HDI software and services to support cross-organisational clinical research

Simon McBride, Stream/Project Leader
John Ryan-Brown, Senior Software Engineer
John O’Dwyer, Implementation Consultant
The Australian e-Health Research Centre
The Australian e-Health Research Centre (AEHRC)

• A joint venture between CSIRO and the State of Queensland, AEHRC is a leading national research facility in ICT for health care innovations.

• Research into: biomedical imaging, health data, smart methods.


Image courtesy the University of Queensland
Need and Benefits of Data Linking

- Typically patient data resides across distributed, heterogeneous databases which are unconnected
  - Different information systems and reporting systems
  - Government vs. Hospital vs. GP vs. Allied health systems
- Health care improvement opportunities flow from linking this data
  - higher levels of patient care due to fuller information
  - extension of evidence-based practice
  - better planning or decision making for specific cases
- Need for integration for research
  - Public health research
  - Clinical and medical research
HDI Features

- Facilitates linking of patient data across multiple databases, at multiple locations, while maintaining data privacy and security.
- Allows data custodians to retain control of their data.
- Does not warehouse data.
- Allows researchers or clinicians to access linked but de-identified data sets for their use (as per ethics agreements).
- Allows privacy, security and ethics restrictions to be met.
Query results show CRC surgical data and chemotherapy data for each de-identified patient.

The virtual data set can be exported to your preferred analysis tool.
**HDI Logical Platform**

**De-identified linked data for analysis**

- Statistical Packages: e.g. R, SPSS
- Reporting Tools: e.g. Crystal Reports
- Custom Applications

**HDI links and integrates data**

**Custodial controlled data**

**HDI Domain**

- HDI Data Source
- HDI Data Source
- HDI Data Source
Deployment Scenario – Virtual Cancer Registry

User access to de-identified, virtual cancer registry
HDI: Process and Roles

Organisation

Domain Administrator
- Create Users, Projects, Data Services

Data Custodian
- Create Data Source. Create Data Views for Project. Publish Views to Project

Project Administrator
- Allocate Users to Projects. Create Linking Algorithms

Project Administrator
- Create Links

Project User
- Interrogate data by creating queries

Project Definition, Ethics and/or Collaboration Agreements. Agree Data Sources and Users
Software supports Delineation of Responsibility

- Domain Administrator
- Project Administrator
- Data Custodian
- Project member
Each project or data integration requirement has a specific focus, intent of use of the data, a number of data sources, and appropriate users eg to integrate cancer information across Area Health Services to populate a virtual registry.

In each case, we must scope the work, the data required, the users who will participate and get ethics agreements and / or collaboration agreements in place.
HDI is installed as required, or for existing installations, the Domain Administrator creates a Project in HDI (eg VCR) and sets up the users for the project. He also sets up the Data Services for each data source.

NB… user accounts and data services may already exist in HDI.
The owners of each data source must then become involved.

If required, the Data Custodians set up their data source in HDI. If their data source has already been used in HDI, they create additional views – specifically to address the needs of the VCR project.

When the view is ready, the Data Custodian publishes it to the project.
The Project Administrator allocates the appropriate users to the new project – ie the users that are bound by the associated ethics or collaboration agreements.

They also set up the linking algorithms for the project, adjusting them from the standard if required.
To find patient information across data sources, the Project Administrator creates links between the data views, or allocates this task to an advanced user.

They schedule the links to run overnight whenever new data has been made available.
Once all the data views are available in the project, and all other configuration is complete, extraction of data sets may begin.

Project users log into their project, and create queries as required to interrogate the data. They obtain de-identified, virtual data sets which can then be exported to other applications for further analysis.
Query results show CRC surgical data and chemotherapy data for each de-identified patient. The virtual data set can be exported to your preferred analysis tool.
Remote Domains

• Allows sharing of data across organisational boundaries

• Still preserves
  • Data custodial control
  • Encrypted linking
  • De-identification of data sets

• The communication between two domains takes place over a mutually authenticated connection that involves the exchange of security certificates.
HDI Architecture

Client machines containing HDC installation

HTTPS

HDI server listens for requests on a HTTPS port via a Java Servlet

HDI server issues SQL queries over JDBC connection against source databases

HDI server deployed as a WAR file to an instance of a JBoss application server

Source RDBMS

Source RDBMS

Source RDBMS
Create remote domain connection:
Name = HDIBrisbane
URL = https://HDIBrisbane:8443/

request connection
<X.509 certificate>
Handshaking Protocol - 2

HDIPerth Domain Administrator

HDIBrisbane Domain Administrator

accept request

accept connection
<X.509 certificate>

HDIPerth
https://HDIPerth:8443/

HDIBrisbane
https://HDIBrisbane:8443/
Handshaking Protocol - 3

HDIPerth Domain Administrator
confirm connection

HDIPerth
https://HDIPerth:8443/ 

confirm connection

HDIBrisbane
https://HDIBrisbane:8443/

HDIBrisbane Domain Administrator
Remote Query Execution

SQL parsed into:
<query join= ... />
&view-list>
  <HospitalB.Patients>
  <HospitalB.TestResults>
</view-list>
</query>

HDI splits query into local and remote queries and combines results

SQL: "select... from HospitalB.Patients join HospitalB.TestResults on ..."

This version of HDI contains a different internal representation of query objects
Linking Algorithms

• HDI provides a linking algorithm based on demographic data (First/last name, DoB, Sex, etc) “out of the box”.
• Algorithms can also be tailored for specific applications.
• Emergency Department example:
  • Three data sources (ED, Ambulance, hospital admissions)
  • Match patient if:
    • Same First and Surname, sex, age within 5 yrs
    • Same date of triage treatment and treatment time within
      • 1 hour between QAS and EDIS
      • 12 hours between EDIS and HBSCIS
  • Compared with a “Gold Standard” manual matching
    • Sensitivity: 99.28% and 93.72%
    • Specificity: 99.99% and 100%
    • Both results acceptable in this case.
Summary

- **Need/benefits of data linking are well known**
  - Higher levels of patient care due to fuller information
  - Extension of evidence-based practice
  - Better planning or decision making for specific cases
  - Research (clinical, public health, etc)

- **HDI can support cross-organisational research by providing data linking and integration capabilities**
  - Facilitates linking of patient data across multiple databases, at multiple locations, while maintaining their privacy and security.
  - Allows data custodians to retain control of their data.
  - Does not warehouse data.
  - Allows researchers or clinicians to access linked but de-identified data sets for their use (as per ethics agreements).
  - Allows privacy, security and ethics restrictions to be met.
Thank you

Simon McBride
Team/Project Leader
The Australian e-Health Research Centre
L7, UQ CCR Building
Royal Brisbane and Women’s Hospital
Herston Queensland Australia 4029
v: +61 7 3253 3631
f: +61 7 3253 3690
w: http://aehrc.com/
e: simon.mcbride@csiro.au