**I don’t know you, so why should I trust you with my data?**

**Gavin Kennedy1,2**

1The Australian Phenomics Network, Canberra, Australia, gkam08@gmail.com

2The Australian Plant Phenomics Facility, Canberra, Australia, gkam08@gmail.com

**OVERVIEW**

At the recent Integrated Biological Systems (IBS) Phenomics Data and Informatics workshop1, key representatives of the biological research, bioinformatics and infrastructure communities came together to discuss the data management needs of the emergent phenomics discipline, and of the broader research community. Whilst issues of infrastructure were presented and discussed, some of the most important discussions were around the issue of trust within the eResearch community, and how the lack of an effective trust framework is an effective barrier to data management, data publication and data reuse.

Public research data policy and practice place an expectation upon researchers that they will freely give up their data, and seeks enforcement through either rewards or punitive measures2. The notion of trust, i.e. that a researcher trusts others to use their data ethically and responsibly, or that a researcher will trust the quality of the data provided by others, is rarely explicitly addressed by the research institutions, the funding agencies or the eResearch infrastructure builders. A trust framework, therefore, is often missing from the value proposition put to researchers when expecting them to comply with data management policy and codes of practice. The danger of policies and practices without trust frameworks is that researchers will feel exposed when publishing or sharing data, and consequently the data infrastructure is under-utilised and fails to provide significant returns on investment.

This birds of a feather session seeks to build upon and broaden the outcomes of the phenomics workshop by providing a forum in which researchers from all domains, both data consumers and data generators, can pose the problems of trust to the eResearch practitioners.

Specifically the session would seek to:

* Identify the dimensions of trust; what trust means to different research and eResearch practitioners.
* Prioritise the dimensions of trust; consider what are the most important in establishing a research data trust framework.
* Look at trust as a product of the researcher’s environment, the researcher’s collaborative network and the data management infrastructure;
* Consider trust as a by-product of not only the quality of the data, but the adherence to metadata standards and practices and the quality and longevity of the infrastructure designed to manage and transmit the data;
* Consider the relationship between data management policy and practice, whether ascribed by the researcher’s institution or funding agency, and the development of data quality and trust;
* Identify how to use a trust framework to increase the value of good management practice to researchers;
* Discuss whether data publishing and data citation mechanisms are valid and measurable means to engender trust; and
* Look at social networking in a research data environment3 and how trust can be built up through social networking tools and/or collaboration tools and services.

**Outcomes**

This session is intended to facilitate discussion about the role of trust in the formulation and application of data management policy and practice.

The intended outcome of the discussion is to formulate a set of key statements that would be considered by research institutions in the formulation of data policy and the procurement of data management systems, as well as by the eResearch community in the design and development of tools, systems and services.

The key statements ensuing from the discussion would be tabled and a follow-up report provided to the participants, the phenomics facilities and published online for other institutions to utilise.

The intention is that these key statements would be reviewed by the phenomics community in order to develop and implement data management policies, systems and practices. For example, the phenomics community would utilise these statements as guidelines for the exemplar data publication processes they wish to implement as part of their infrastructure development.

**Audience**

This session is aimed at the “coal face” researchers and data generators, bioinformaticians and eResearch practitioners who wish to negotiate data management policies and practices or at least consider these issues within their own domains. Whilst it is a follow on from a phenomics community initiated discussion, the intention is to broaden the scope and gather input from other research domains.

**Format**

The intended format is to have two or three oral presentations that cover issues of trust, policy and practice, and provide examples where trust has been a barrier to online data publishing or where trust has been successfully established and led to successful outcomes for the original investigators and the data re-users. The presentations would then be followed up with a chaired discussion.

**References**

1. Integrated Biological Systems Phenomics Data & Informatics Workshop 2010 Available from: <http://www.plantphenomics.org.au/IBSPDIW2010>
2. Schofield, P.N., et al., *Post-publication sharing of data and tools*, Nature 2009 **461**, pp171-173.
3. Dyer, C., *Science is not social networking.* Eur J Immunol. 2009 **39**(12) pp3276-7.