# Managing sensitive data across the data life cycle

**David Eyers[1] and Russell Butson[2]**
[1]University of Otago, Dunedin, New Zealand, david.eyers@otago.ac.nz
[2]University of Otago, Dunedin, New Zealand, russell.butson@otago.ac.nz

## Background | Rationale

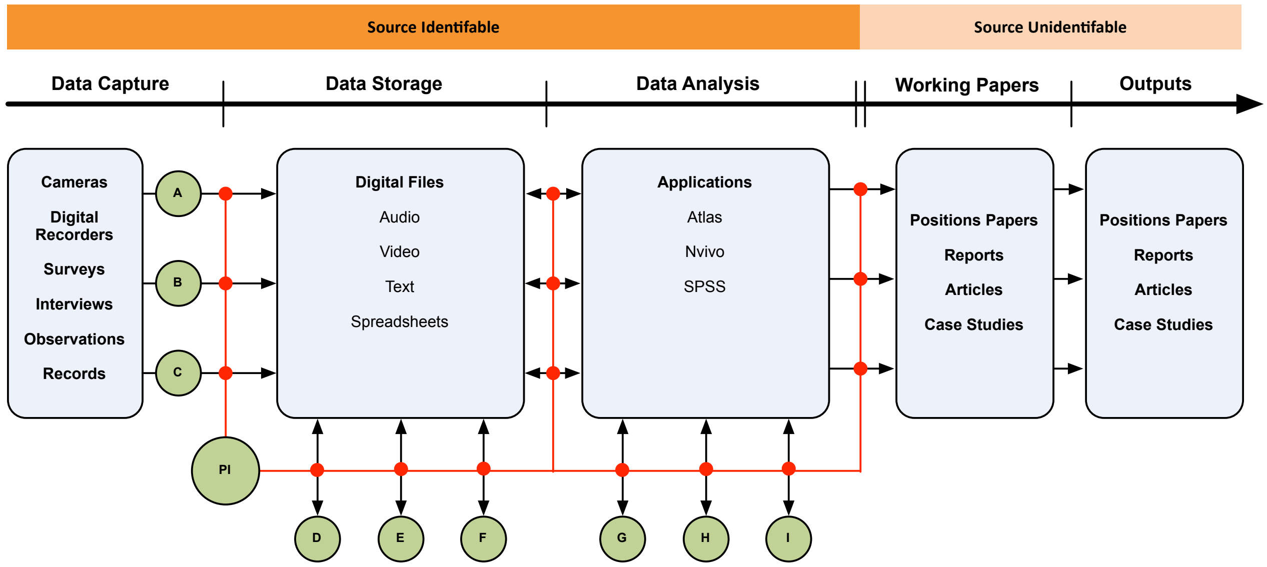
The management of and responsibility for raw data is a central aspect of empirical research. Most traditional texts on the practice of research have sections outlining various approaches for categorizing, storing and recovering information from concrete artefacts like paper and tape. While it would be difficult to say if the arrival of digital media has made this process any easier, it may well be that for many the adoption of the more abstract digital media meant a shift from autonomous control to one reliant on technologists. Over the years researchers have become more self-reliant through the widespread use of personal computing. The proliferation of cheap, high-capacity storage technology has made it possible for researchers to store large amounts of data. However, the ethical responsibility on principal investigators requires the management of raw data beyond the task of storage: particularly in the current climate of collaboration. Many types of research projects require collaborative sourcing, management or sharing of sensitive datasets. Often researchers make do with an ad hoc approach to sharing data, without fully appreciating (or even considering) the risks involved, often because of the perceived inaccessibility of higher quality solutions.

While many institutions are addressing the demand for improved joint access to raw data through centralised data storage schemes, these innovations are being driven by technologies that are focused solely on data storage. Our goal in this presentation is to show a different approach to designing managed storage spaces, based on researcher workflows of the capture, storage and analysis of highly sensitive data (Figure 1).

Provided that eResearch tools can present a convenient (or at least intuitive) interface to researchers, and provided that there is a route to fund an independent technology host, there are many benefits to be gained from offloading research data into a layered storage infrastructure. However, the significant economies of scale to be gained in terms of having backup and on-line redundancy of physical media managed independently from research data create difficulties in cases where the repositories contain highly sensitive data. In these contexts the technology host that was previously able to remain agnostic to the application specifics of researchers, now must partition their infrastructure in a complementary manner in order to provide security assurances.

## The Project

The University of Otago is currently exploring the implications of developing a secure storage capability that aligns with the workflow needs of researchers working with patient data within the healthcare sector (Figure 1). The project aims to achieve a workable model and a set of guidelines for controlling the access, storage, retrieval, replication and analysis of highly sensitive data within a secure environment.



**PI** = Principal Investigator/s (responsible for data management under the project ethics approval) **A-I** = Various project members ● = PI management of access-transfer-replication

**Figure 1:** Blueprint for Management of Sensitive Data—Researcher perspective

The underlying technology of the data storage component is a typical enterprise-grade SAN (Storage Area Network). The SAN is managed by the central IT Services division (ITS) of the University of Otago. It is the responsibility of ITS to ensure that the storage layer is insulated from hardware failures of the hard disks and network fabric. Typically the storage within a SAN will be divided up into security 'zones'. However, for our use this would involve an inconvenient coupling of ITS system administration with the dynamic needs of the research projects using this storage. Instead, a storage middleware is deployed to effect decentralisation of the access control to the data.

This project employs the **i**ntegrated **R**ule **O**riented **D**ata **S**ystems iRODS [1] middleware to provide the required level of abstraction between the researchers' needs and the infrastructure provided by ITS. At this layer, we introduce the capability for the research project PI to map their project members to project resources. The research workflow can inform the PI of the exposure that they have, care of their project members, to the data that they are legally responsible for.

Our goal is to ensure that iRODS can provide a security layer that is at least as expressive as standard Role-Based Access Control (RBAC) [2] models. This facilitates allocation of security privileges on the basis of functional and organisational roles, avoiding the difficulty of maintenance caused when privileges are linked to individual users. RBAC provides a straightforward basis on which to implement more advanced functionality, such as the delegation—and subsequent revocation—of access rights.

## CHALLENGES

The project poses numerous challenges. First, there is the need to meet research user requirements. Some participants in the project will be unwilling to move away from their current data management practices, even if these practices are likely to fall short of documented project requirements. Second, there is the technical problem of security surface area—i.e. how much of the software is critical to maintaining security. Traditional, comparatively static configuration of access permissions, for example at the SAN level or at shared filesystem volumes directly above it, will have a smaller surface area than this project's infrastructure. That is because our security layer is added on top of the existing access control layers, and will compute whether access requests are appropriate based on its own policies. We will need to determine the level of assurance that we can provide over the software's behaviour.

In this presentation we will discuss the importance of the data lifecycle, introduce our current blueprint, describe the challenges and share lessons we have learnt within the design phase and initial implementation of work. We will conclude with an outline of where we see likely future opportunities and challenges.

## REFERENCES

1 Introduction to rods. From http://www.irods.org/index.php/Introduction_to_iRODS

2 Ferraiolo, D., & Kuhn, R. (2009). Role-Based Access Controls. From http://arxiv.org/pdf/0903.2171

## ABOUT THE AUTHOR(S)

**David Eyers** is a lecturer in Computer Science at the University of Otago. He does research into wide-area distributed systems, with particular interests in security and efficient network communication. Security topics of relevance to this project include distributed access control systems, and data management techniques that facilitate the effective tracking and protection of sensitive information. David is interested in the evolution of policy within access control systems, i.e. their ongoing management, in addition to what they are able to enforce at any point in time. As academic projects in the aforementioned fields can sometimes drift away from the actual needs of users, David is keen to undertake projects that examine how such research topics can be grounded in reality. Undertaking eResearch projects often involves researchers being required to incorporate and evaluate new technologies, and thus is a particularly appropriate target community for his research.

**Russell Butson** is a lecturer in Higher Education at the University of Otago. His primary area of interest is the practice of collaboration, particularly within virtual research environments. The emerging nature of this field has meant that Russell has had to design and create a number of virtual research environments in order to advance his research area. While there is a technical component to this work, the actual research is more sociological than technical; with an emphasis on the nature of knowledge working through the analysis of the environments, communities, and practices involved in the practice of researcher collaborations. Russell has been in the eResearch domain for some years, promoting the uptake of ICT to support the practice of research both nationally and locally at the University of Otago. He is a member of the University of Otago's eResearch Advisory Group and an advocate for eResearch generally.