

Summary of the Enabling Data Sharing in Cancer Research Workshop

Cancer Research UK, Holborn, London

Friday 30th November 2007

Background

Since its formation the NCRI Informatics Initiative has endeavoured to address the issues surrounding data sharing and integration in cancer research. In 2003, a NCRI Data Sharing Group was formed who subsequently developed a data sharing policy that was formally adopted by the NCRI Board in July 2005.

The NCRI Data Sharing Policy¹ served as the first step in encouraging wider access to data generated in cancer research by encouraging the NCRI organisations to introduce a requirement for scientists to include data sharing strategies in funding applications and for provision to be made by funders to support the cost of data sharing where necessary.

Since the adoption of the data sharing policy the NCRI Informatics Initiative has been working with member organisations to implement data sharing through developing individual data sharing policies and guidelines that set a clear direction to allow scientists to carefully consider data sharing when applying for funding. The successful implementation of data sharing will allow the scientific community to maximise the potential of their research by:

- Reinforcing open scientific inquiry
- Encouraging diversity of analysis and opinion
- Promoting new research
- Enabling the testing of new or alternative hypotheses and methods of analysis
- Facilitating the education of new researchers
- Enabling the exploration of topics not envisioned by the initial investigators
- Avoiding unnecessary duplication of data gathering
- Permitting the creation of new data sets when data from multiple sources are combined.

Currently the NCRI member organisations are at different stages of the implementation process and the 'Enabling Data Sharing in Cancer Research Workshop' was organised to bring together as many of the NCRI organisations as possible to provide a mechanism to review progress in the implementation of the data sharing policy, discuss best practices and share experiences.

Summary of Presentations

Welcome and Introduction

Abi Ajose-Adeogun from the NCRI Informatics Initiative provided an introduction that covered the current international data sharing activities and the work the NCRI have been doing to encourage data sharing in cancer research.

Different Approaches to Data Sharing

This session of the workshop provided an opportunity for attendees to hear from those organisations that have carried out a great deal of work in data sharing. It allowed for an in-depth look at their data sharing strategies and how they plan to sustain data sharing in the research that they fund.

¹ http://www.cancerinformatics.org.uk/Documents/DATA_SHARING_POLICYapr2006.pdf

Allan Sudlow from the Medical Research Council (MRC) gave an overview of the MRC's Data Sharing Initiative which started with the drafting of their data sharing policy, continued with the formal adoption of the policy by the MRC and currently involves their implementation of that policy. He stressed that throughout their initiative the MRC consulted widely with researchers, funding partners, study participants and other stakeholders to ensure that all those who would be affected by their data sharing policy had the opportunity to contribute their views. He also touched on some of the attitudes and practices within the research community that they had to consider while putting together their policy and then outlined the MRC's Data Sharing Policy Principles and Data Access Principles. Finally, Allan covered the MRC's ongoing work to support their scientists in their data sharing efforts, this included:

- Web based resources.
- Pilot projects to enhance data curation and access for wider research use.
- Establishment of a Data Support Service.

More information about the MRC's Data Sharing Initiative can be found at:

<http://www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing/index.htm>

Alan Doyle from the Wellcome Trust then provided an overview of the work that led to the development of the 'Wellcome Trust Policy on Data Management and Sharing'. Their data sharing activities stemmed from the Fort Lauderdale agreement of 2003 which encouraged the rapid release of sequence data to public databases and recommended that this principle of rapid pre-publication release should apply to other types of data. With this in mind the Wellcome Trust developed their policy which aims to maximise the availability of data with as few restrictions as possible. Scientists applying to the Wellcome Trust for funding are now expected to produce a detailed data management and sharing plan with their applications which will be subject to scrutiny and review. The Wellcome Trust have also put together a set of guidelines on good research practice which scientists can refer to when putting together their data management and sharing plans.

Alan then covered some of the issues that they had to tackle to ensure that scientists' concerns were addressed, these included:

- The cost of data sharing
- The timing of the release of data
- Intellectual Property
- Ethics Approval
- Resources for the scientific community
- Responsibility of users – terms and conditions of data use

Following this Alan then mentioned the development of their Data Access Policy which is based on the recommendations of the Lowrance Report published in March 2006 and finally he mentioned that the Wellcome Trust will be extending their data sharing activities internationally. This is due to the fact that 10% of their funding focuses on research conducted overseas. This initiative will have the added challenge of overcoming cultural differences and one of the ways they are hoping to address this is by actively engaging with the local scientific community through workshops.

More information about the Wellcome Trust's data sharing activities can be found at:

<http://www.wellcome.ac.uk/node3309.html>

Amanda Collis from the BBSRC started her talk by providing some background information about the BBSRC and the type of research that they fund putting in to context what their requirements were when developing a data sharing strategy. She stated that the development of the BBSRC's data sharing policy was based on four general principles:

- Data generated from public research investment should be publicly available
- BBSRC is committed to obtaining the best value for the funds we invest
- Data sharing reinforces open scientific inquiry
- New bioscience research methods mean increasingly large amounts of research data.

She mentioned that the BBSRC's approach to data sharing was that it should be community-led and driven by scientific need so their role has been to encourage, facilitate and stimulate data sharing. Therefore, in the early development stage of their data sharing policy they actively and extensively consulted with all interested parties to ensure that whatever policy the BBSRC adopted would meet the needs of their researchers.

The BBSRC's Data Sharing Policy Project identified that a data sharing policy alone would not be sufficient in ensuring success in data sharing, there needed to be a plan for implementation and sustainability. Therefore not only did they focus on developing a data sharing policy but also an implementation plan, both were formally adopted by the BBSRC's Strategy Board in March 2006. The formal launch of the data sharing policy was early this year and the final policy was based on the information received from the consultation process and also on the OECD guidelines. The formal launch saw the release of two documents a high level Data Sharing Policy Statement, which should be read in conjunction with the second document, the Implementation Guidance.

Amanda then went on to say that there were two themes to the implementation of the policy:

1. The integration of data sharing into the funding process.

Applicants applying for funding have to include a detailed statement on data sharing in their grant proposals and this would be considered separately from scientific merit. Data sharing would be monitored by a Data Sharing Policy Monitoring Group and scientists will be graded on their compliance and any non-compliance will affect their future funding.

2. Facilitating and encouraging data sharing in the bioscience community through the provision of:
 - Funds to support data sharing
 - Information and guidelines
 - Training

Following the launch of their data sharing policy earlier this year they have received their first data sharing statements with a number being of a high quality which has been very encouraging.

More information about the BBSRC's Data Sharing Policy can be found at:

http://www.bbsrc.ac.uk/funding/news/2007/0704_data_sharing.html

Summary of the discussion session

A round table discussion was convened which included the speakers and workshop attendees and was facilitated by Allan Sudlow. The aim of this session was to identify the current challenges that NCRI member organisations that are yet to implement the NCRI Data Sharing Policy were experiencing and to identify ways of overcoming these challenges.

The following challenges and issues were identified:

- How to enforce data sharing
- Grants are short lived so when should the data sharing responsibility for funders end
- Intellectual property issues
- Concerns over supporting the cost of data sharing
- How should compliance be monitored
- Technology is needed to implement data sharing – who should provide this?
- There is a general reluctance from scientists to share data until it has been published – how do we encourage scientists to share their data?
- Ownership of data is an issue – who ultimately makes the decision about the data?
- Infrastructure issues
- Concerns about the quality of data that is to be shared. There needs to be some quality control – who is responsible for assessing the quality?
- Decisions on what data should be shared. Is it necessary to share all data? Who makes this decision?
- How to deal with the resistance in the research community to properly archive data
- Lack of trust – scientists doesn't trust that their data will be used wisely; they feel that data will be mined without a hypothesis which is not useful.

The following suggestions were offered on how to overcome some of the challenges:

- The process of sharing data needs to be practical
- Ensure that there is a scientific need for the data being shared
- More archives need to be developed and the responsibility for maintenance decided
- Study specific policies should be developed as a lot of policies are too general and don't meet the needs of every scenario
- Give scientists credit for sharing their data
- Data mining should be in a controlled environment
- Data should only be shared with a group that has a plan to enhance that data

The group were asked how the NCRI Informatics Initiative could assist and the following responses were received:

- A demonstration project that provides examples of good practice and how scientists can share their data
- Develop robust cost models to share data
- Facilitate the formation of a data sharing consortium
- Training courses for charities and small funding agencies
- Case studies with scientists that have physically shared their data
- Organise a meeting in the next 9 months with NCRI organisations to assess the data sharing plans submitted by research scientists

Summary

The Enabling Data Sharing in Cancer Research Workshop provided a forum for the NCRI member organisations to come together to:

- Learn from each others experiences
- Discuss the challenges of data sharing
- Identify ways of overcoming these challenges.

The NCRI Informatics Unit will continue to work with the NCRI member organisations to ensure the data sharing become the norm in cancer research.