BARRIERS

NATIONAL VS INTERNATIONAL

Scientific knowledge discovery efforts exist at the national and international level, e.g. the virtual observatory in astronomy. Ad-hoc data standards and protocols are sometimes developed and adopted at the national level

INTEROPERABILITY

Infrastructure solutions to scientific knowledge discovery often do not interoperate. Overall, a Google-like search and management service for scientific data is not in place.

COST

Data hosting, management, mining as well as cloud computing solutions are expensive.

PUBLISHABLE ARTIFACTS

What is it ok to publish? What should be published? We need to consider data, software, algorithms, and other supplemental materials as publishable artifacts

PROPRIETARY ISSUES

Some data may be locked (e.g., rich genomic data owned by pharmaceutical companies) while other may be entirely unlocked (e.g., most astronomical data). Even in the locked scenario, allowing data annotation, adds value to the data and makes it reusable.

MANY EFFORTS

Multiple solutions and tools for scientific knowledge discovery and data management exist. These efforts are either not connected to one another, or if they are, they are connected in a variety of ways.

DATA CITATION

We have no standardized, widely-adopted mechanism to cite scientific data.

SOCIO-CULTURAL BARRIERS

How do SKD researchers fit in within the academic/industry job market?

REWARD STRUCTURE

The role and shape of the academic paper is changing, as well as reward and recognition systems (citation, authorship, etc.). Today there are no clear incentives to change in the direction of open scholarship (often, it conflicts with the culture of independence and competition)

COMPUTER SCIENCE CURRICULUM

To refocus scientific disciplines and their curricula around computer science and data issues will allow not only to solve problems a lot more quickly and in an automated fashion, but also to ask novel questions

SCIENCE IS NOT THE BIG KID ANYMORE

World wide web data (e.g. user-generated content on social networking sites) is also very large and very well funded

DATA SEARCH

Indexing, searching and retrieving scientific data is hard

KNOWLEDGE EXTRACTION

Extracting hypotheses and results from scholarly articles is hard

DIVERSITY

Computer-mediated scientific knowledge discovery changes significantly between different communities, disciplines, and institutions. Barrier or opportunity?

DISCIPLINARY VS INSTITUTIONAL

Should data and literature repositories be disciplinary or institutional? Enabling scientific knowledge discovery across different platforms and repository philosophies is crucial.

PHYSICAL BARRIERS

Distance matters: large cyber-infrastructure initiatives imply remote computer-supported collaboration, yet scientific output (and trust) often exists (and improves) with face-to-face interaction

TEMPORAL BARRIERS

Software ecologies and metadata requirements change with time. What you collect today may not be what you want tomorrow.

LAW DOES NOT HELP (MUCH)

Difficulty of achieving simplicity(??). The legal system is designed to promote securing things, not sharing knowledge.

REPRODUCIBILITY

As scientific enterprises become progressively more computational, data and code sharing at the time of publication is crucial to enable reproducibility of scientific results.