

Crowdsourcing Cultural Heritage

Mia Ridge and Ben Brumfield

HILT 2014

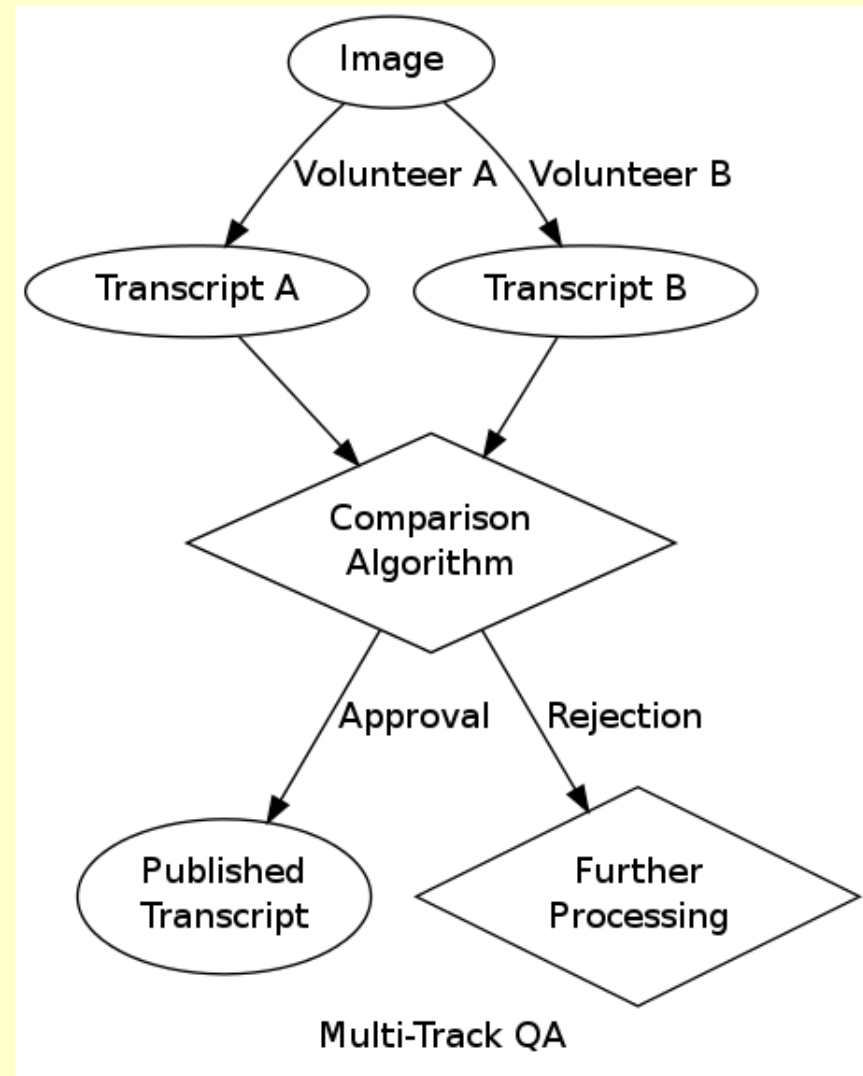
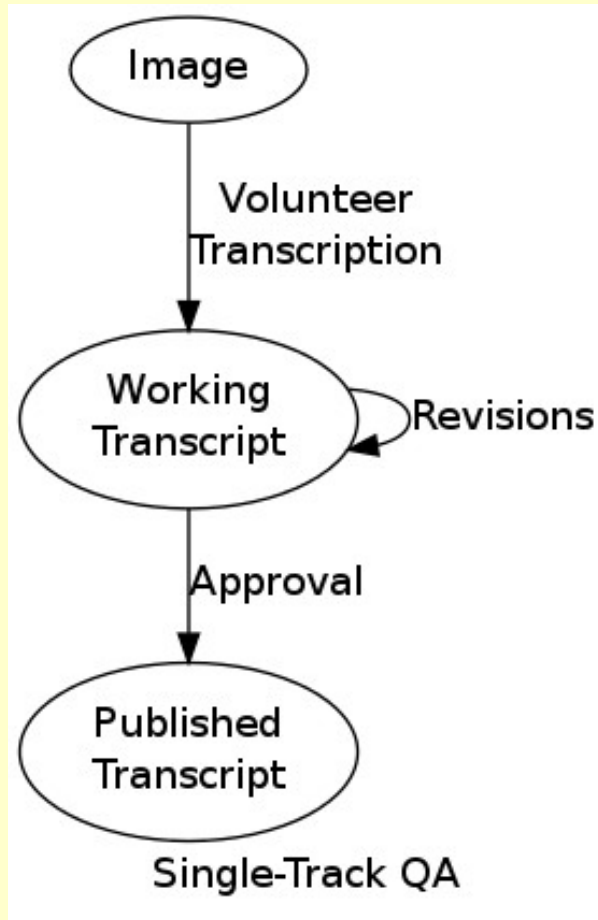
August 4-8, Maryland Institute for
Technology in the Humanities

Welcome!

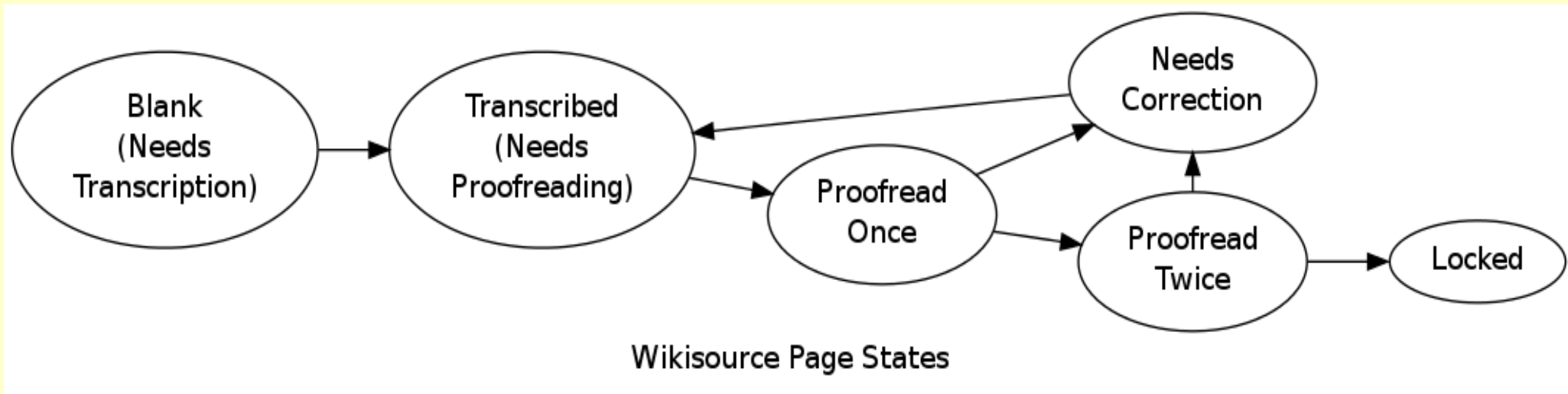
Session 13: Thursday, 10:45 am – 12:00 pm

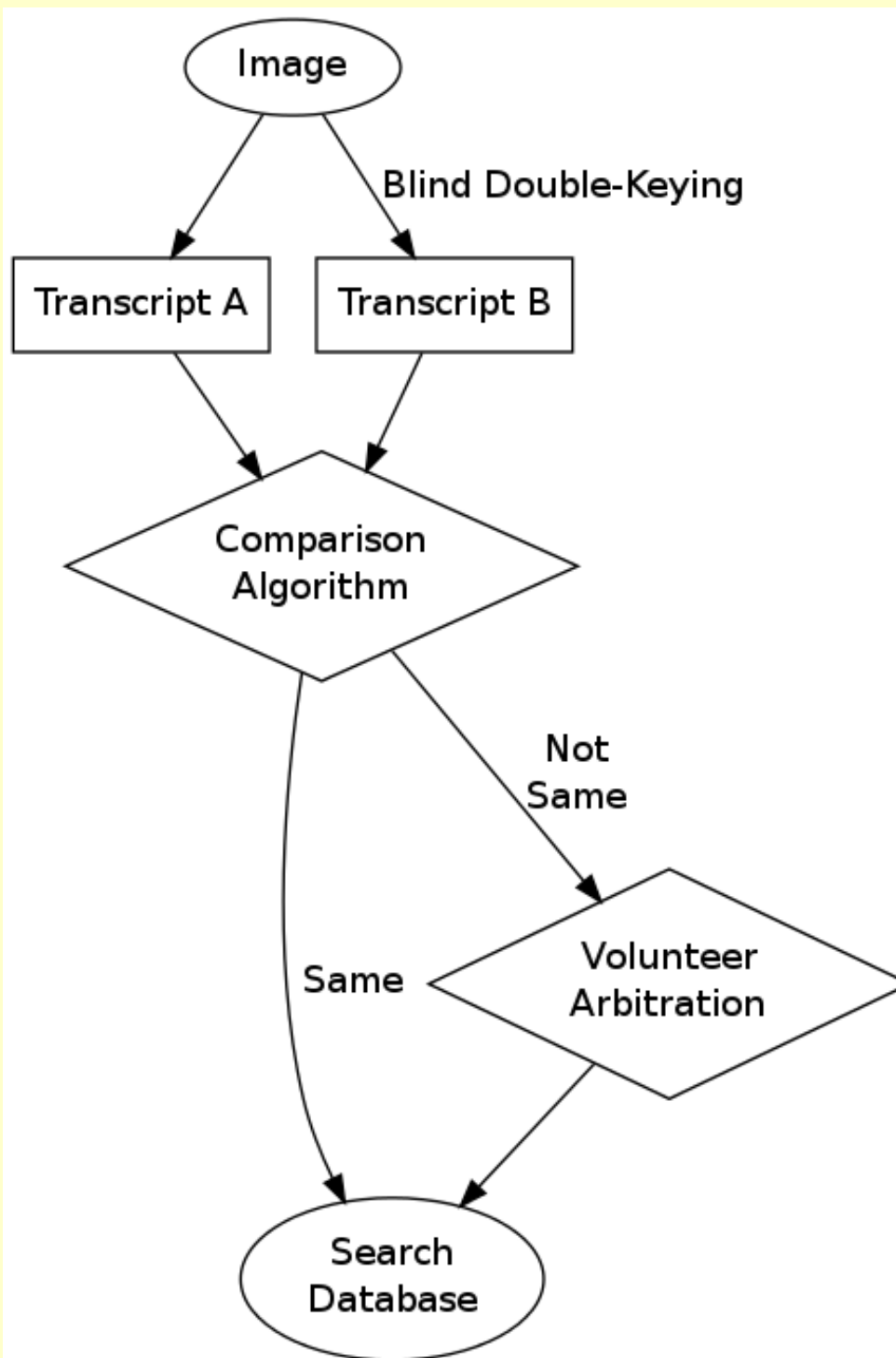
Quality Control for Crowdsourcing
Community Ecosystem and Core Tasks
Special Guest

Quality Control

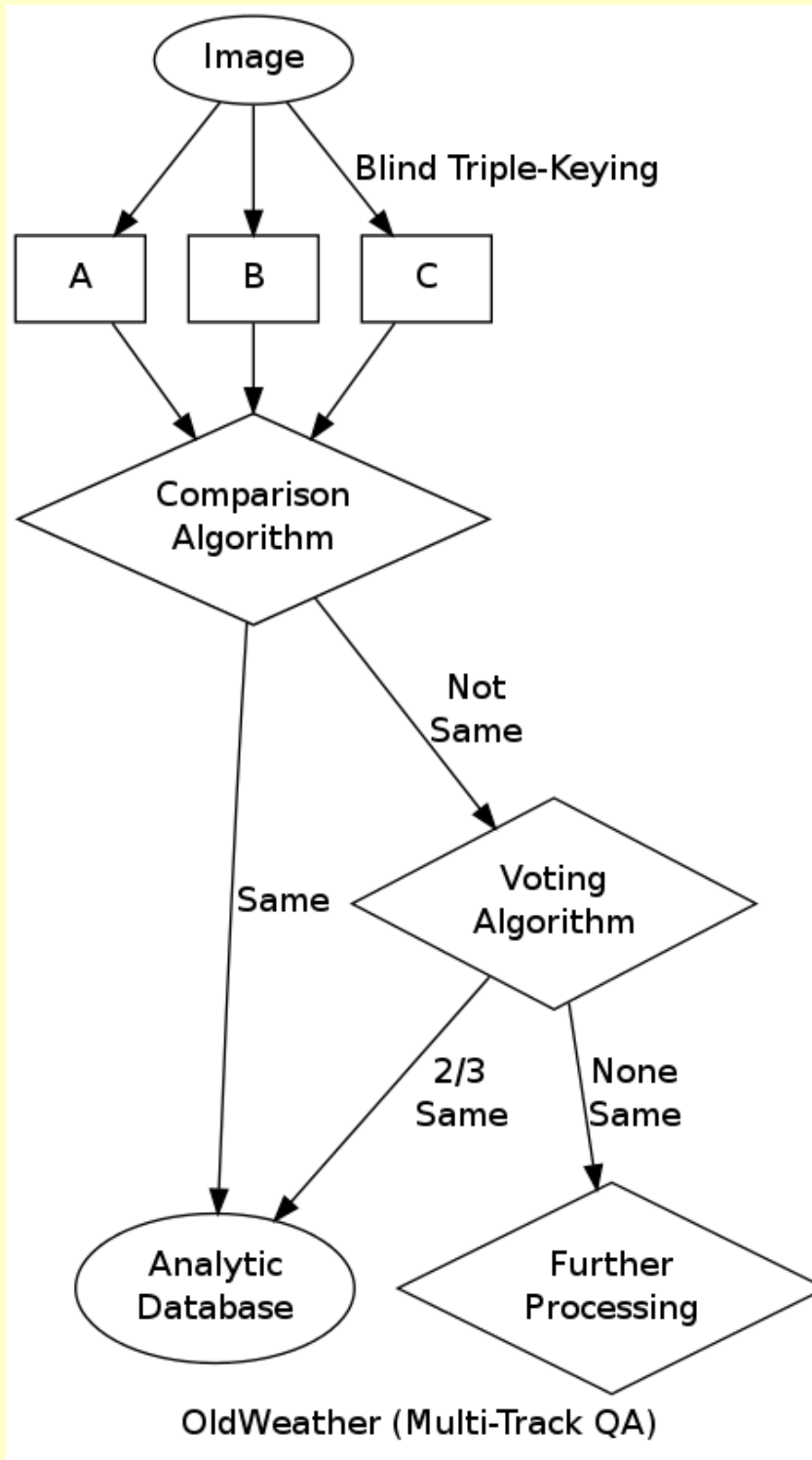


Quality Control





FamilySearch Indexing (Multi-Track QA)



Quality Control

What's a unit of work?

- ROIs

- Lines or Words

- Fields

- Pages

How do you split?

- Algorithmic B-box/ROI identification

- Clicking and Drawing

Quality Control

How do you identify the “same” contribution?

How do you compare contributions?

Quality Control

Which is correct?

Geryong magnirostris tenebrosa

Gerygone magnirostris tenebrosa

gerygone magnirostris tenebrosa

Gerygone magnirostris tenebrosa

gerygone magnirostris tenebrosa

Gerycong Magnirostris Renebrosa

Quality Control

Which is the same entry?

User ID	Distance From Top of Page
LaGeographe	0.05992010652463382
estron21	0.08183538315988648
tdf4638	0.085219707057257
estron21	0.09839167455061495
tdf4638	0.10075454948956947
KikkerLG	0.10267005721551176

Exercise

Hands-on: designing community ecosystem around core tasks - how can people grow their skills while helping the project?

Holistic design for a virtuous circle

- Ecosystem of crowdsourcing projects
 - Crowdsourcing tags or transcription to support discoverability and other crowdsourcing tasks
 - Validate data created in other projects
 - Help surface most interesting items
 - Help find most difficult items for advance challenges
- Supports different abilities, skills, types of challenge

Special Guest

Dr. Meghan Ferriter

Project Coordinator for the Smithsonian
Transcription Center

@meghaninmotion