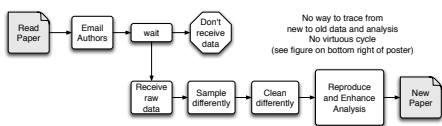## Research Community

Those researching free and open source software and its development, from Software Engineering, Management, Organizational Studies, Economics and Sociology.

This community is interested in questions like:
- How do projects coordinate and make decisions?
- How does software 'evolve' over time?
- How is software complexity managed?
- What is the social structure of the projects?
- How are people socialized into these communities?
- How do the projects change their practices over time?
- Can open source practices be used in other online collaboration communities?

Currently there is **duplicated collection and analysis effort** and **low replicability** which is hampering the community's ability to building each others work and therefore progress the field.

Without Collaboratory

No way to trace from new to old data and analysis
No virtuous cycle
(see figure on bottom right of poster)

Data on Open Source

Project "Demograpics"
Email lists
Bug Trackers
Code repository logs (CVS)

FLOSSmole
Spiders 8 repositories (inc Sourceforge, ObjectWeb, Savannah).
Data on 130K projects and 250K developers
180G + 15G Month

Notre Dame Dumps
Sourceforge backend db dumps
Includes all Trackers and Forums
140G + 10G month

CVSanalY
CVS logs for 30K projects 25 Million atomic commits
10G raw, 3G db

Surveys

Interviews & Transcripts

Observation Journals

Content Analysis Schema
Coded Text

Building a collaboratory for research on open source software development

floss.syr.edu

ossmole.wiki.sourceforge.net
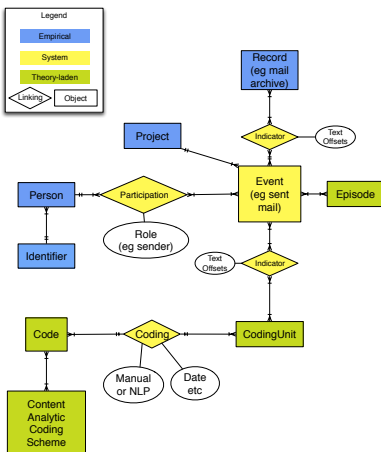
Advice Wanted :)
1. data models
2. workflow tools
3. community building

FLOSSmole++ project:

Megan Conklin (Elon University)
Kevin Crowston (Syracuse University)

Poster presented by James Howison,
PhD Candidate, Syracuse University
School of Information Studies
NSF CRI grant #0708767 and #0708767

Possible Data Model

Legend
- Empirical
- System
- Theory-laden
- Linking / Object

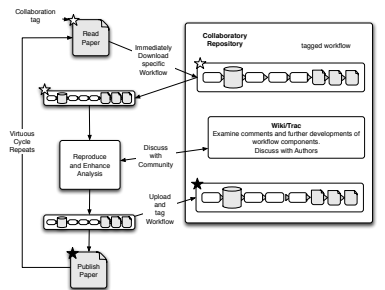The goal is to share:

Data and Data formats

Data annotations

Research Workflows

Algorithms for calculating metrics
(eg effectiveness, implicit roles)

Make it possible to discover publications from data, and data from publications

Workflows

What workflow tools?
How to represent? How to share,
How to create (GUI tools?), How to annotate?