



NSF Campus Cyberinfrastructure PI and Cybersecurity Innovation for Cyberinfrastructure PI Workshop

October 3-4 | Albuquerque, New Mexico

NSF Program: CiCi

Program Area: Data Provenance

Award Number: 1547360

PI: Zachary Ives

Co-PIs: Susan Davidson, Sampath Kannan, Val Tannen

Project Title: Provenance-Based Trust Management for Collaborative Data Curation



Zachary Ives

Professor of Computer
& Information Science
University of
Pennsylvania
zives@cis.upenn.edu



Susan Davidson

Professor of Computer
& Information Science
University of
Pennsylvania
susan@cis.upenn.edu



Sampath Kannan

Professor of Computer
& Information Science
University of
Pennsylvania
kannan@cis.upenn.edu



Val Tannen

Professor of Computer
& Information Science
University of
Pennsylvania
val@cis.upenn.edu



NSF Campus Cyberinfrastructure PI and Cybersecurity Innovation for Cyberinfrastructure PI Workshop

October 3-4, 2017 | Albuquerque, NM

Provenance-Based Trust Management for Collaborative Data Curation

Challenge:

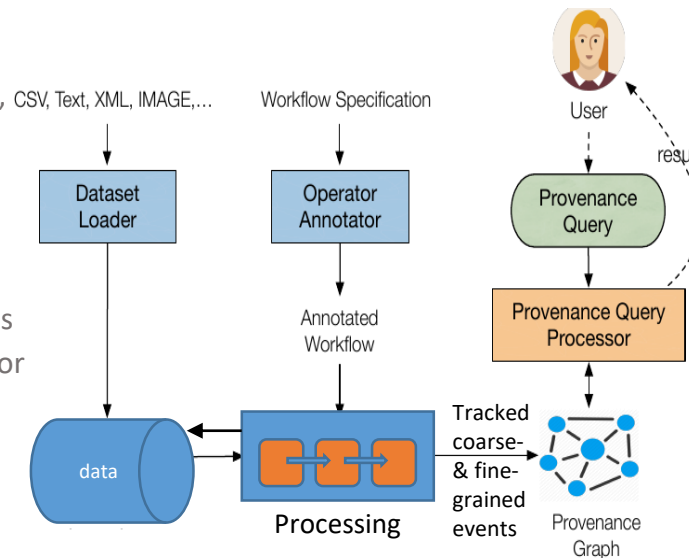
- How to make human curation “scale” by taking contributions from many

Approach:

- Make it easy to instrument raw data through computation, across many data modalities, types, analysis stages
- Develop more powerful formalisms for tracking data provenance
- Learn to map from provenance to trust

Solution(s) :

- Dropbox-style **provenance tracker** for any OS
- Instrumented **Jupyter Notebook** for analysis
- “**Gray box**” modeling – *take file -> code -> file* provenance, turn into *record -> operation -> record* provenance!



Scientific Impact:

- Provenance + reproducibility without requiring new tools and styles of work
- New techniques for tracking how data, not just files, relate!
- Can determine utility, conflicts, and impact of individual annotations, extractions, contributions

Status:

- *Early tools for tracking, browsing provenance!*
- *<https://pennprovenance.net>*