Brief Bio and (PR)²: Problems & Pitches – Raves & Rants by Mike Pollard

Biography (about 250 words)
Mr. Pollard has extensive managerial and technical consulting experience in the planning, review, analysis, design, construction, and implementation of information systems for both commercial and US Government organizations. His experience includes planning, managing, and reviewing full life cycle large-scale Federal Program Management Office (PMO) and Support Office initiatives, technology and system reengineering programs; IT Architecture (ITA) and Enterprise Architecture (EA) planning and management; IT Investment Management/CPIC and Business Case analysis; Federal IT Privacy and Security Life-Cycle Management, Internet, Intranet, Extranet, Web, and Electronic Commerce technologies; and digitization projects.

Mr. Pollard is an expert in business intelligence and knowledge management and discovery. Mr. Pollard evangelized early adoption of XML as founder and CEO of venture-backed thinkXML. Pollard supported commercialization for Lockheed Martin’s Aerotext (www.aerotext.com), an important text mining and knowledge discovery tool. Mr. Pollard currently leads Discovery Logic’s Knowledge Discovery (KD) Practice.

Mr. Pollard has supported the launch of several new initiatives for the US National Institutes of Health (NIH) including the NIH Undergraduate Scholarship Program, the NIH Training Website (www.training.nih.gov), the NHGRI Mouse Genome Website, and transformation of the NIH Loan Repayment Programs to an automated office. Additionally, Mr. Pollard has published in peer review journals and served as a board member for technology companies.

General Questions
What is your main interest in attending the workshop?
- We solve business-critical problems for our clients.
- Our most critical data revolves around people.

What is your main interest in ‘data sharing’ and/or ‘data integration’?
- Helping our customers make logical decisions.

Scholarly Datasets and Data Integration Challenges

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

Dataset/Project Description
CRISP, Medline, News, USPTO, ClinicalTrials.gov, and similar databases.

Preservation – what processes or procedures are in place to keep data useable over time?
- Weekly updates for all major databases.

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?
- All data is in a database
Database Technology – what database management system is used (Oracle, MySQL, etc)?

**Microsoft SQL Server 2005.**

Storage Technology – what hardware is used to store the data?

**Dell Storage Area Network (SAN)**

Backup Strategy – with what frequency is the data backed up? Are copies kept off site? Do you have a disaster recovery plan?

**Yes**

Partners – what other organizations participate in this project?

**Not Disclosed**

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

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?

Please provide references/links/pointers to relevant work, papers, and efforts.

*Please send the completed document by August 22, 2006 to Katy Borner <katy@indiana.edu> and Stacy T. Kowalczyk <skowalcz@indiana.edu> using the subject header ‘ScholarlyDB&I’.*