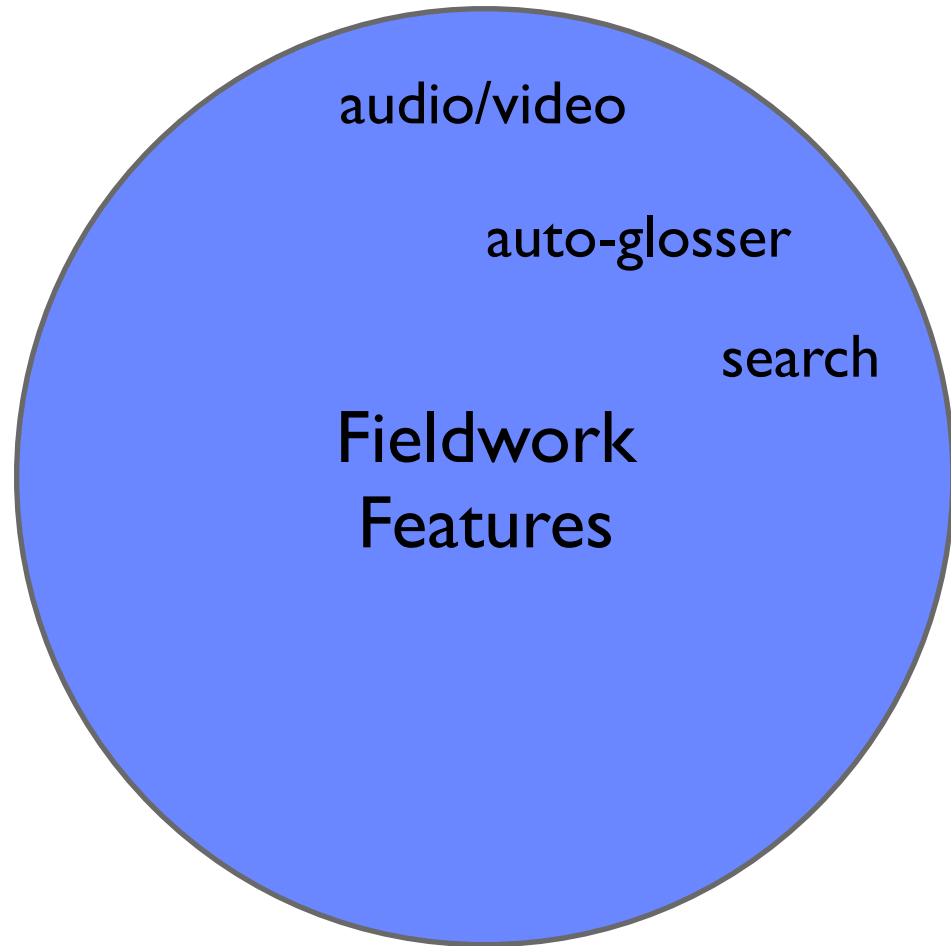
Search

- Basic search Spread
- Advanced search Proto
- Save searches as data lists Proto



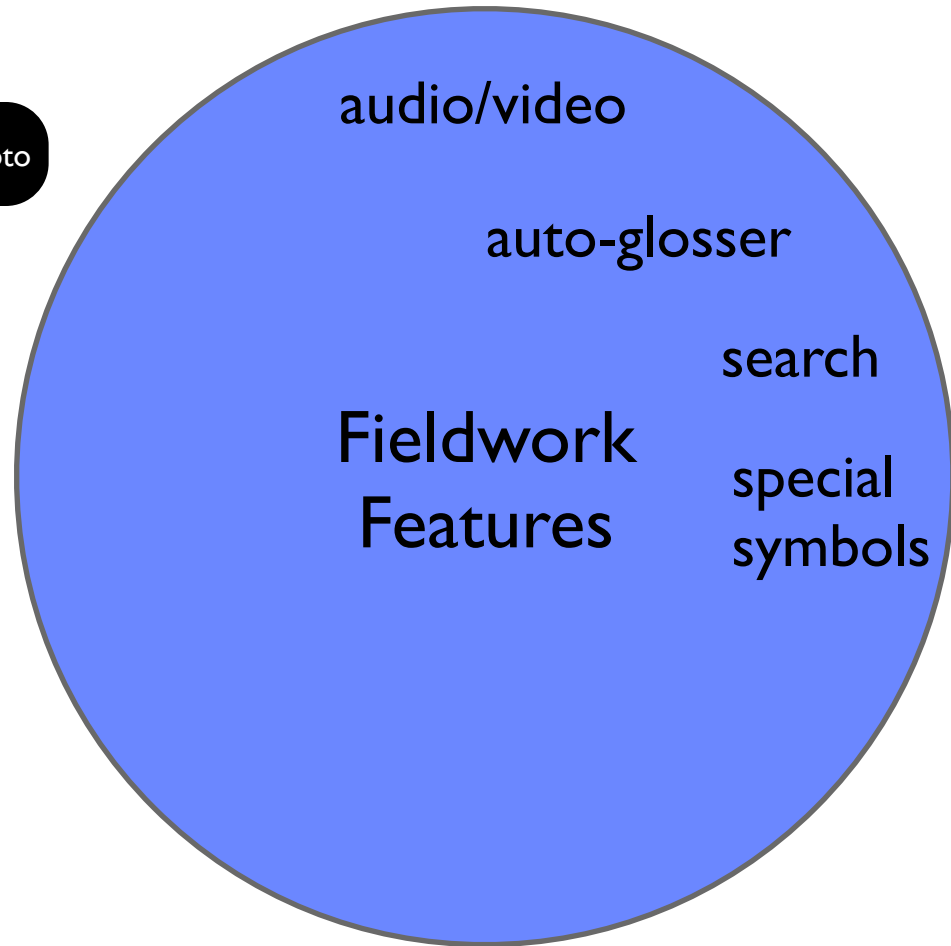
Search

- Basic search Spread
- Advanced search Proto
- Save searches as data lists Proto
- Search over user-defined tags Spread Proto



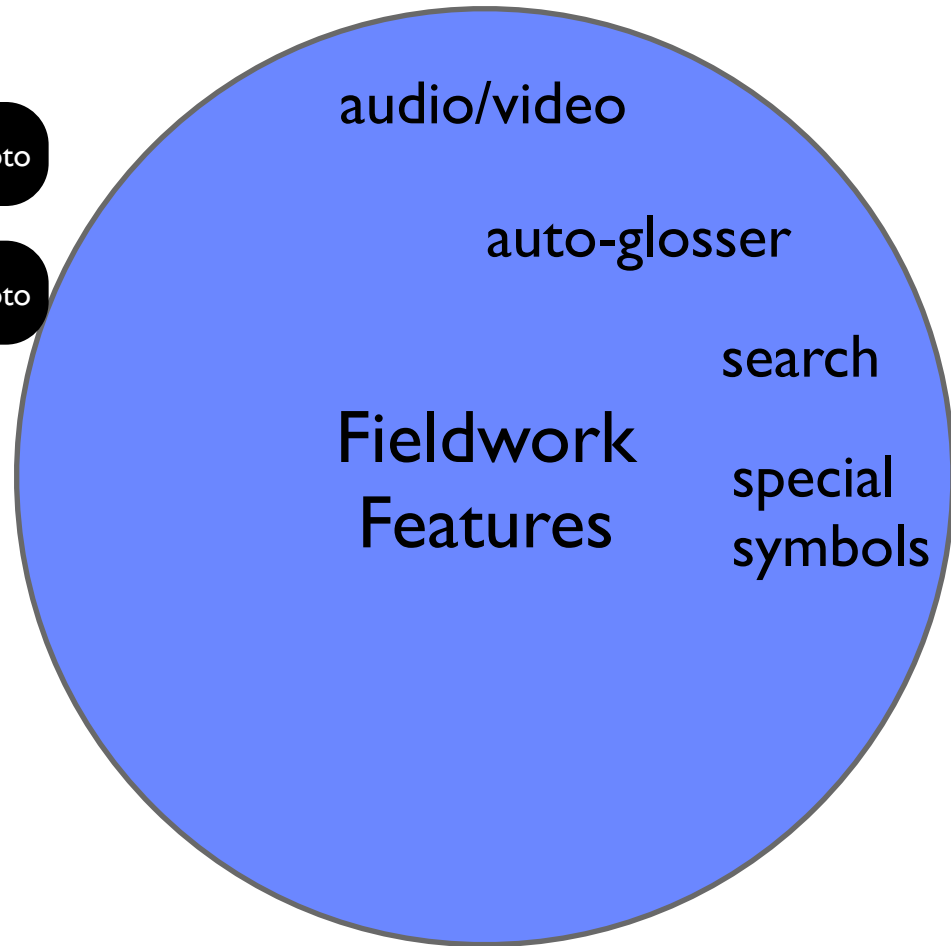
Special Symbols

- Unicode throughout Spread Proto



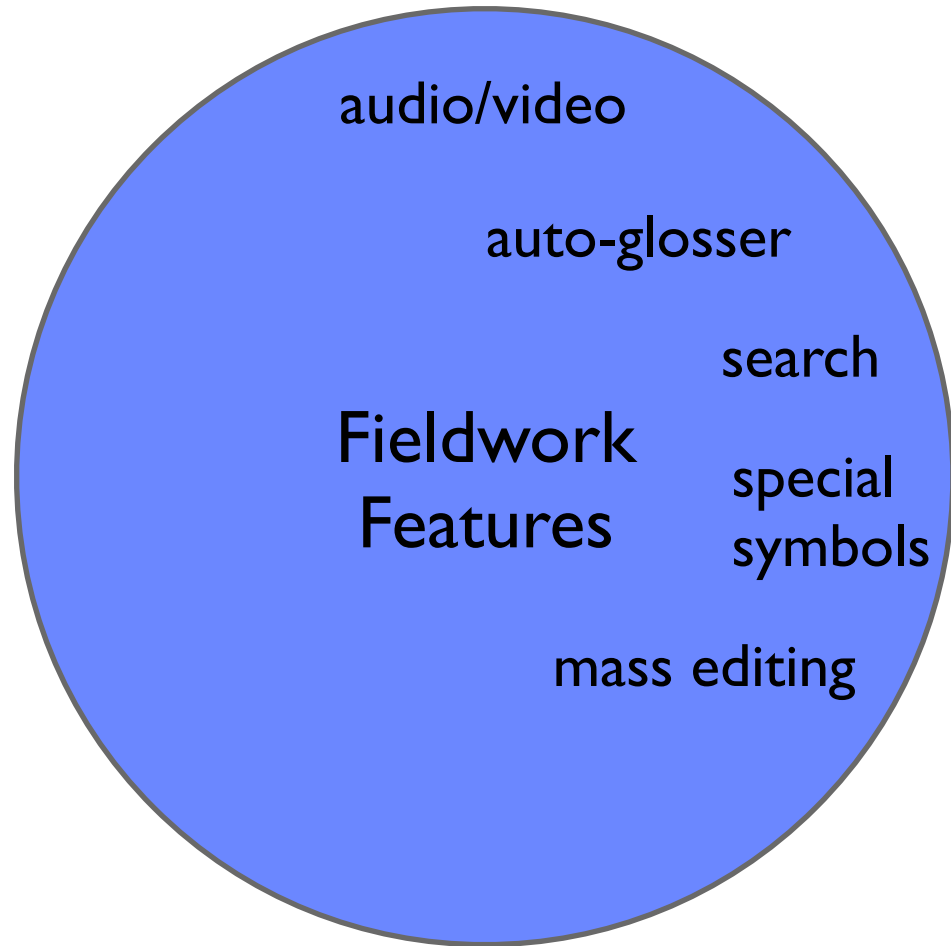
Special Symbols

- Unicode throughout Spread Proto
- Drag-and-drop IPA entry Proto



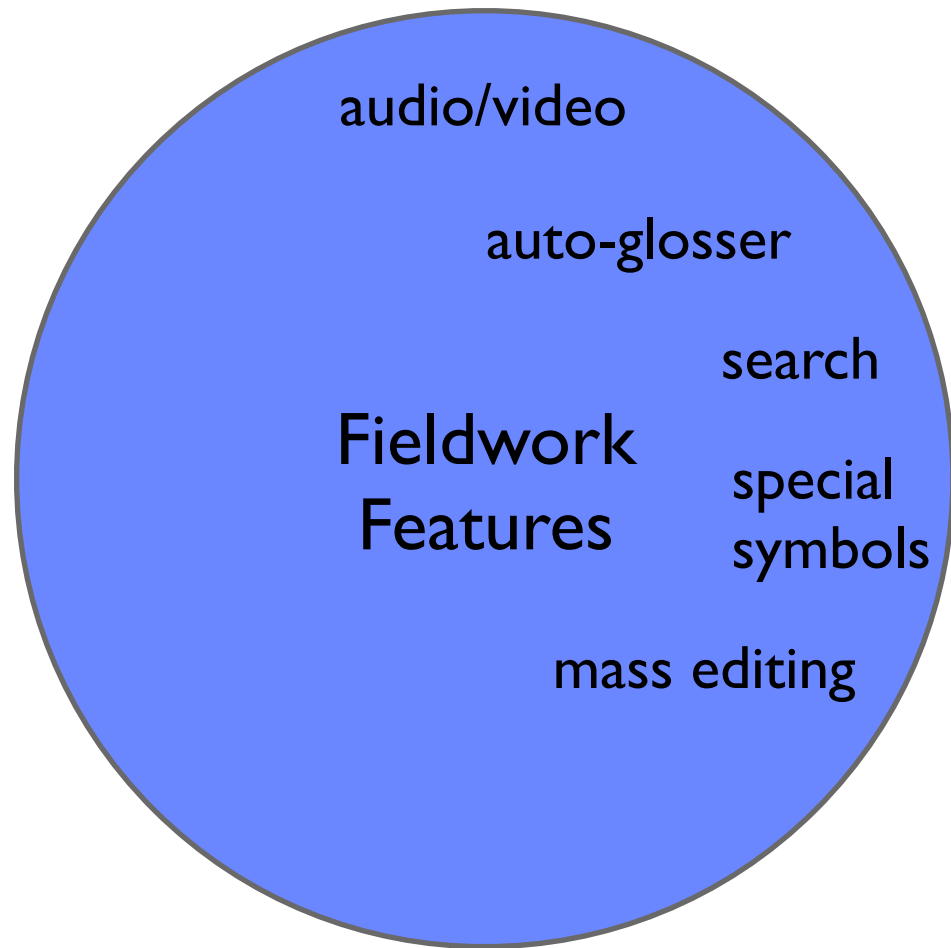
Mass Editing Proto

- Write scripts (bots) to change data *en masse* (JavaScript)
- E.g., re-analyze a tense morpheme as aspectual



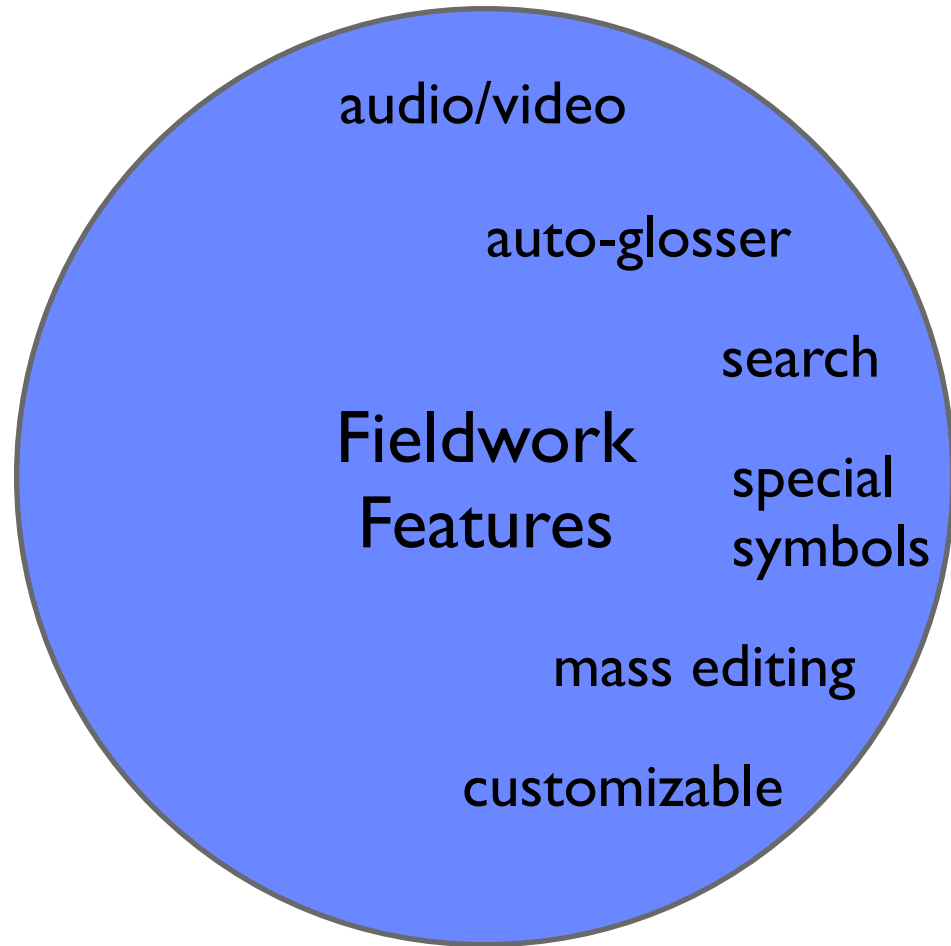
Mass Editing Proto

- Write scripts (bots) to change data *en masse* (JavaScript)
- E.g., re-analyze a tense morpheme as aspectual
- Scripts/bots can also archive, run statistics, and create visualizations



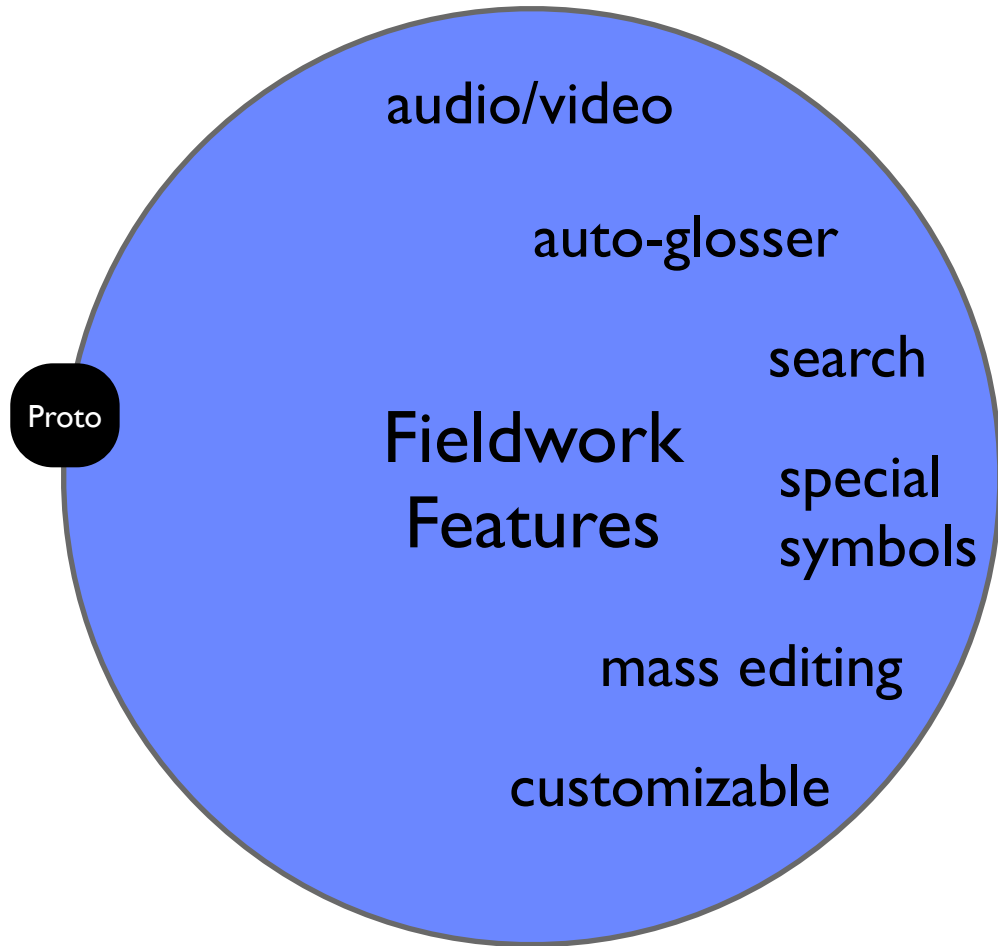
Customizable

- Commentable Spread Proto



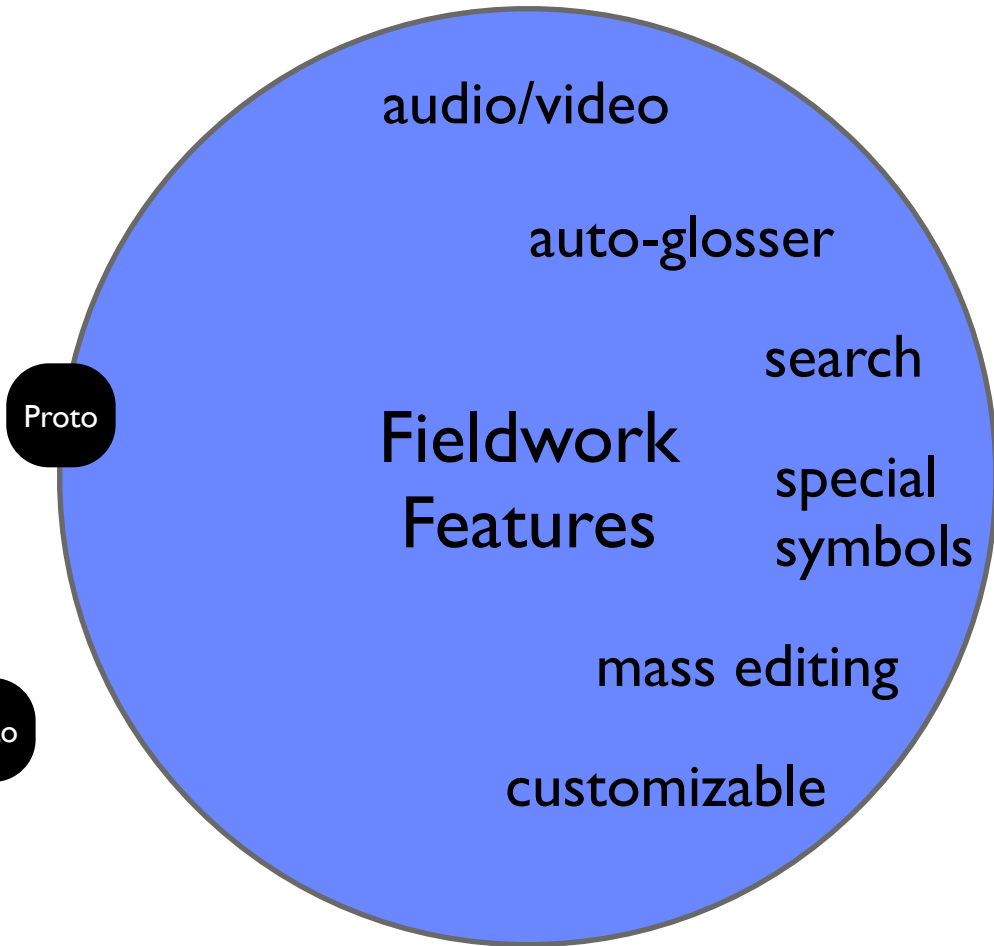
Customizable

- Commentable Spread Proto
- Alterable data structure Proto



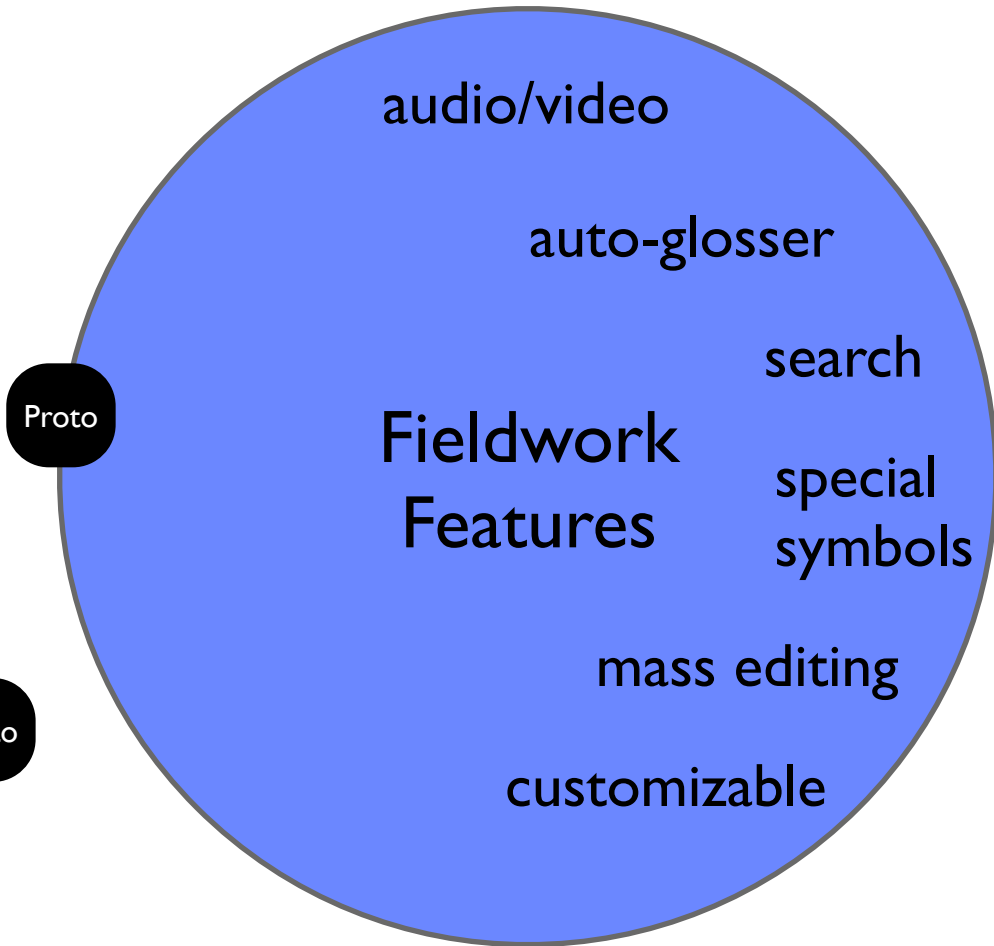
Customizable

- Commentable Spread Proto
- Alterable data structure Proto
- Learns which fields you use most and displays these by default Proto



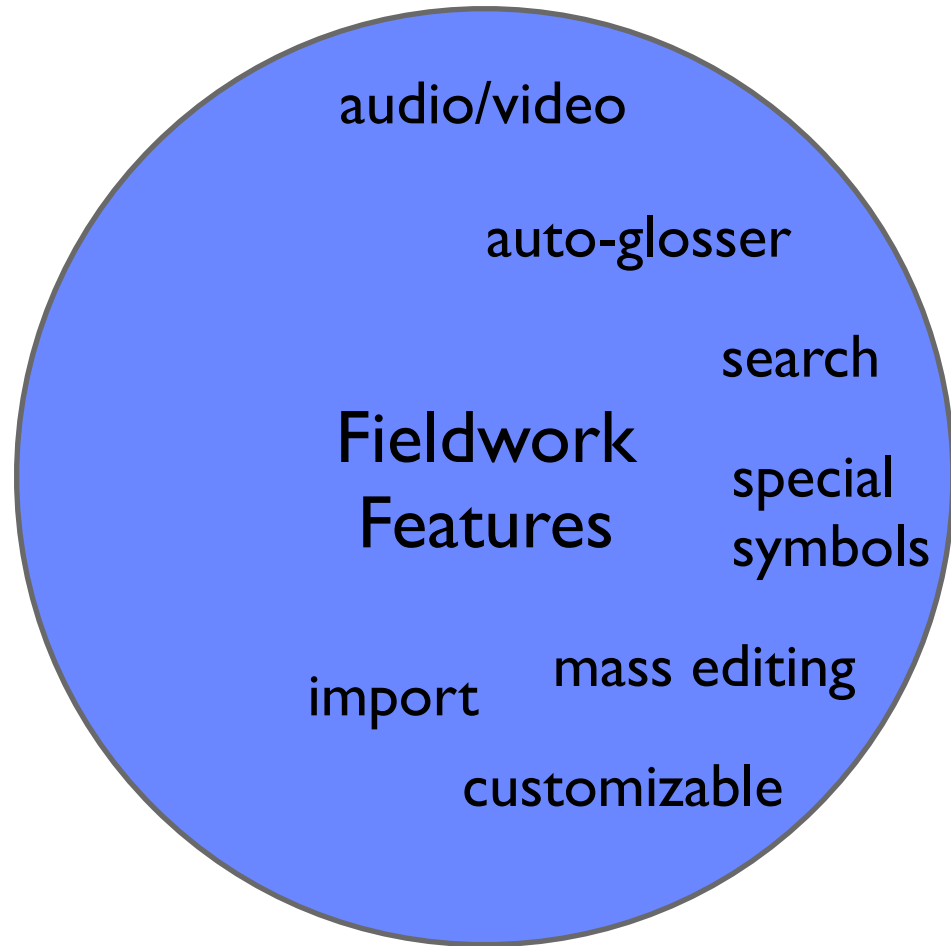
Customizable

- Commentable Spread Proto
- Alterable data structure Proto
- Learns which fields you use most and displays these by default Proto



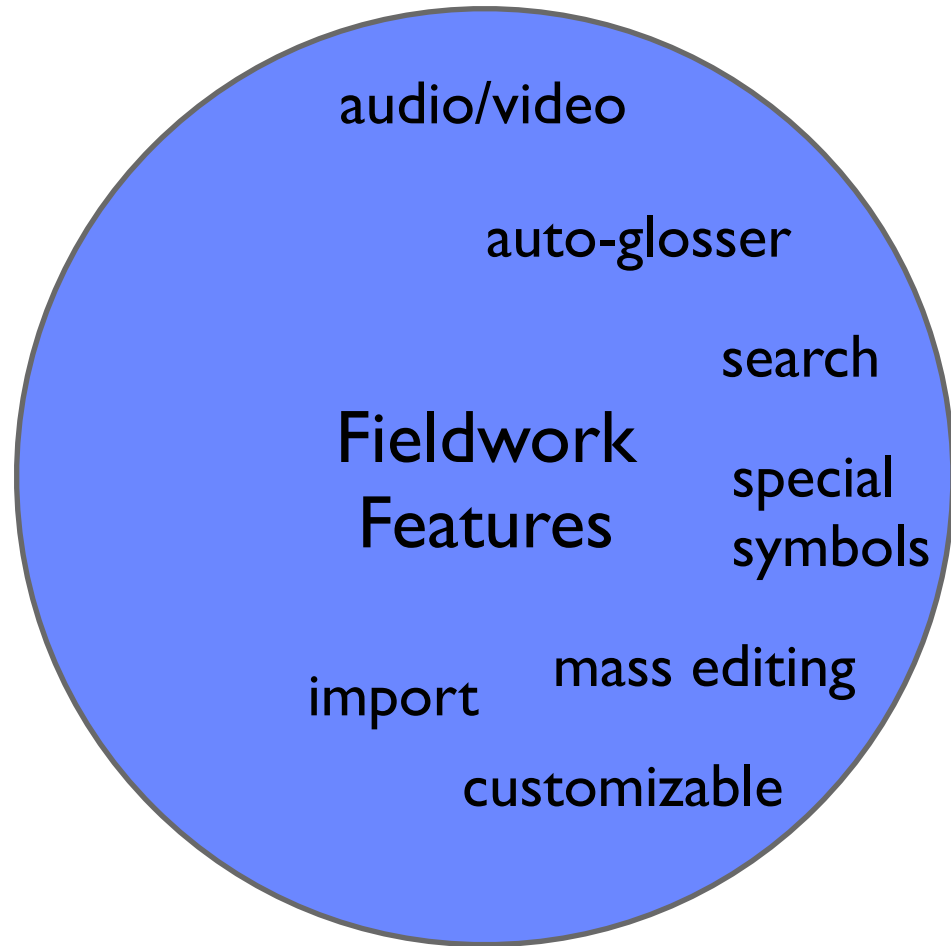
Import Proto

- Smart import analyzes the structure of your file
- Non-lossy: preserves whatever information you have; you decide which field it belongs in



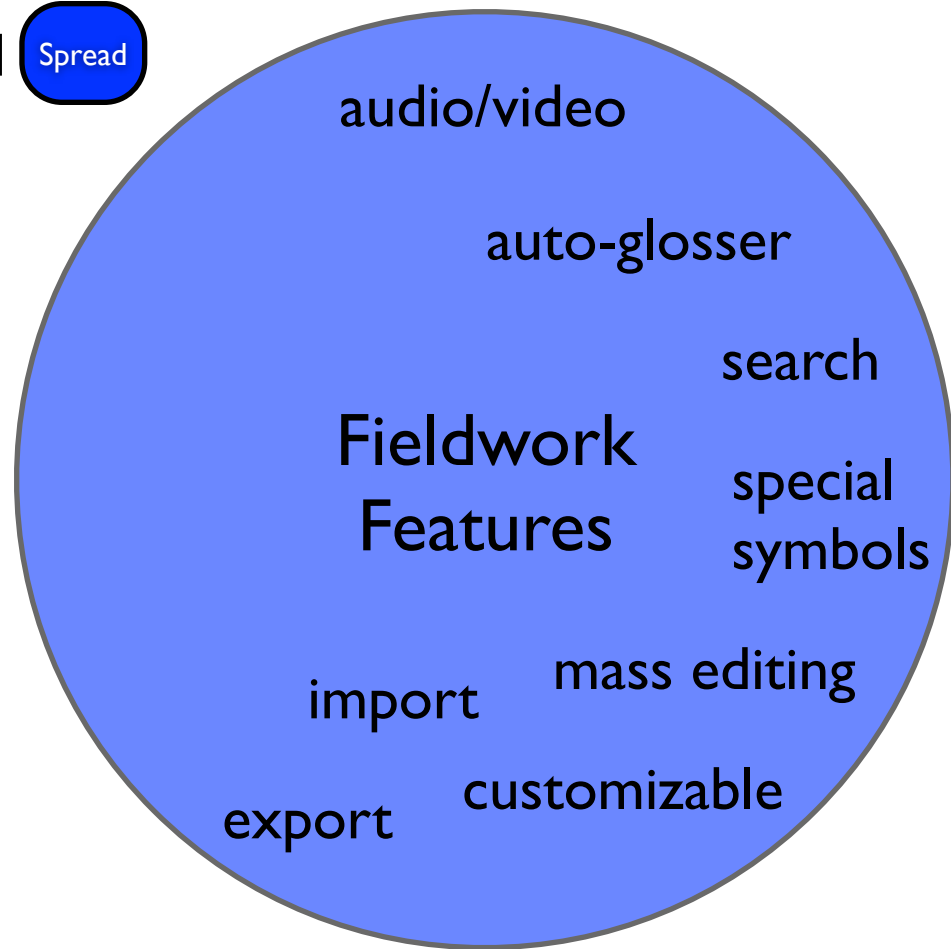
Import Proto

- Smart import analyzes the structure of your file
- Non-lossy: preserves whatever information you have; you decide which field it belongs in
- Works with CSV (FileMaker Pro, Excel) and Praat TextGrid formats



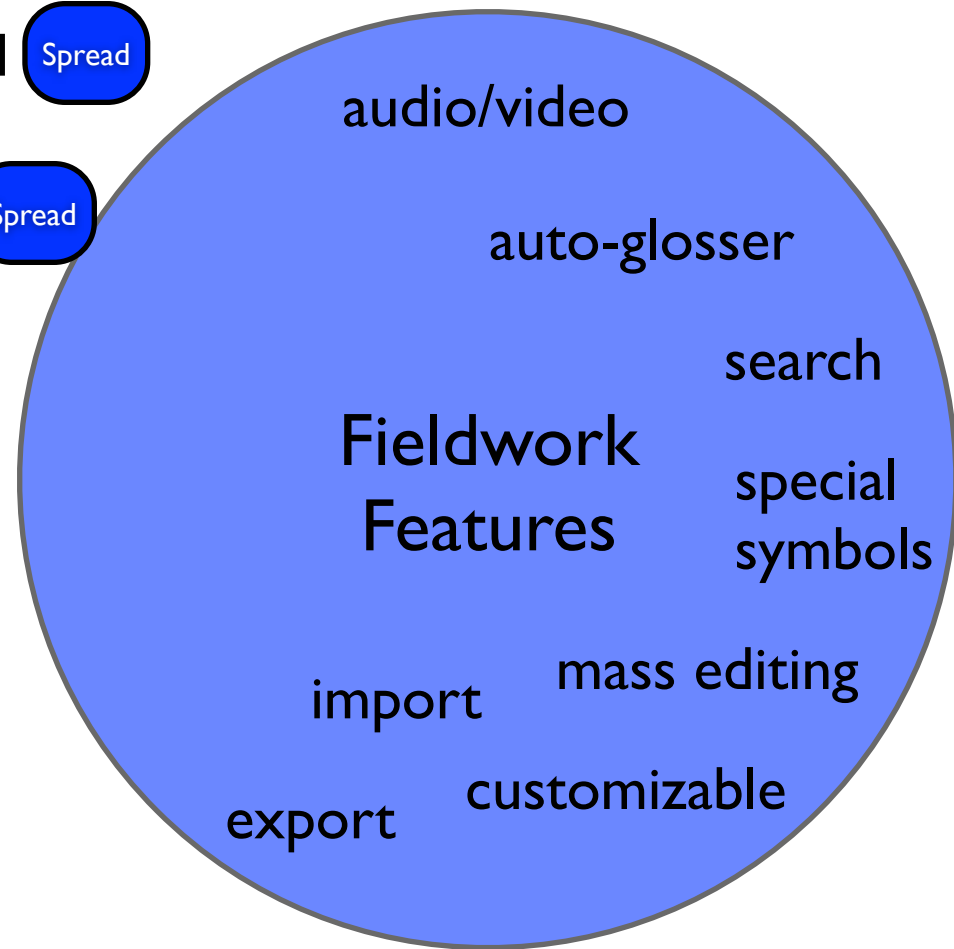
Export

- Export corpus to CSV or JSON Spread



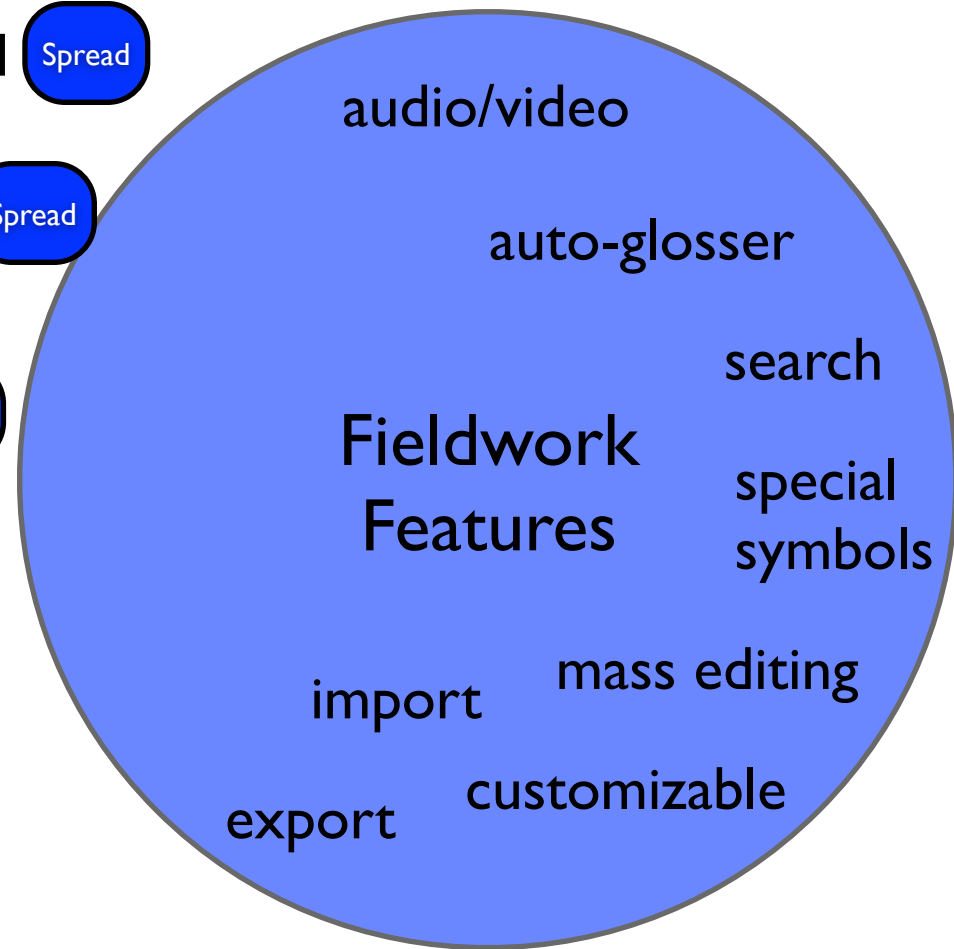
Export

- Export corpus to CSV or JSON Spread
- Export word and morpheme lists with frequencies Spread



Export

- Export corpus to CSV or JSON Spread
- Export word and morpheme lists with frequencies Spread
- Export selected entries as plain text Spread



Export

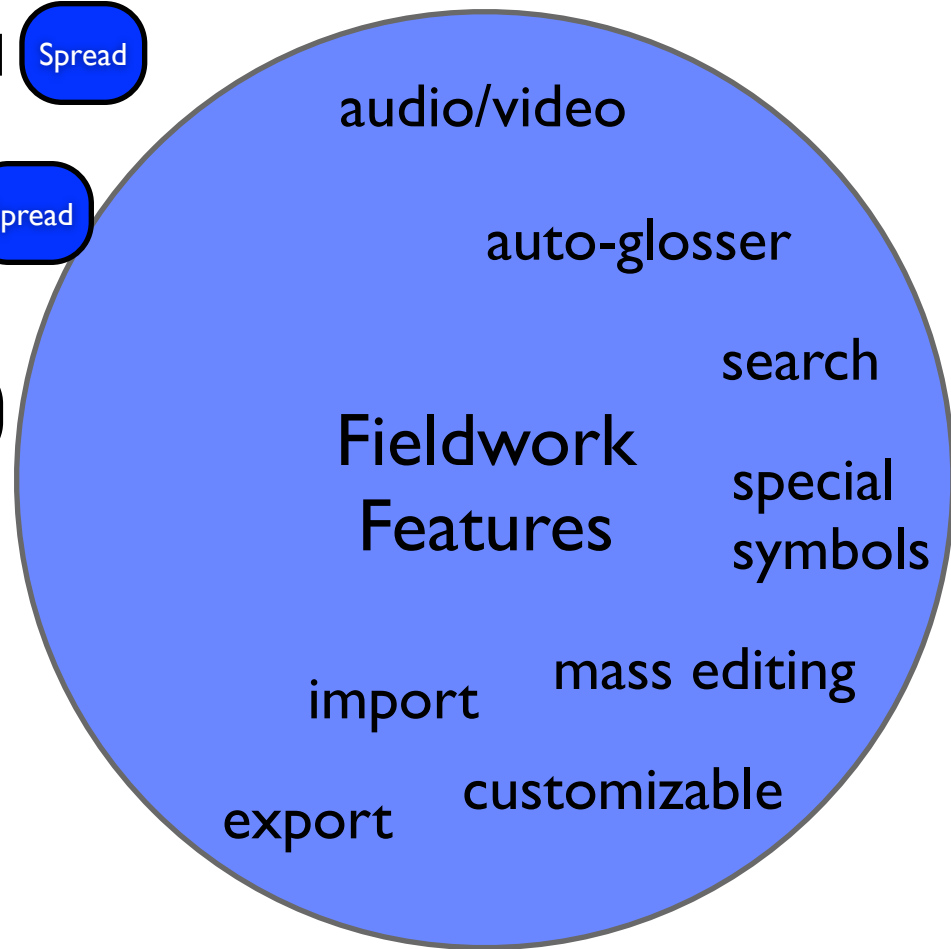
- Export corpus to CSV or JSON
- Export word and morpheme lists with frequencies
- Export selected entries as plain text
- Export to CSV, plain text, and JSON

Spread

Spread

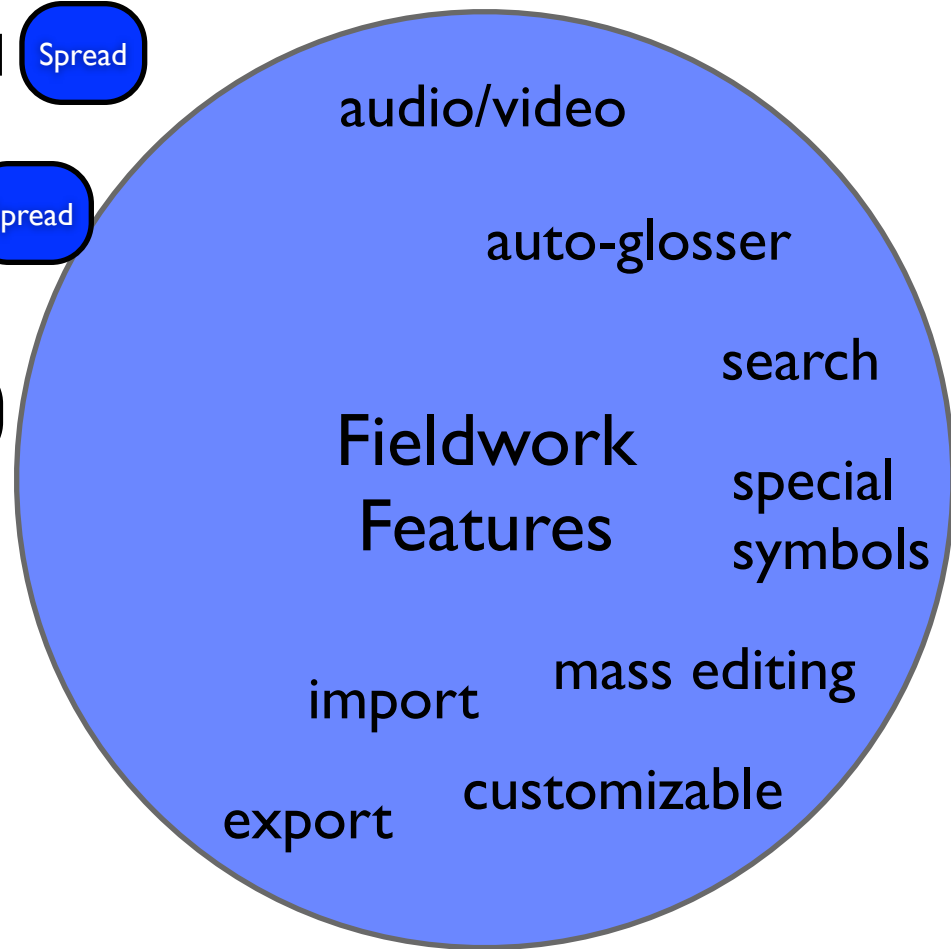
Spread

Proto



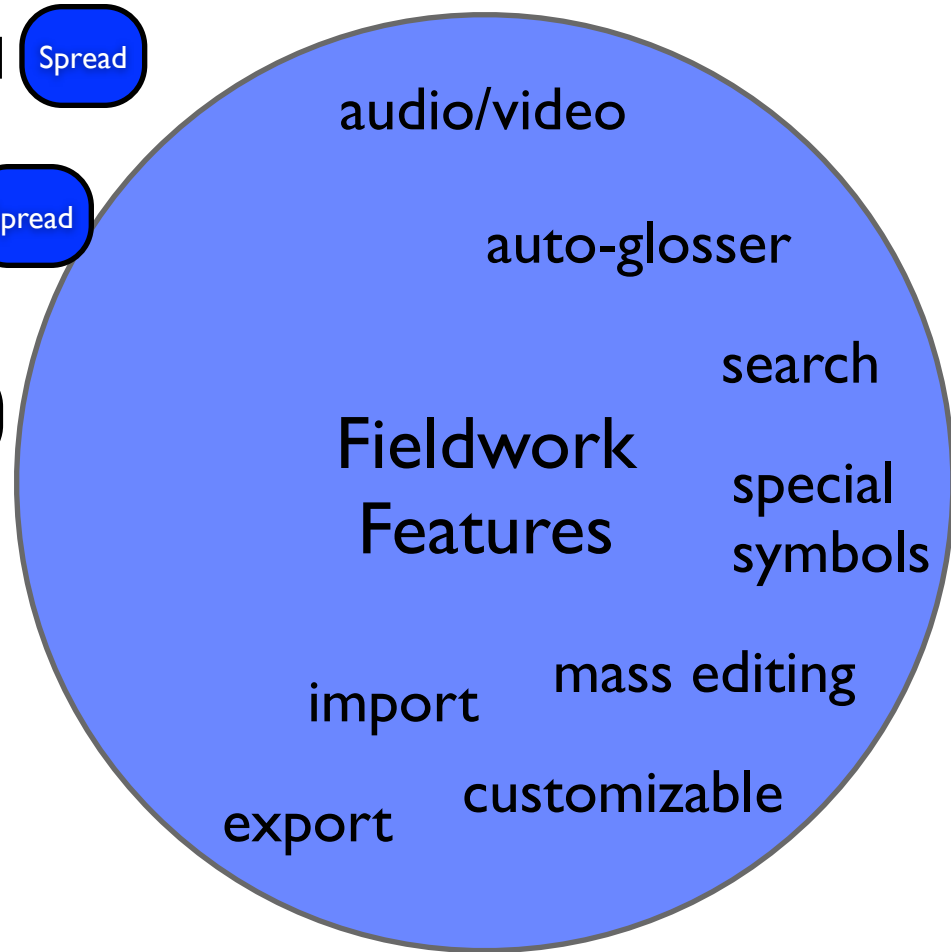
Export

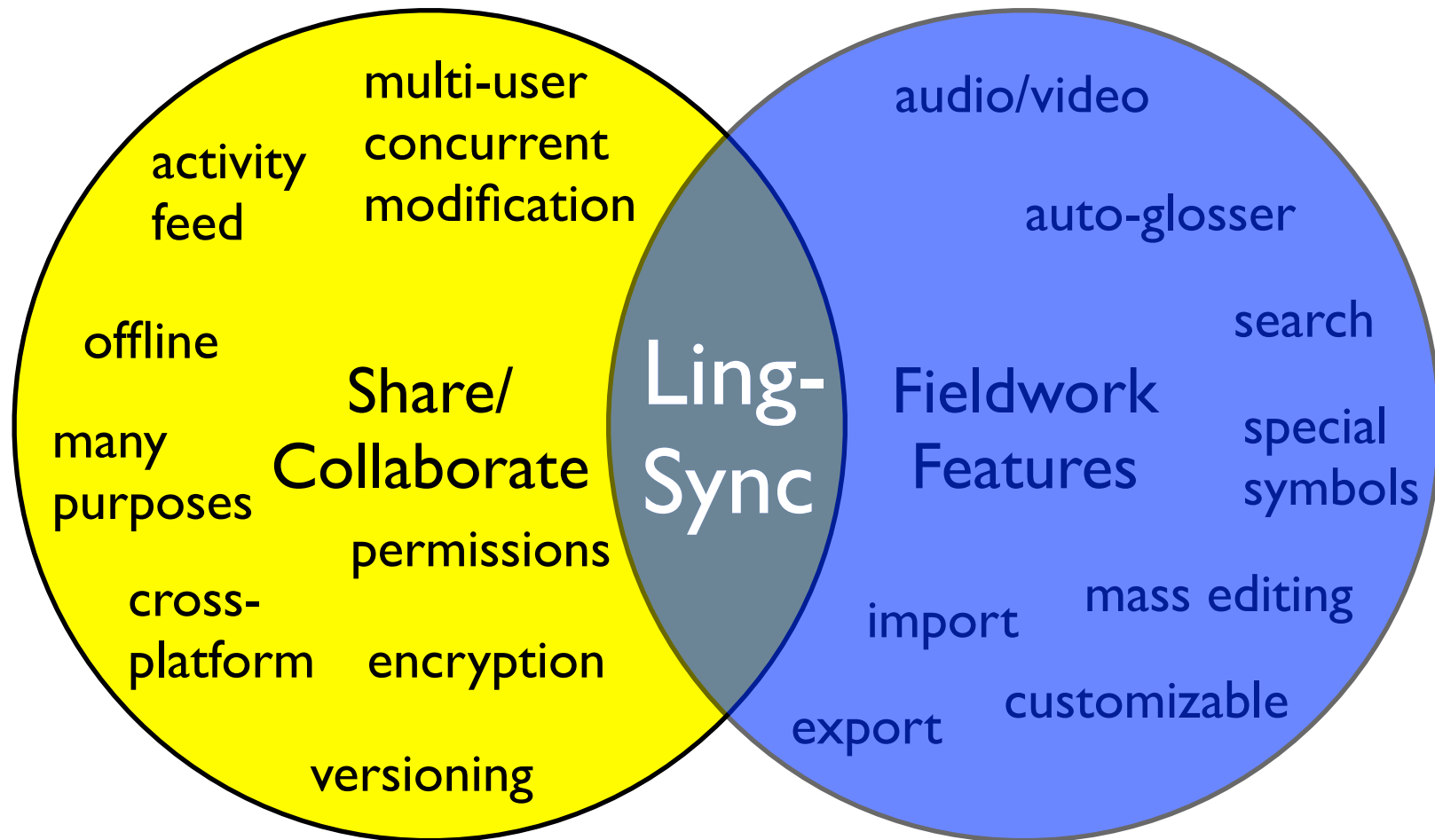
- Export corpus to CSV or JSON Spread
- Export word and morpheme lists with frequencies Spread
- Export selected entries as plain text Spread
- Export to CSV, plain text, and JSON Proto
- LaTeX export with IGT-formatted entries Proto

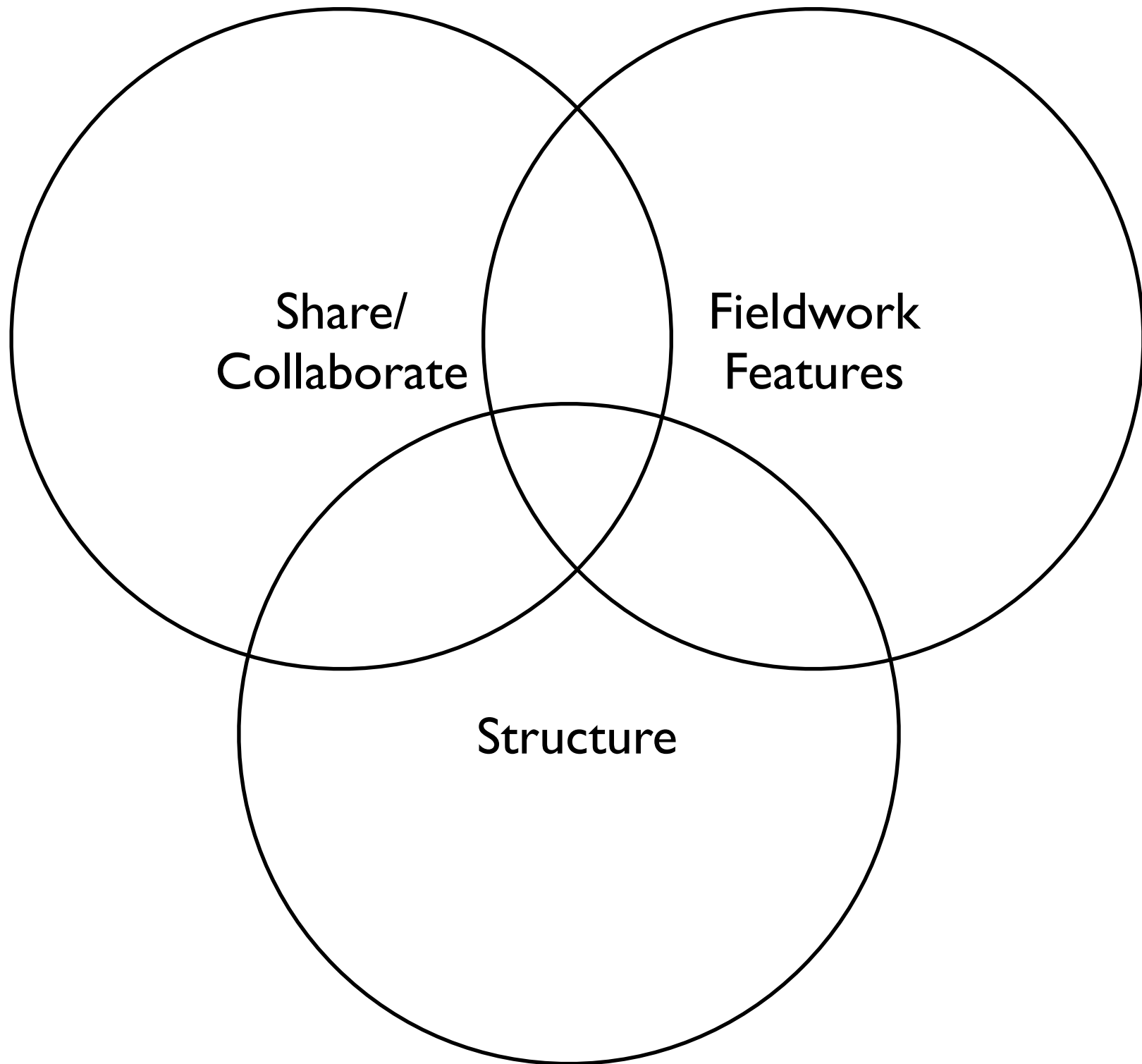


Export

- Export corpus to CSV or JSON Spread
- Export word and morpheme lists with frequencies Spread
- Export selected entries as plain text Spread
- Export to CSV, plain text, and JSON Proto
- LaTeX export with IGT-formatted entries Proto
- Export all text data with associated audio/video data as a single .zip file Proto



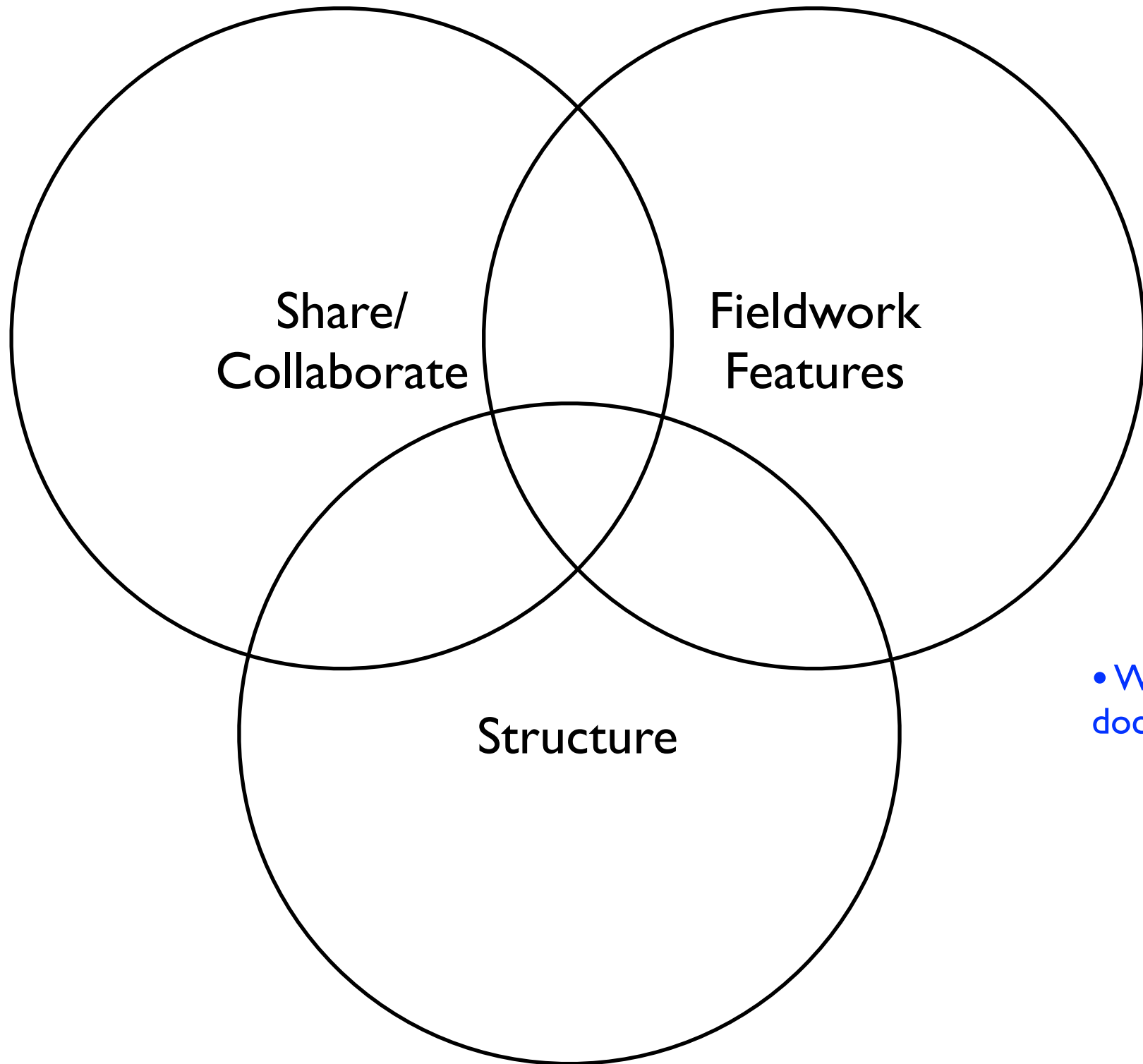




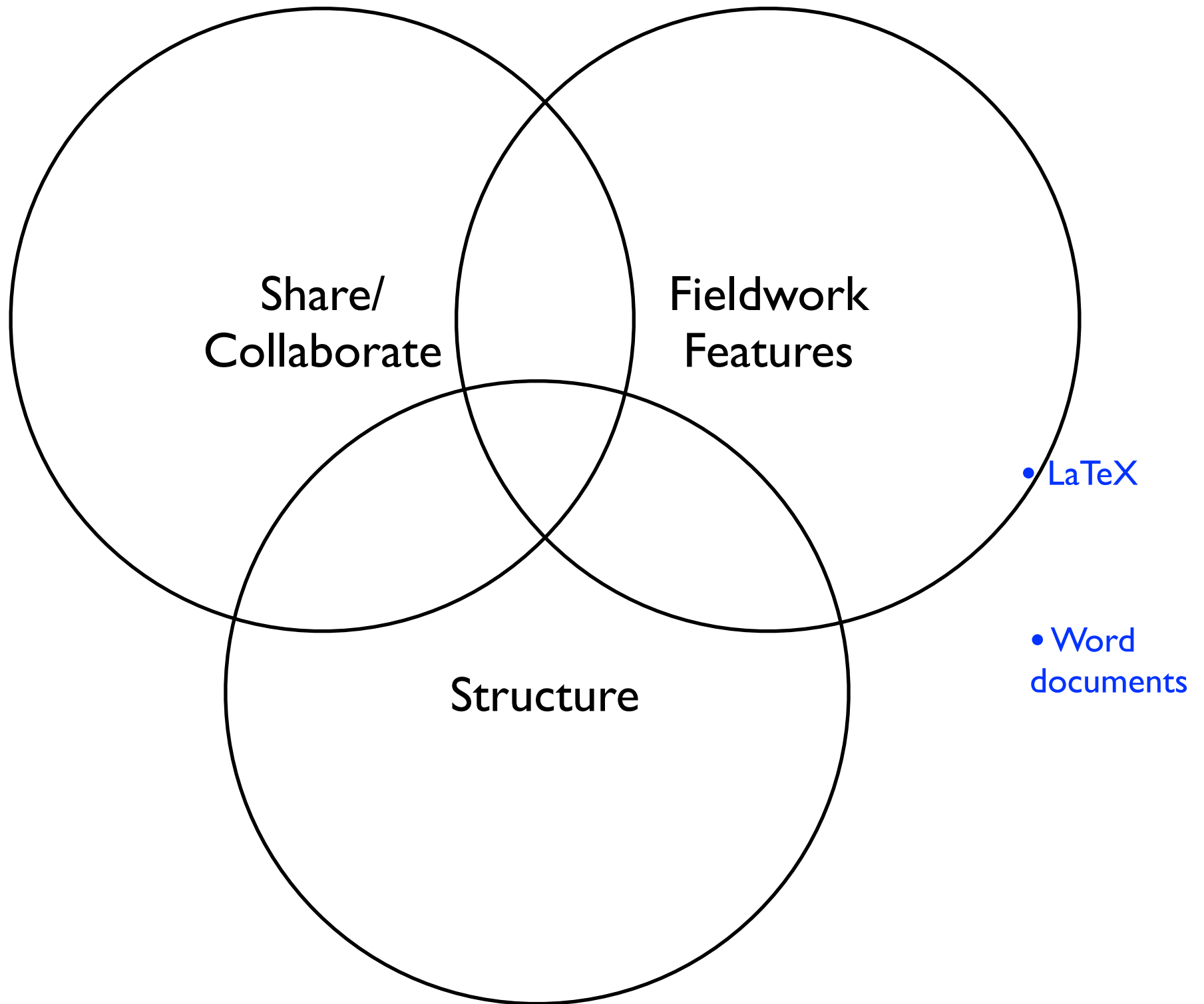
Share/
Collaborate

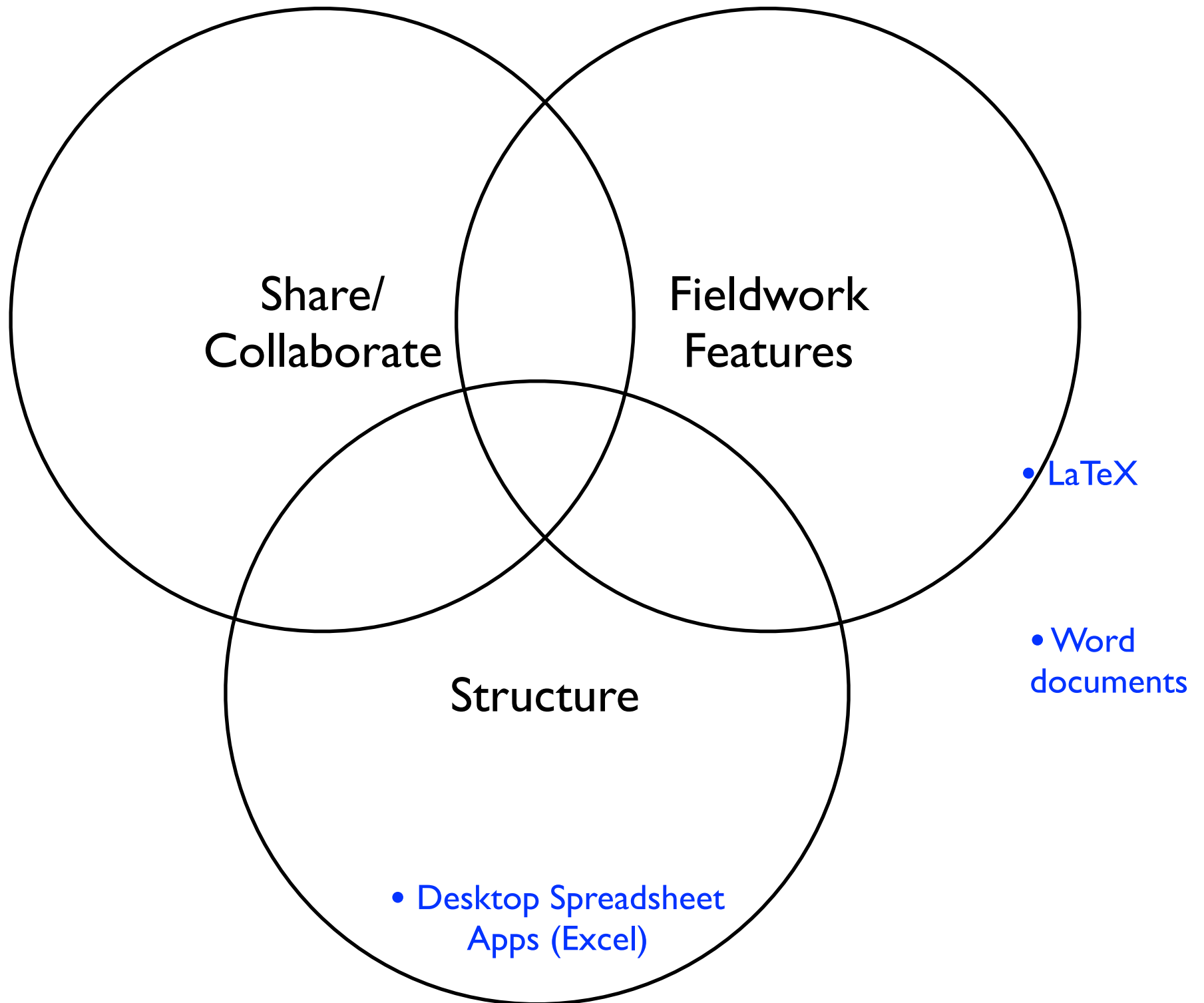
Fieldwork
Features

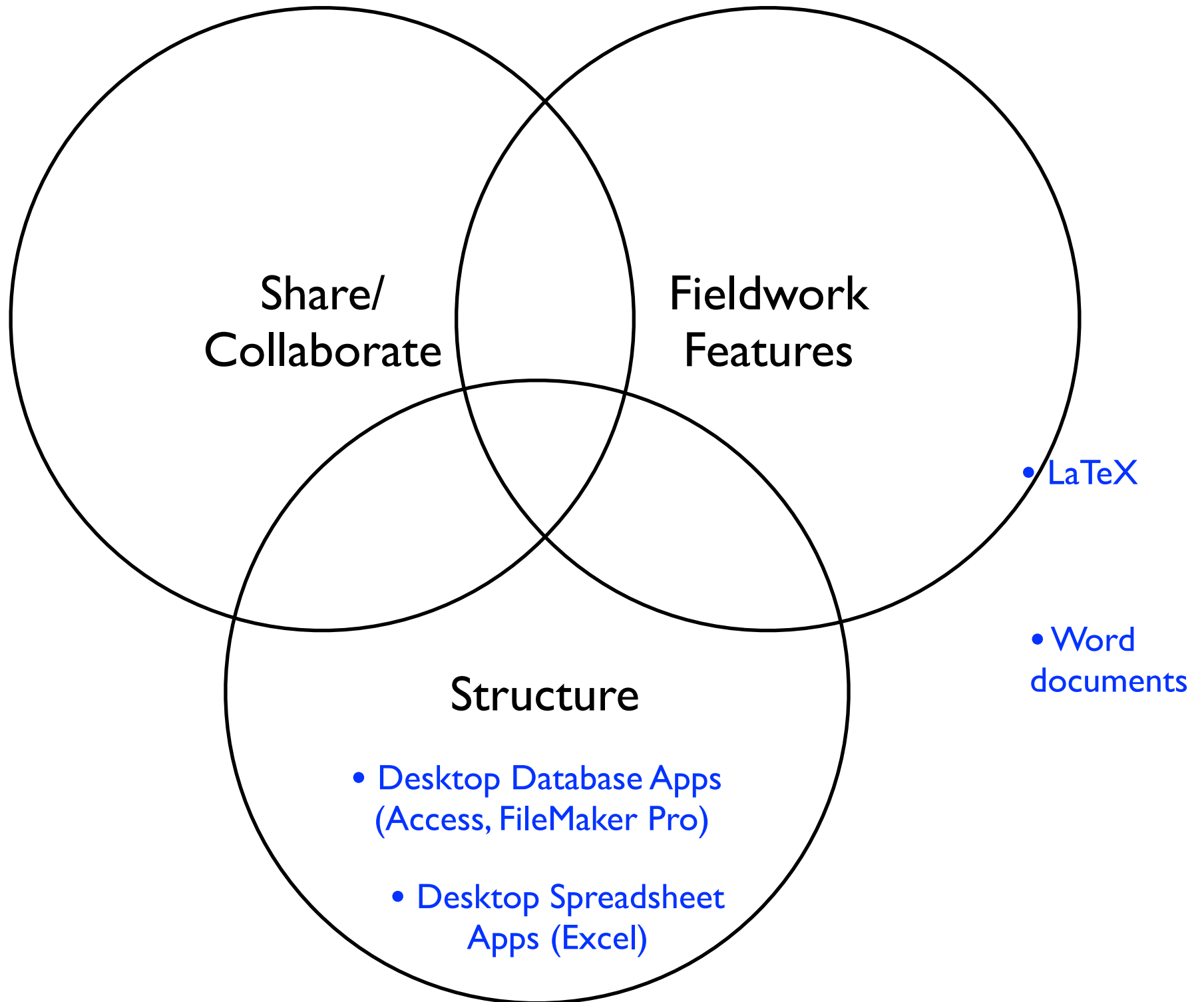
Structure

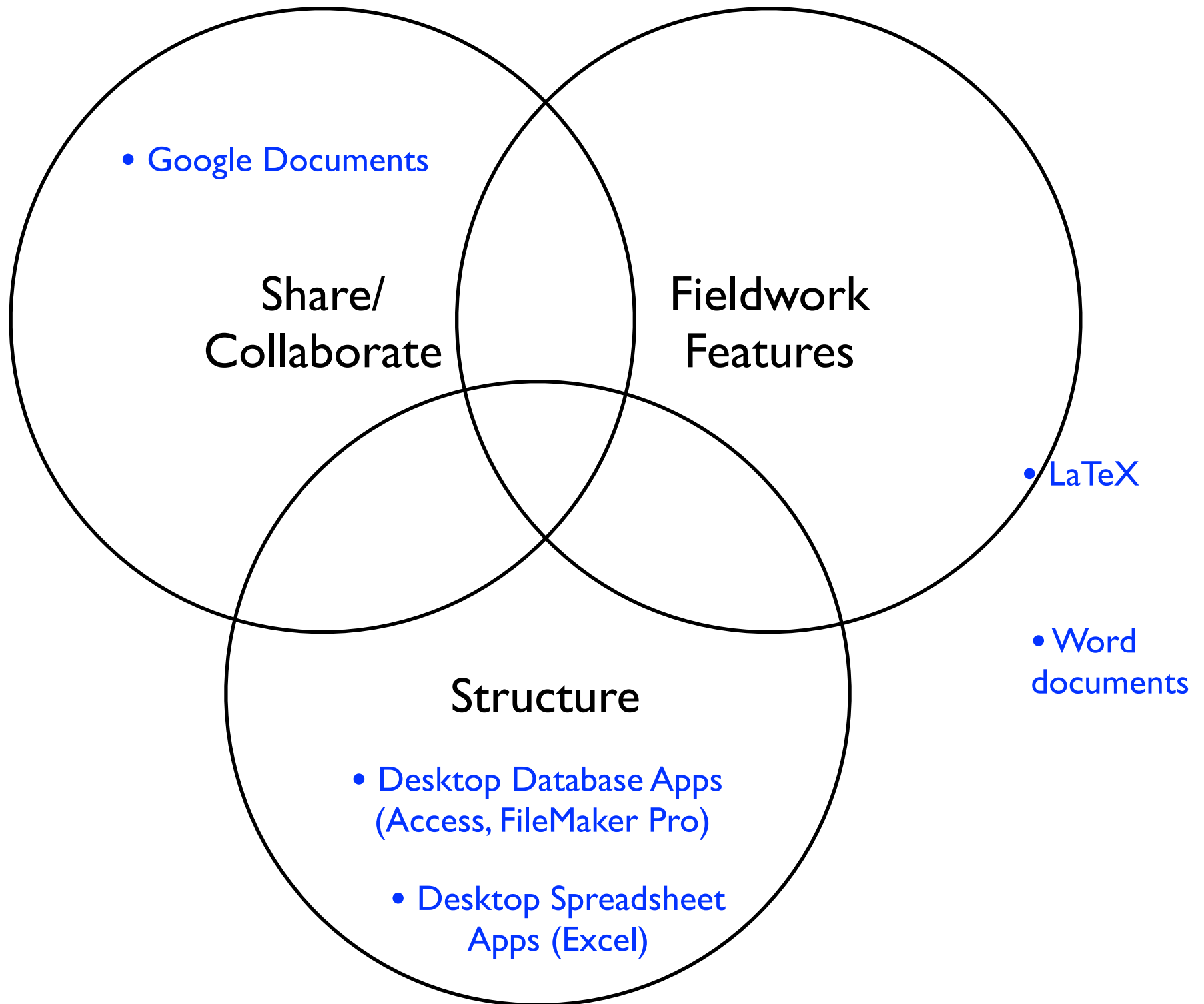


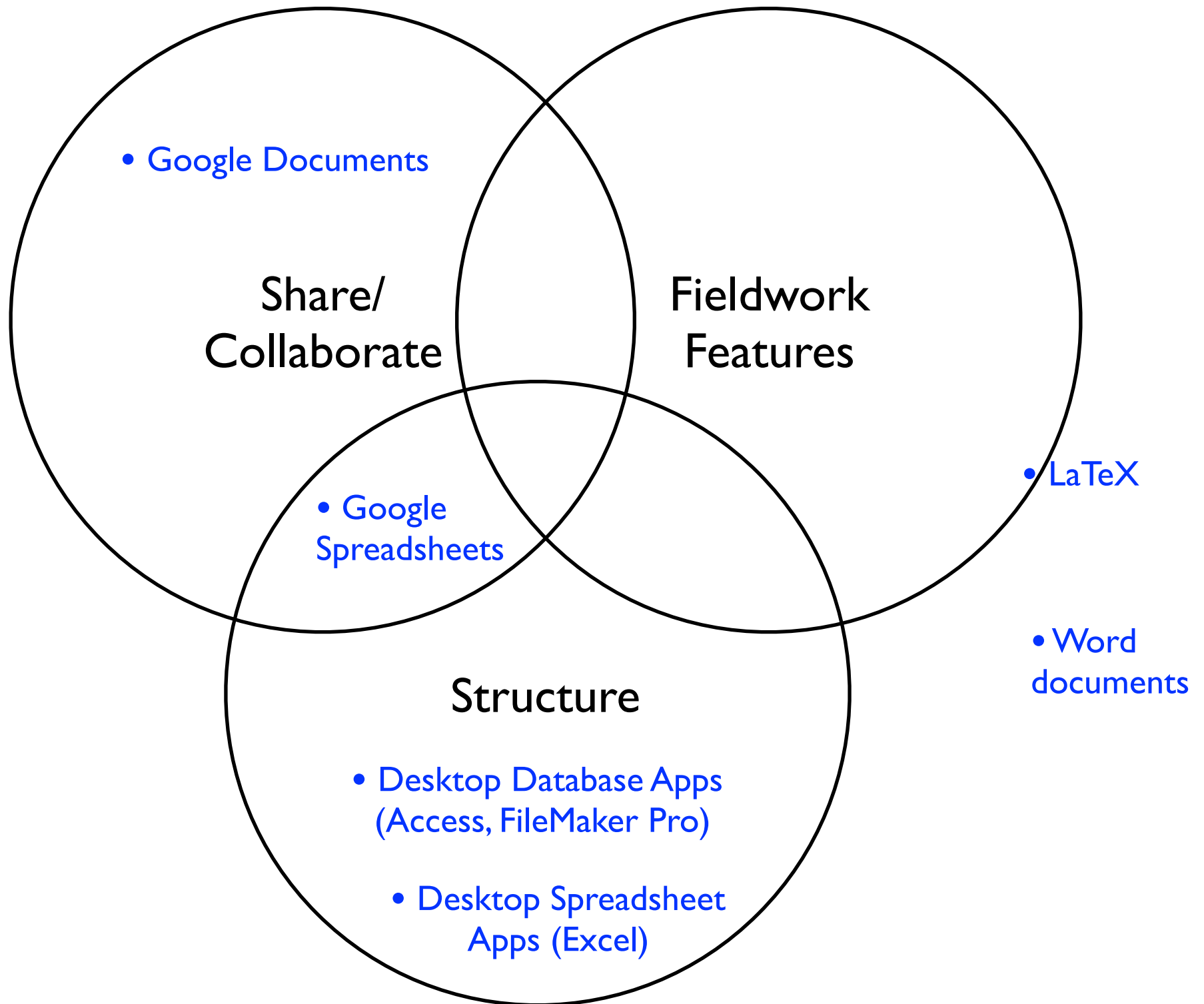
• Word documents

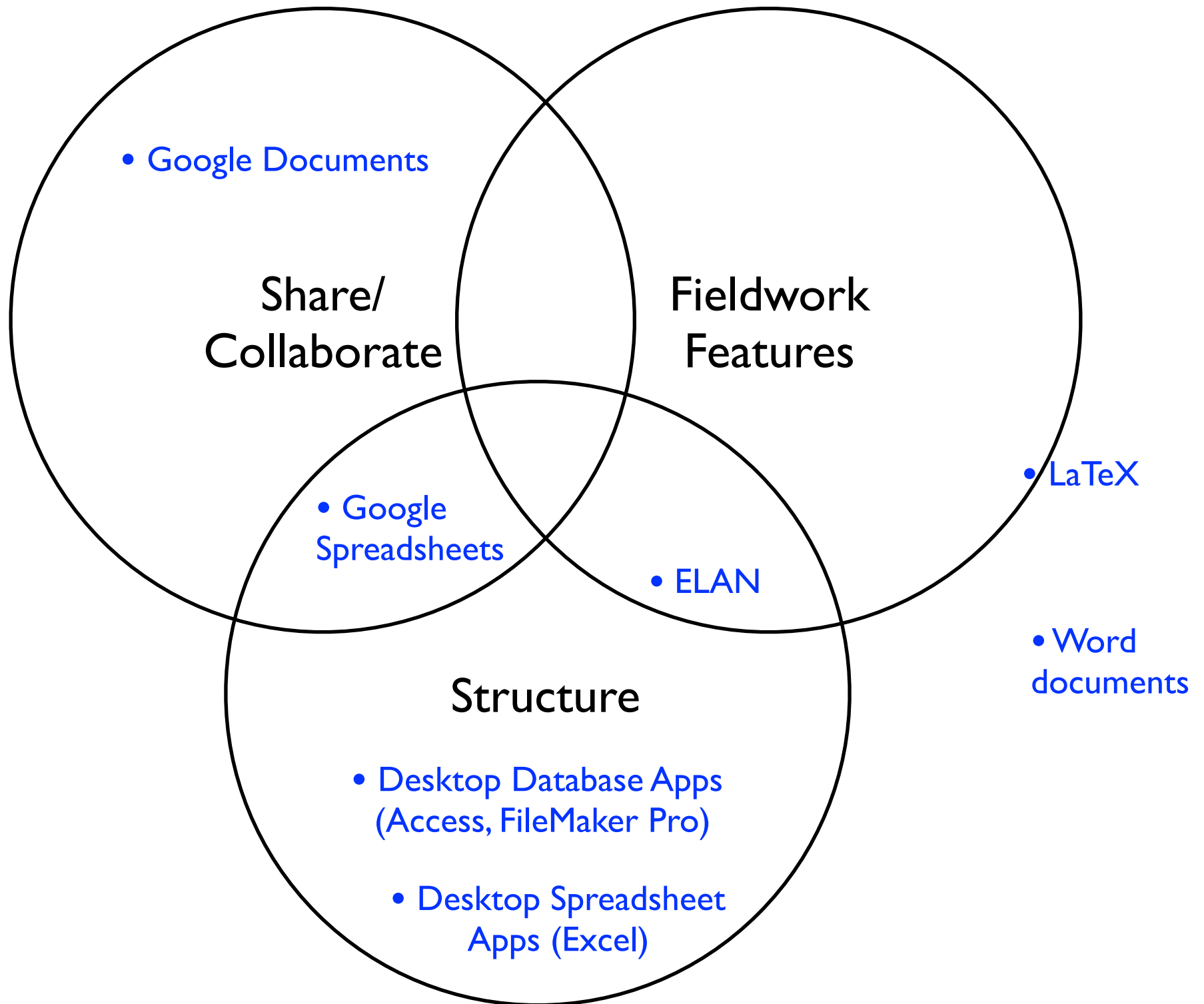


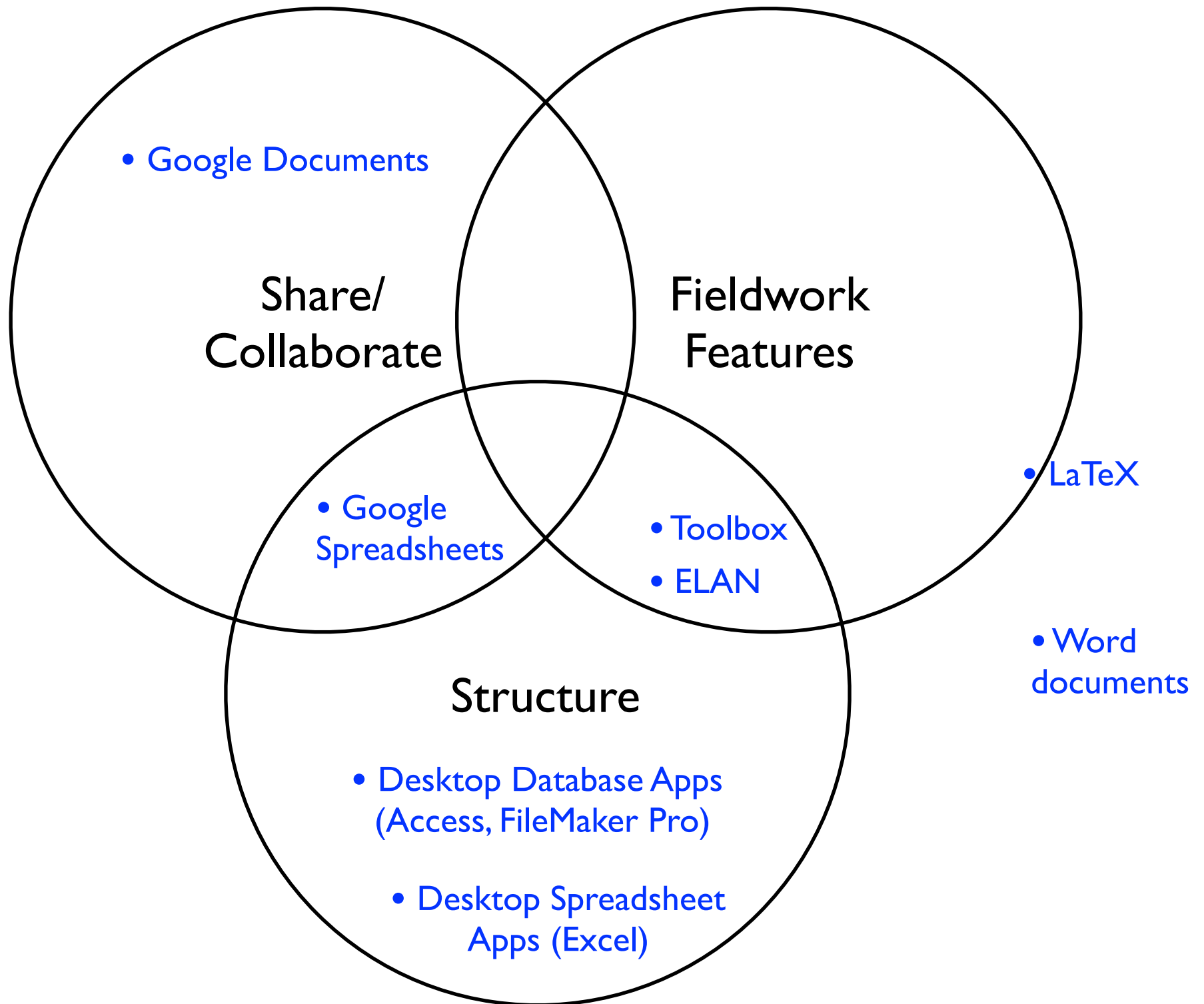


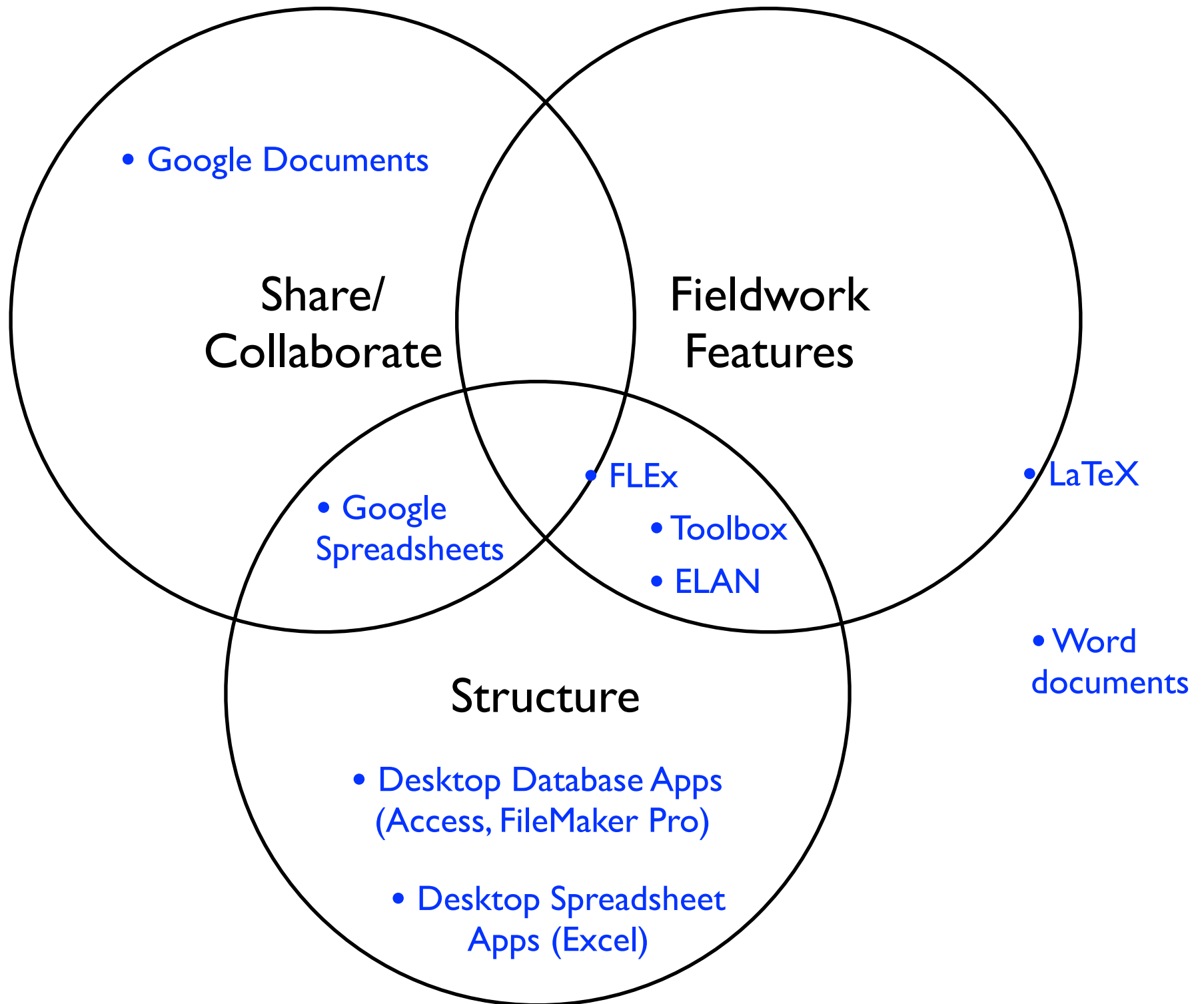


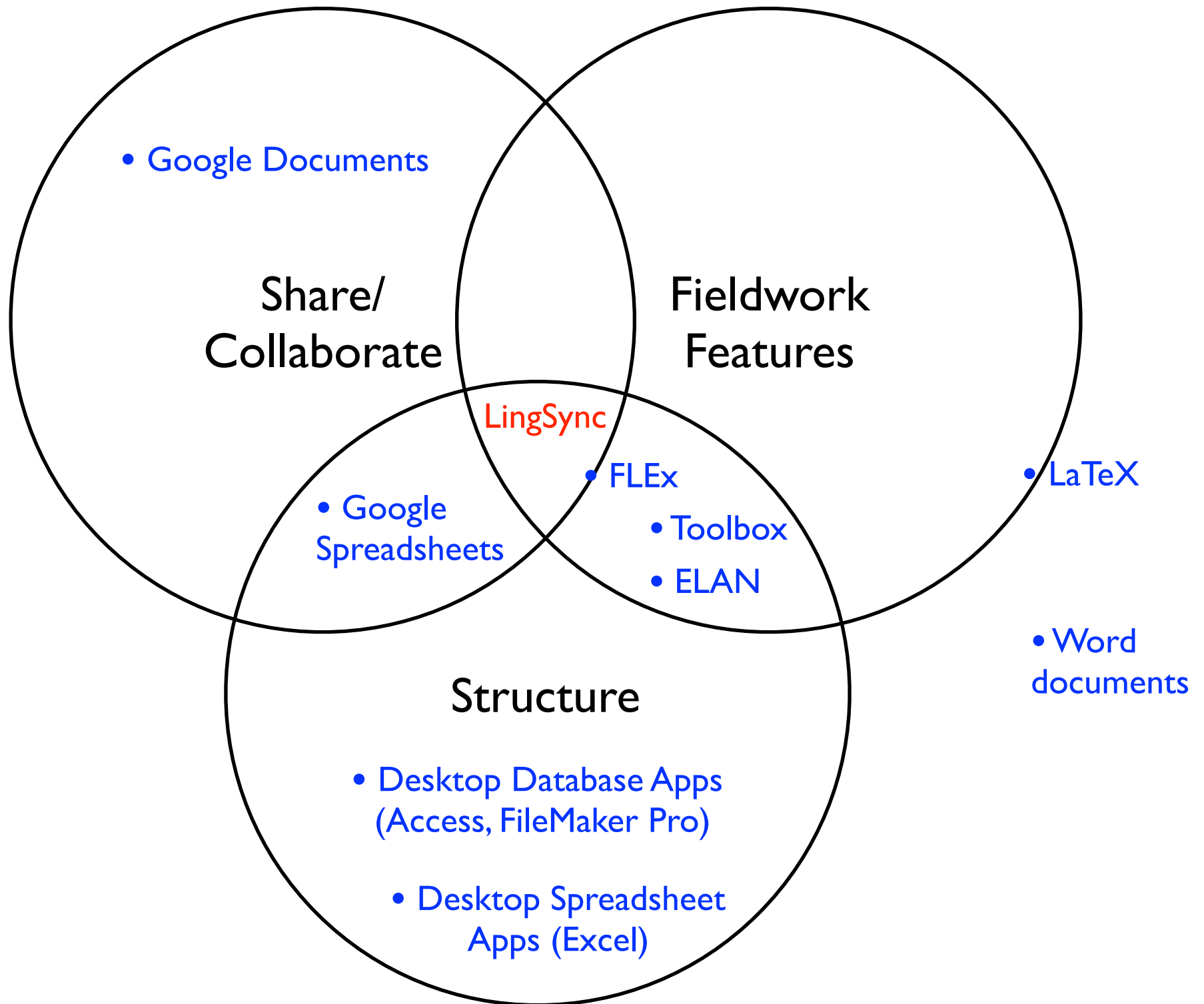












By linguists, for linguists

- We are linguists and fieldworkers: we want the same things you do
- Open-source, by volunteers & interns
 - Feel free to join us on [Github](#)!
- Always free
- We have experience using computational linguistics tools for low-resource languages
- Contact us if you have questions or feature requests, or if you'd like to get involved, we are friendly!

More about LingSync

- [Website](#)
- [Spreadsheet App](#)
- [Prototype App](#)
- [Video Tutorials](#)
- [White Paper for Grant Committees](#)
- [Community Wiki](#)
- [Milestones and Future Modules](#)
- [Open Source Code](#)
- [Privacy Policy](#)
- Advisory Board email:
info@lingsync.org
- Technical Members
email:
support@lingsync.org

LingSync-related Demo Videos

- [Georgian Together](#)
- [Lexicon Browser](#)
- [Audio Web Service](#)
- [Ecoute Alpha Demo](#)
- [Lexicon Browser demo](#)

Acknowledgements

- Alan Bale (Concordia)
- M.E. Cathcart (U Delaware)
- Gina Cook (iLanguage Lab Ltd)
- Jessica Coon (McGill)
- Theresa Deering (Visit Scotland, iLanguage Lab Ltd)
- Joel Dunham (UBC)
- Josh Horner (Amilia, iLanguage Lab Ltd)
- Yuliya Manyakina (McGill)
- Elise McClay (McGill)
- Gretchen McCulloch (McGill)
- Hisako Noguchi (Concordia)
- Michael Wagner (McGill)
- Jesse Pollak (Pomona College)
- Tobin Skinner (Acquisio, iLanguage Lab Ltd)
- Xianli Sun (Miami University)
- More...



Thank you!

References

B. Bickel, B. Comrie, M. and Haspelmath. 2008. The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. <http://grammar.ucsd.edu/courses/ling120/leipziggloss.pdf>

Lynnika Butler and Heather van Volkinburg. 2007. Review of Field- Works Language Explorer (FLEX). *Language Documentation & Conservation*, 1(1):100–106.

MaryEllen Cathcart, Gina Cook, Theresa Deering, Yuliya Manyakina, Gretchen McCulloch, and Hisako Noguchi. 2012. LingSync: A free tool for creating and maintaining a shared database for communities, linguists and language learners. In Robert Henderson and Pablo Pablo, editors, *Proceedings of FAMILI II: workshop on Corpus Approaches to Mayan Linguistics 2012*, pages 247–250.

Henry Davis. 2006. A Teacher's Grammar of Upper St'át'imcets. MS, University of British Columbia, Vancouver, BC.

Joel Dunham, Gina Cook, and Joshua Horner. 2014. LingSync & the Online Linguistic Database: New models for the collection and management of data for language communities, linguists and language learners. To appear in *The Proceedings ComputEL: The use of computational methods in the study of endangered languages*, workshop at the 52nd Annual Meeting of the Association for Computational Linguistics.

Gretchen McCulloch. LingSync.org: A revolutionary new app for field linguistics and language communities. ([Slides](#).)

Stuart Robinson, Greg Aumann, and Steven Bird. 2007. Managing fieldwork data with ToolBox and the Natural Language Toolkit. *Language Documentation & Conservation*, 1(1):44–57.

Chris Rogers. 2010. Review of Field- Works Language Explorer (FLEX) 3.0. *Language Documentation & Conservation*, 4:78–84.

Plan of the Tutorial

9:00-9:10	Introduce the LingSync team
9:10-9:40	<i>Why LingSync?</i> (presentation)
9:40-9:50	Learn X/ Georgian Together video demo
9:50-10:10	Question period 1
10:10-10:20	<i>coffee break</i>
10:20-10:35	<i>LingSync Practical Tutorial</i> (presentation)
10:35-10:45	Question period 2
10:45-11:00	Participatory collaborative data entry
11:00-11:30	Question period 3

Georgian Together

