Research Data Appliance – Key Features

The **Research Data Appliance (RDA)** is an electronic data catalogue system that allows for Dataset collection, management, documentation and local linkage. It has a range of applications such as audit, performance, benchmarking and commissioning.

Brings together diverse forms of data

The RDA aims to integrate diverse forms of data stored within multiple records systems, such as those that reside in hospitals. The RDA has also been designed to be as low a unit cost as possible to maximise the potential for RDAs to be used in a much larger number of organisations in future.

Sophisticated, deterministic and probabilistic data linkage

Integrated matching technologies are incorporated into the RDA, utilising state-of-the-art solutions such as those from Curtin University in Australia (an existing Farr Institute partner). This approach adds additional assurance to the de-identification process by allowing data owners to 'encrypt' the identifiers supplied as an input to our TTP system before supply, and for the RDA linkage module to perform high-quality matching on the hashed data, rather than on the identifiable data as currently happens.

Information governance controls

The RDA has built-in information governance controls that allow users to define roles, access controls and project permissions. Each RDA is able to interact with other RDAs (where agreement is in place) so that identity matching can occur across RDAs, and therefore across organisational boundaries and geographies, without the need to move sensitive data across. When a Dataset is transferred from one RDA to another, the Dataset documentation is also transferred, so that when the received data is loaded into the end system it is fully documented.

Dataset collection and versioning

The RDA provides the data-management infrastructure required to efficiently manage a variety of datasets. The RDA makes it easy for users of all abilities to submit Datasets through its simple user interface. It also allows users to publish different versions of a Dataset and to document the variations between each version.

Metadata capture and publishing

The RDA provides systematically gathered and shared information about the origins of each Dataset. All published Datasets can be viewed in the Local Data Catalogue. This part of the user interface provides a standard way of viewing key Dataset information. It also has links to relevant reference tables to facilitate interoperable studies.

Data quality measurement

RDA design includes basic automated quality measurement of all Datasets that are loaded by a data supplier. The system also lets users define simple rules about variable validation that will be viable as a marker of quality. The automated data quality validation checks take place during the Dataset publishing process and generates a report outlining the results of the analysis.

Stakeholders