

# A Data Network for Health e-Research \*

Kerry Taylor, Christine M. O’Keefe, John Colton,  
Rohan Baxter, Ross Sparks,  
Mark Cameron, Laurent Lefort  
CSIRO ICT Centre  
firstname.lastname@csiro.au

Uma Srinivasan  
PHI Systems  
uma@phisystems.com.au

**Abstract:** Sharing health data for research purposes across data custodian boundaries poses technical, organisational and ethical challenges. We describe a service oriented architecture for a proposed Health Research Data Network (HRDN). The HRDN architecture supports services to manage data access and use by researchers in accordance with individual data custodian policies.

The capabilities of the HRDN architecture are described using a layered service model. The four abstract layers from lowest level to the highest level are 1) *Preparing*, 2) *Storing*, 3) *Sharing* and 4) *Using*. Two additional groups of services are interfaced with the services in each of the four layers. They are 1) *Describing*, with services for collecting and managing metadata, and 2) *Protecting*, with services for ensuring confidentiality and privacy protection, as well as services and tools implementing information security functions. In addition to these HRDN service groups, client-side applications are used by data custodians, service providers and researchers. Following a reference implementation of most services and the Researcher’s Workbench, a commercial version of the software has been developed and is being trialled.

## 1 Introduction

The Health Research Data Network (HRDN) is a collection of data and software services, connected via a high-bandwidth communications infrastructure and standardised interfaces, enabling use of data collections by authorised participants.

The overall HRDN goal is to minimise the time to insight for answering health and community sector research questions. Although rich and growing data collections already exist in Australia, the task of performing any kind of analysis across them is beset with technical difficulties, organisational and ethical constraints. For example, the data collections are geographically dispersed, are built on a variety of technologies and have differing data quality related to their varying primary purposes. Data custodians have ethical and legislative requirements for managing research use of their data. The HRDN design addresses the technical difficulties in information integration, while supporting complex organisational and ethical requirements arising from use of diverse health data sources.

The two key HRDN features of *custodial control over access and use of resources* and

---

\*This work was conducted at CSIRO. Rohan Baxter is currently affiliated with the Australian Tax Office.

*privacy protection integrated into a secure end-to-end system for data sharing and analysis* distinguish HRDN from other service oriented architectures for distributed data sharing and analysis. In HRDN, data custodians have the ability to specify access and use policies for each data collection: who may access it, what can be done with it, what sort of results can be released, for what purpose and with what approvals, conditions and obligations.

The paper starts in Section 2 by using a Researcher’s Workbench example to motivate a key subset of the HRDN capabilities. In Section 3, the capabilities are described using a layered service model, with increasing capabilities offered by higher layers building on lower ones. Section 4 discusses the architectural principles for the design of component services. The network services and their interfaces are described in Section 5.

## 2 Using the HRDN

This section provides a motivating example showing the HRDN benefits for a researcher. In this example, we use the names of HRDN network services without defining them.

Consider an epidemiologist, Emma, who is interested in studying the relationship between the post-operative survival times of colo-rectal cancer patients and their MYC gene amplification level [AWC<sup>+</sup>97].

**Researchers:** Emma must first obtain HRDN membership through the *Member Registration* service. Emma then uses the Researcher’s Workbench to browse available resources, including the metadata descriptions of the available data collections. We assume Emma finds some useful resources and develops a detailed research plan. Emma may need to apply to data custodians for specific access to their data for the purposes of her research and receive the data custodians’ individual access agreement(s) for signing via the *Agreement Facilitator* service. Emma’s project involves access to individuals’ medical data and so will require approval from one or more Human Research Ethics Committees (HREC).

If the data custodians are satisfied and the approvals are granted, Emma designs the MYC data resource (or model) for her study using the Researcher’s Workbench. The model includes mappings between her purpose-built data resource and the selected existing data sources to be accessed. For example, source *A* may have details of the patient clinical care events, while source *B* contains the Births, Deaths and Marriages Register and source *C* is Emma’s own register of MYC amplification levels for patients. Obviously, Emma only requires a subset of data from each source for her project. The mappings describe the required subsets and any desired transformations (e.g. from date of birth to age).

Emma uses the *Exploratory Data Analysis (EDA)* service tools in the Researcher’s Workbench to understand data coverage and quality issues, which may effect the validity of her intended analysis. The EDA service interacts with the *Disclosure Risk Estimator* and *Disclosure Risk Mitigator* to ensure presented results do not disclose confidential or personally identifiable information unless Emma has specific approval.

Emma’s model requires linkage of the records for individuals from three data sources. The workflow in the development of linking methods and any other aspects of the MYC

data resource developed by Emma become part of her data model specification. The Researcher's Workbench records Emma's data model specification for any future audits, for verification of results and for re-use by Emma and other authorised HRDN users.

Eventually, Emma executes the resultant queries for her MYC data model. The Researcher's Workbench interacts with the *Planner* and *Orchestration* services to plan and then schedule requests for data from the three sources *A*, *B* and *C*, linking and other transformation services. The resulting data is stored remotely in a secure cache.

The data custodians' conditions of use constrain the invocation of HRDN services as Emma calls them to explore the cached data. Within the imposed constraints, Emma explores the distribution of her response variable (survival time) and the relationships between the response and potential explanatory variables, such as age, gender, comorbidities and MYC level. All the while, disclosure risk assessments are made and confidentiality transforms invoked, before any results are presented to Emma. Emma will refine the MYC data model and her choice of analytical models. Version control will document iterations of the project workflow, to enable reuse of her analyses. Emma can choose to share her data models and project workflows with colleagues and the broader HRDN community.

**Benefits and Costs:** The benefits and costs for Emma arise from the key consequence of using the HRDN: Emma never sees any confidential data unless approved and authorised to do so. Emma's research plan, as submitted to the relevant Human Research Ethics Committees (HRECs) and data custodians, will emphasise the ability to answer the research questions without obtaining any confidential data as a part of the HRDN research approach. We expect that this will allow wider data access approvals and faster research plan approvals than is currently available. Three examples of the benefits are now given. In some cases, data governance requirements, whether legislative or organisational, may enable remote access and use of confidentialised unit record files, with strict limits on the number of such records which can be downloaded [ABS06]. HRDN provides the infrastructure for data custodians to achieve this. In other cases, data custodians are prepared to share the data for linking provided the identifying data is stripped before delivery to the researcher [KBH02]. HRDN also provides the infrastructure to do this. The third example is where a data custodian can release data for research purposes, but has had negative experiences in the past of apparently bona fide researchers using the data for commercial purposes contrary to the data release agreement. HRDN reduces the risks of abuse of data by requiring the remote use of the data. Emma's actions on the data are checked for consistency with the release policy.

A disadvantage for Emma is the need to perform data cleaning, linking and the resulting analyses remotely. This requires a change from the current practice of working in one's own local workspace with one's expertise in particular standalone analytical packages, such as SAS, Stata or SPLUS.

**Data Custodians:** HRDN offers some benefits to data custodians willing to share their data, too. Currently, they vet research proposals and, if they meet data release policy requirements, they physically release the data to a researcher. In HRDN, the data custodian's data release policy is present and enforced within the network, allowing an auditable trail of what is happening. In addition, confidentialisation and disclosure control services com-

bined with remote analyses of the data may allow research questions to be answered with data that otherwise would be unavailable due to privacy concerns.

Costs for providing HRDN infrastructure support could be amortised over the network participants. Currently, on-line access control and publishing is done on an individual *ad hoc* basis by the better resourced data custodians. Only a few data custodians, such as National Statistical Offices, are implementing remote data access and analysis laboratories.

### 3 Layer model

Our HRDN layer model in Figure 1 divides the functionality of an HRDN network into layers of successively greater functionality, with the top layer closer to an end user’s need. Higher level functionality depends on the lower level functionality, however not all layers will be needed for all users.

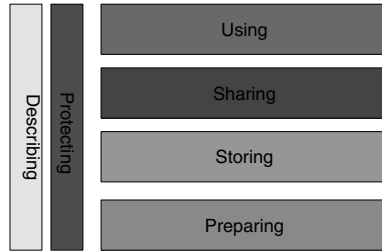


Figure 1: HRDN Architecture: Abstract Layer Model

The *Preparing* layer is about collecting data by data custodians. The challenges and software services required for managing data collections are out of the scope of the present project. The *Storing* layer is about adding layers of software and standards around the fundamental data held by a data custodian to enable its contribution as a network resource. The *Sharing* layer addresses the capabilities beyond enabling access to individual data custodian data resources. Sharing services provide capabilities for integrating resources offered by the *Storing* layer. The resources include software services, such as transformations, linking, and analytical functions, as well as virtual views of data and shared workflows. The *Using* layer supports value-added interaction with the network and includes user-oriented services that interface with the lower layers. Some groups of services deliver functions to multiple layers. The *Protecting* and *Describing* service groups are both presented as vertical bars rather than horizontal layers, because there are ‘protecting’ and ‘describing’ requirements at every horizontal layer of the network capability.

*Protecting* services provide the functions for implementing information security and privacy protection, partitioned into two sub-groups: 1) Information Security, which involves the protection of data from unauthorised access, modification and disclosure. In this sub-group, the content of the data and the purpose for which it is going to be used are irrelevant.

The tools used are standard information tools such as encryption, PKI, digital signatures, and digital certificates. 2) Privacy Protection, which involves protecting the privacy of individuals or organisations. In this sub-group content and purpose are critical. The tools here include a metadata language for expressing data custodian policies for data access and use, and disclosure control processes.

*Describing* services provide the functions for implementing metadata management. Metadata for these purposes is broadly interpreted as descriptive metadata that assists in finding and applying resources, but does not include metadata for access rights and privacy protection since that is handled in the *Protecting* layer.

**Relationship between Layers in the HRDN Layer Model:** A schematic representation of the relationship between the layers is shown in simplified form in Figure 2. The three square boxes represent the HRDN members: data custodians, researchers, and service providers. The two lowest layers, *Preparing* and *Storing*, interact only with the Data Custodian at the top of the figure. The highest layer, *Using*, interacts with the Researcher, but also the vertical *Protecting* and *Describing* services. A sub-layer within the *Sharing* layer is *Linking*. These are shown separately in the figure. The *Sharing* service publishes integrated data from one or more data custodians. It may use services from the *Linking* sublayer to do this. Outputs from both groups of services are provided to the *Using* layer.

## 4 Architectural Principles

We now describe the architectural principles used to guide the HRDN architecture design, which is described in the next section. A service-oriented architecture requires decisions on the services to be provided and the messages to be exchanged. At the heart of a service-oriented architecture is service composition planning and execution. Some architectural principles for the choice of services to aid service composition are: 1) *Services should be composable* 2) *Composition should be scalable* 3) *Composition should be fast* 4) *Composition should be flexible*. The desired composition properties just described are greatly promoted by an appropriate design of service types using the following principles: 1) *Adopt a common, reference model for service types*: Service capabilities, their interfaces, and the types used in message exchange, will be defined in a small number of uniform ways. Within this standard model, non-standard service instance-specific parameters may be supported through a universal parameter data type. 2) *Define a standard set of service features*: This would make it (syntactically) possible to substitute one service for another in composite workflows. A common type model would encourage economical and flexible data flows in composite services, and minimise the need for prohibitive data conversion technology. 3) *Canonical metadata*: Metadata is fundamental to supporting and attaining a semantically correct outcome for consumers of HRDN services. All HRDN services (whether they be custodian data services, analytical services or composed services) will provide canonical metadata. Finally, the services themselves should comply with available non-proprietary, open standards. Relevant Web Service standards used currently include XML, SOAP, and WSDL.

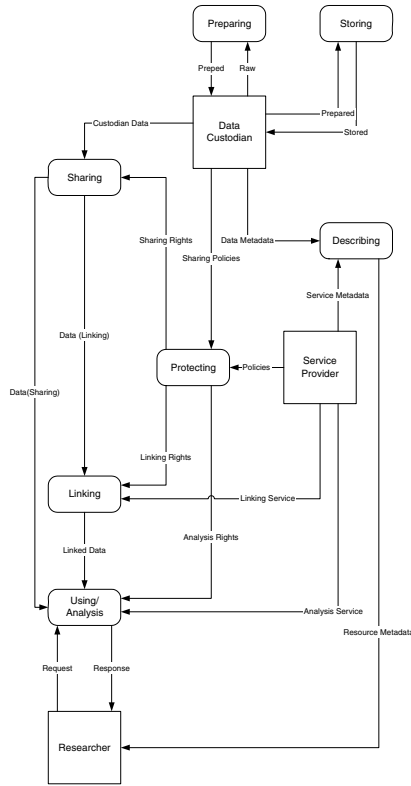


Figure 2: Schematic diagram of the main relationships between the layers in the HRDN layer model

## 5 Architecture

In order to realise the capability outlined in the layer model, the following basic services are proposed. Table 1 outlines these basic services and their major role grouping.

The provision of each of the services described in this section is subject to the user having been satisfactorily registered and authenticated, and having all the appropriate approvals and authorisations. These functions are supported by services in the *Protecting* layer, described below, where the interface between the two layers requires secure tokens to be included in the service requests. For brevity, we do not include all the details of the interactions between the services in the *Sharing*, *Using* and *Describing* layers and the *Protecting* layers in the descriptions of the services in this Section.

**Preparing and Storing:** The Preparing and Storing layers are not described in detail. The specific services available are specific to the data custodian and are not determined by the HRDN architecture.

**Sharing:** A *Data service* is the access point for data resources in the network. Primarily, a

Layer	Services
Preparing	Specific to data custodian organisation requirements and out of scope
Storing	Specific to data custodian, but typically a commercial or open-source database engine
Sharing	<i>Data Service; Orchestration Service; Cache Manager; Planner; Transformer service; Structure Registry.</i> <i>Linking sub-layer: Standardisation; Indexing; Comparison; Decision Classifier; Evaluation.</i>
Using	<i>Surveillance Analysis Service; Exploratory Statistical Analysis Service; Statistical Model Building; Analytical Information Management.</i>
Protecting	<i>Member Registration; Authentication; Authorisation; Agreement Facilitator; Disclosure Risk Estimator; Disclosure Risk Mitigator; Audit Logger; Audit Log Analyser.</i>
Describing	<i>Registry of Registries; Descriptive Metadata Service; Annotation Service; Metaflow Interpreter; Provenance Tracking.</i>

Table 1: Basic HRDN Services

data service accepts requests for data framed as a query against a data schema, and returns the requested data as a single message. Secondly, and optionally, a data service might also respond to requests for structural (schema) metadata and return that in a single message. An *Orchestration Service* is an infrastructure resource to manage multiple composite invocations of network resources. Its primary input is a workflow plan, which is represented as a graph of data-dependent service invocations. It will produce a composite message, which is the final result of the workflow execution, and an execution trace. A *Planner* accepts requests for data and data transformation services. Transformation services include linking and type transformations phrased as queries over a data integration schema, as described in [CT05]. These queries are presented to the planner. The planner produces a workflow plan. A *Cache Manager* offers a persistent and secure data storage service on behalf of clients. It manages both data and associated metadata that have been generated through interaction with network resources. At all times, it operates in concert with the *Authorisation Service* to ensure compliance with the custodian-specified conditions of use. It will be used together with the *Statistical Model Building* and *Analytical Tools* services to enable access to summaries or approved analyses of protected data. The persistent data is not directly accessible to the authenticated client as it is effectively placed inside a security and privacy screen that ensures the conditions of use are applied. A *Transformer Service* provides a method for translating between data types, data representations and standard coding schemes. It accepts an input message containing input data and a description of that data, and provides an output message containing the transformed data and, optionally, metadata describing the statistical properties of the transformation. A *Transformer Service* is represented by a description in the *Structure Registry* service, and, optionally, may itself respond to requests for its description. A *Structure Registry* service tracks formal metadata about data services and transformer services in the network. The linking services group are a sub-layer of the *Sharing* layer. The linking services have been derived from the components of a process model for linking consisting of the following six steps: data selection;

standardisation; indexing, also called blocking or clustering; comparison; decision modelling; and evaluation [BCC03, EVE02]. The resulting linking services: *Standardisation*, *Indexing*, *Decision Classifier*, *Comparison* and *Evaluation* are specialist transformer services, and adopt the same service interface. Data selection is provided by a *Data Service* or a *Cache Manager*.

**Using:** Analytical services operate in concert with the *Disclosure Risk Estimator* service and *Disclosure Risk Mitigator* service to ensure any specified confidentiality and privacy protection requirements are enforced. Data inputs are sourced from a *Data service*, or a *Cache manager*. The *Analytical Information Management* service takes data, control parameters and supporting data as input in order to generate metadata relating to the quality and fitness-for-purpose of data. Some specialist functions provide a transformed version of the data. The *Exploratory Data Analysis* service produces a range of graphical summaries of the data given some input data and control parameters. The *Surveillance Analysis* service provides some specialist analysis functions appropriate for health surveillance. The *Statistical Model Building* service provides ways for developing functional equations that describe how the response can be ‘best’ predicted from explanatory variables.

**Describing:** The *Describing* services can all be modelled as specialised data services in terms of their interface, but they perform a very different role in the network, according to the semantics of the data they store. Unlike data services, they are not themselves described in the *Structure Registry*. A *Descriptive Metadata* service delivers descriptive metadata about network resources. The descriptive metadata are viewed by users, so that the applicability of data resources to user needs are assessed. The service accepts requests for metadata phrased over a metadata schema, and returns a metadata description in one of a number of standardised schema formats. A single metadata schema is used for each metadata service. Metadata services describe both data resources and algorithmic services within the network, such as analysis and linkage services. A *Registry-of-Registries (RoR)* service is designed to respond to user-level interactive queries over metadata resources in the network. It handles interoperability over *Descriptive Metadata* services. It conforms to a request/response interface pattern and so can be used interactively via a web browser client. A *Metaflow Interpreter* service implements reasoning over formal metadata descriptions generated by *Descriptive Metadata* services in the network. Given a set of metadata documents as input, together with a specification, it produces a metadata document that corresponds to the application of the specification over the inputs. It is a specialist instance of a *Transformer Service*, operating over metadata documents. An *Annotation* service supports informal markup of HRDN resources by the user community. It accepts input of annotation remarks associated with content references; and, given a context reference, can respond with the relevant annotation remarks. It may be used when retrieving data or metadata to provide an extra level of descriptive metadata that reflects community opinions or practice. A *Provenance Tracking* service provides lineage tracking that helps to retrace the path of discovery and causality, by providing information about which processing intermediaries have touched the data during the course of its discovery. It uses descriptive metadata service, annotation service and with the help of the metaflow interpreter provides formal data provenance records for replay and analysis. Table 2 summarises the metadata services using the capability layer model.

<b>Abstract Layer</b>	<b>Functionality</b>	<b>Metadata service</b>
Preparing	Recording what happened	<i>Descriptive Metadata service.</i>
Storing	Storing and describing	<i>Descriptive Metadata service. Annotation Service</i>
Sharing	Track information flow	<i>RoR, Metaflow Interpreter</i>
Using	Fitness for purpose of resources, Discovery of available resources, Relationships, such as derivations, versioning, between resources.	<i>Metaflow Interpreter, Provenance tracking</i>

Table 2: Describing or metadata services summarised using the capability layer model.

**Protecting:** The security and privacy services assume deployment of all network services over a secure transport layer, such as SSL. They also assume that network services are securely deployed and protected from interference through interfaces other than the specified service interface. A *Member Registration* service accepts HRDN member descriptions, and optionally policy statements and purpose statements, and responds with authentication tokens designed for authentication in subsequent interaction with network resources. An *Authentication service* is primarily responsible for user authentication. An *Authorisation service* is responsible for making authorisation decisions when presented with a request from an authenticated user. The service consults the network and custodial policies when making authorisation decisions. An *Agreement Facilitator* service oversees, validates and records a user’s acceptance of a data custodian’s access agreement, including conditions and obligations. The service takes a user’s identity claim, security token and request details and consults the data custodian’s conditions of use to generate an agreement which the user must ‘sign’ before the request can proceed. A *Disclosure Risk Estimator* estimates the risk of identification or a disclosure occurring. A *Disclosure Risk Mitigator* applies statistical disclosure control techniques to reduce the risk of identification or of a disclosure in a data product before release to a researcher. The aim is to reduce the risk to an acceptable level with a minimum of information loss. An *Audit Logger* service maintains an audit trail at the request of other network services. The *Audit Log Analyser* conducts analysis and data mining of the audit logs in order to verify compliance with network conditions.

## 6 Implementation Status

Within the *Sharing* layer, we have: an implementation of a *Data Service* wrapper enabling data custodians to mount and share datasets; a basic *Orchestration* service enabling execution of a workflow; a *Planning* service enabling automatic generation of workflows in response to user queries; and a *Structure Registry* enabling static documentation of ser-

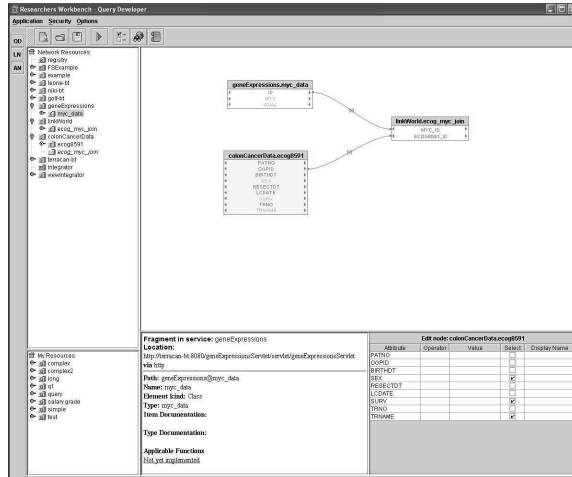


Figure 3: Screenshot of Query Developer screen in the Researcher's Workbench

vices and their interfaces. Within the *Linking* sub-layer of the *Sharing* layer, we have developed *Standardisation*, *Indexing*, *Comparison*, and *Classification* services. The *Using* layer currently has Exploratory Data Analysis tools and model builders for survival analysis. These are implemented in an *Rweb* [Ban] web server environment rather than through WSDL interfaces. These services and data resources are made available through the prototype Researcher Workbench. Figure 3 shows the workbench layout for interactive graphical query development. In the figure, network and local data resources are presented on the left-hand side, while a query under development is displayed in the graphical pane on the right-hand side. On completion of query development, a researcher can request materialization, during which the workbench converses with the *Planning* and *Orchestration* services to develop and execute a workflow. In addition, the Researcher's Workbench has an interface to the *Linking Services* and a subset of analytical methods available. The *Protecting* and *Describing* services are the next implementation focus. These services are needed before the Custodian Workbench application can be developed as the access point for data collection publishing and access management by data custodians. A commercial version of the prototype software has been developed and is being trialled for cancer studies in a government health agency and partner hospitals [HDH<sup>+</sup>05].

## 7 Related Work

The proposed Health Research Data Network requires multidisciplinary research and our work is related to other work in many areas. In the following we highlight some of the key relationships. Our service oriented architecture is aligned with generic architectures for Web Services systems such as the abstract pyramid model of [PG03]. Our overall goals are similar to those of the Clinical E-Science Framework (CLEF) [KSI<sup>+</sup>03], a research

project using Grid Services for sharing clinical data for research purposes. We share a strong concern for privacy, but are aiming for a broader network participation, leading to an emphasis on distributed data integration and custodial control of data release. Our distributed service architecture resembles that of myGRID [WSG<sup>+</sup>03]. However our services are designed more directly to meet the needs of the statistically-savvy researcher, and to implement strong privacy protection.

Turning to our design for describing services, our architecture is strongly influenced by developments in shared statistical metadata such as arising from the FASTER project and its precursors [FAS]. Distributed multi-party privacy protection protocols have been recently proposed for the linkage of clinical data at an individual level [KBH02, Chu03]. A trusted third party is required to provide these linkage services. Our architecture supports the implementation of these protocols in a form more readily available to researchers. Currently, we have developed linking services based on the algorithms implemented in the Febrl package[CC03]. Remote access data laboratories offering on-line access to "Confidentialised Unit Record Files" can also be mapped onto our service oriented architecture [ABS06]. In contrast, we are proposing to enable linking and analysis of raw datasets in a secure environment with strict authorisation procedures and confidentialisation of released data products.

## 8 Conclusions

We have proposed a service-oriented Health Research Data Network architecture. The unique contribution of HRDN is its end-to-end application of custodian specified data access and use conditions, from data custodian to researcher, as well as in-built confidentialisation and privacy protection. Description of the specific details of the component technologies used in HRDN services, such as the planner for service composition, and distributed linkage algorithms[CTB04], are outside the scope of this paper.

There are some crucial tests to be passed before HRDN could be widely deployed. One test would be to gain endorsement and adoption by several key data custodians. Another test for the HRDN is acceptance by a broad community of researchers. We are using the current implementation to test these user adoption issues.

**Acknowledgments:** We thank Dr Dave Abel and Dr Simon Hawkins for helpful discussions. We thank Mike Kearney, Deanne Vickers, Lifang Gu, Gavin Walker, Bella Robinson, Catherine Daly, and John Donnelly for their assistance with architecture ideas and implementation. The work was co-funded by CSIRO's Preventative Health National Research Flagship Program. An extended abstract was published in [TOC<sup>+</sup>04].

## References

- [ABS06] ABS. Remote Access Data Laboratory, [www.abs.gov.au](http://www.abs.gov.au), 2006. Australian Bureau of Statistics.

- [AWC<sup>+</sup>97] Leonard H Augenlicht, Scott Wadler, Georgia Corner, Christine Richards, Louise Ryan, Asha S Multani, Sen Pathak, Al Benson, Daniel Haller, , and Barbara G Heerdt. Low-Level c-myc Amplification in Human Colonic Carcinoma Cell Lines and Tumors: A Frequent, p53-independent Mutation Associated with Improved Outcome in a Randomized Multi-institutional Trial. *Cancer Research*, 57:1769–1775, 1997.
- [Ban] Paul Banfield. *Rweb*, <http://www.math.montana.edu/Rweb/>.
- [BCC03] R. Baxter, P. Christen, and T. Churches. A Comparison of fast blocking methods for record linkage. In *Proc. of ACM SIGKDD'03 Workshop on Data Cleaning, Record Linkage, and Object Consolidation*, pages 25–27, Washington, DC, USA, August 2003.
- [CC03] P. Christen and T. Churches. *Febrl: Freely extensible biomedical record linkage Manual*, release 0.2 edition, April 2003.
- [Chu03] T. Churches. A proposed architecture and method of operation for improving the protection of privacy and confidentiality in disease registers. *BMC Medical Research Methodology*, 3(1):1–13, 2003.
- [CT05] Mark Cameron and Kerry Taylor. First-Order Patterns for Information Integration. In *Web Engineering: 5th International Conference, ICWE 2005*, pages 171–184, Sydney, Australia, July 27–29 2005. Springer LNCS 3579/2005.
- [CTB04] Mark Cameron, Kerry Taylor, and Rohan Baxter. Web Service Composition and Record Linking. In *VLDB Workshop on Information Integration on the Web, IIWeb 2004*, Toronto, Canada, August 2004.
- [EVE02] M.G. Elfeky, V.S. Verykios, and A.K. Elmagarmid. TAILOR: A Record Linkage Toolbox. In *Proc. of the 18th Int. Conf. on Data Engineering*. IEEE, 2002.
- [FAS] FASTER. <http://www.faster-data.org>.
- [HDH<sup>+</sup>05] D.P. Hansen, C. Daly, K. Harrap, J. Jacquet, M.A. O'Dwyer, C. Pang, and J. Ryan-Brown. HDI: Research Software to Commercial Product. In *Australian Software Engineering Conference (ASWEC 2005)*, Brisbane, Australia, 29 March–1 April 2005. Industry Experience Report.
- [KBH02] C. W. Kelman, A. J. Bass, and C. D. J. Holman. Research use of linked health data - a best practice protocol. *Aust. N.Z. J. Public Health*, 26:251–5, 2002.
- [KSI<sup>+</sup>03] D Kalra, P Singleton, D Ingram, J Milan, J MacKay, D Detmer, and A Rector. Security and confidentiality approach for the Clinical E-Science Framework (CLEF). In *Proceedings of UK e-Science All Hands Meeting 2003*, volume 83, Nottingham, UK, September 2003.
- [PG03] M. Papazoglou and D. Georgakopoulos. Service Oriented Computing. *Communications of the ACM*, 46(10):24–28, October 2003.
- [TOC<sup>+</sup>04] Kerry L. Taylor, Christine M. O'Keefe, John Colton, Rohan Baxter, Ross Sparks, Uma Srinivasan, Mark A. Cameron, and Laurent Lefort. A Service Oriented Architecture for a Health Research Data Network. In *16th International Conference on Scientific and Statistical Database Management SSDBMS04. Proceedings*, pages 443–444, Santorini Island, Greece, 21–23 June 2004. IEEE.
- [WSG<sup>+</sup>03] Chris Wroe, Robert Stevens, Carol Goble, Angus Roberts, and Mark Greenwood. A Suite of DAML+OIL Ontologies to Describe Bioinformatics Web Services and Data. *Int. J. of Cooperative Information Systems*, 12(2):197–224, 2003.