The Intellectual Property Management for Data Sharing in a German Liver Cancer Research Network

Shan He, Matthias Ganzinger, Petra Knaup
Institute for Medical Biometry and Informatics
Medical Informatics Section
Introduction

- efficient and reliable data storage
- platform for data sharing
- benefits
  - efficient use of research resources
  - avoiding duplicate experiments / trials
  - gathering adequate sample sizes
  - rapid release of prepublication data
**SFB/TRR77 - Liver Cancer**

**From Molecular Pathogenesis to Targeted Therapies**

**Research Area A**
From Chronic Liver Disease to Tumor Formation
- Project A1
- Project A2
- Project A3
- Project A4
- Project A5
- Project A6
- Project A7
- Associated Project

**Research Area B**
Basic Molecular Mechanisms / Novel Targets
- Project B1
- Project B2
- Project B3
- Project B4
- Project B5
- Project B6
- Project B7
- Project B8

**Research Area C**
Novel Approaches to Therapy
- Project C1
- Project C2
- Project C3
- Project C4
- Project C5
- Project C7

**Project Area Z**
Central Projects of the SFB/TRR77
- Project Z1
- Project Z2

TRR 77 is funded by the German Research Foundation (DFG)
www.livercancer.de
Methods

- survey among all 22 projects of network (TRR)
  - potential of data sharing
  - concerns
- interviews with project leaders
- literature review on promoting data sharing
- development of a concept within a Service Oriented Architecture (SOA)
Results

• surveys and interviews
  – current data sharing patterns
    • known colleagues
    • negotiated on an individual basis
    • based on mutual trust and interest
  – concerns
    • data confidentiality
    • access control issues

→ difficulty of getting proper acknowledgement for data contribution
Results Survey (n=19, 14 projects)

Who would be interested in your data?
- No answer: 21%
- All TRR projects: 22%
- A number of TRR projects: 57%

Would you provide your data in a common TRR information system?
- Yes: 44%
- Only if my confidentiality requirements are met: 56%
- No: 0%
Results Survey (n=19, 14 projects)

Are you concerned about the confidentiality of your data?

- Yes, TRR on request: 67%
- No, anybody can access my data: 6%
- No answer / depends: 11%
- Yes, no sharing: 11%
- Yes, access for all TRR: 5%
- Any answer: 8%
Literature Review

- intellectual property issues of data sharing
  → not unique in biomedical domain
- previous efforts in biomedicine mostly policy level strategies → prerequisite for funding
- mandatory data sharing does not address major concerns → give credit to investigators
- Idea in literature: share metadata only → not distribute the data set itself
Key System Requirements for Intellectual Property Management

- enable data producers to track who is accessing their data *for what purpose*
- providing data consumers a formal and easy way to acknowledge data producers
- allowing impact of data set to be tracked
- establishing policies that recognize and reward data producers
Metadata Collection

• metadata for
  – data queries from data consumers
    • project description
    • research question or hypothesis
    • expected timeline
  – data sets generated by data producers
    • brief description of data source
    • related data sharing policies
      → e.g. ‘Are review process and approval required before data are released to requester?’
Data Citation

• review process for sensitive data sets
  → approve or deny data requests
• data consumer has to agree on
data citation statement
• citing data in the same way
  as articles and books
  – digital object identifier (DOI) allocated
  – standard metadata scheme for
    proper data citation
Incentives for Data Sharing

• tracking the impact of a data set via unique ID
• establishing policies that recognize and reward data producers
• high impact data set may be basis
  – funding renewal
  – searching for financial support
The diagram illustrates a workflow for data registration and access. The process begins with a Digital Object Registry, where a data set is registered with a DOI. The metadata is then retrieved and can be accessed through the TRR pelican Server. The data access permission is requested by the data consumer, and the access logs are audited. If permission is granted, the data and related metadata are retrieved, and the data citation statement is generated. The data producer grants or rejects the permission.
Discussion and Future Work

- Integrate data citation idea from other domains to TRR
- Novel approach
  - Easy data citation
  - Metadata collection
  - Incentive strategies
- Not sufficient to avoid intentional data plagiarism
Take home message

• data sharing in biomedical research
• researchers are concerned about getting credit
• IT platforms for research networks should support
  – metadata for
    • queries
    • data sets
  – easy data citation
  – tracking impact of a data set