In preparation for the Workshop on Scholarly Databases & Data Integration on August 30th and 31st, 2006 at the School of Library and Information Science at Indiana University in Bloomington, Indiana, we ask you to provide a brief bio and your input to the questions below.

We plan to make your input available via the workshop web page before the meeting to ease introduction of participants. We will also use your feedback to more effectively structure the workshop.

Thank you for your time.

Biography (about 250 words)

**Smalheiser** - MD-PhD (Neuroscience), Albert Einstein College of Medicine, 1982.

**Torvik** - PhD (Engineering Science), Louisiana State University, 2002.

Please provide a photo of yourself (you have these)

and a link to your home page  [http://arrowsmith.psych.uic.edu](http://arrowsmith.psych.uic.edu)

General Questions

*What is your main interest in attending the workshop?*

Author name disambiguation

*What is your main interest in 'data sharing' and/or 'data integration'?*

this is a central interest of the Arrowsmith project, and several of our tools are devoted to these goals
Scholarly Datasets and Data Integration Challenges

Please provide information on the datasets you serve/use in your research.

**Dataset/Project Description** - various customized text and genomics databases

**Preservation** - what processes or procedures are in place to keep data useable over time? Periodic updating

**Provenance** - what processes or procedures are in place to insure that the data contributed to the project is authentic and authoritative?

**Privacy Policy** - what policies are in place to insure the privacy of contributors and/or end researchers?

**Data formats / types** - what type of data is stored and in what format?

**Database Technology** - what database management system is used (Oracle, MySQL, etc)? MySQL

**Storage Technology** - what hardware is used to store the data?

**Backup Strategy** - with what frequency is the data backed up? Are copies kept off site? Do you have a disaster recovery plan? Some databases are replicated on multiple machines, and we have a backup hard drive for some data.
Partners - what other organizations participate in this project? 
Arrowsmith consortium involves investigators across 5 different universities.

Funding / Ownership - which organization owns this project, and how is this project funded? NIH

What data integration issues are you facing/addressing in your research? Please address policy and technical level.

What data integration solutions do you use/have you developed? 
Arrowsmith, Anne O'Tate, Author-ity.

Please provide references/links/pointers to relevant work, papers, and efforts. See References link on our homepage.