

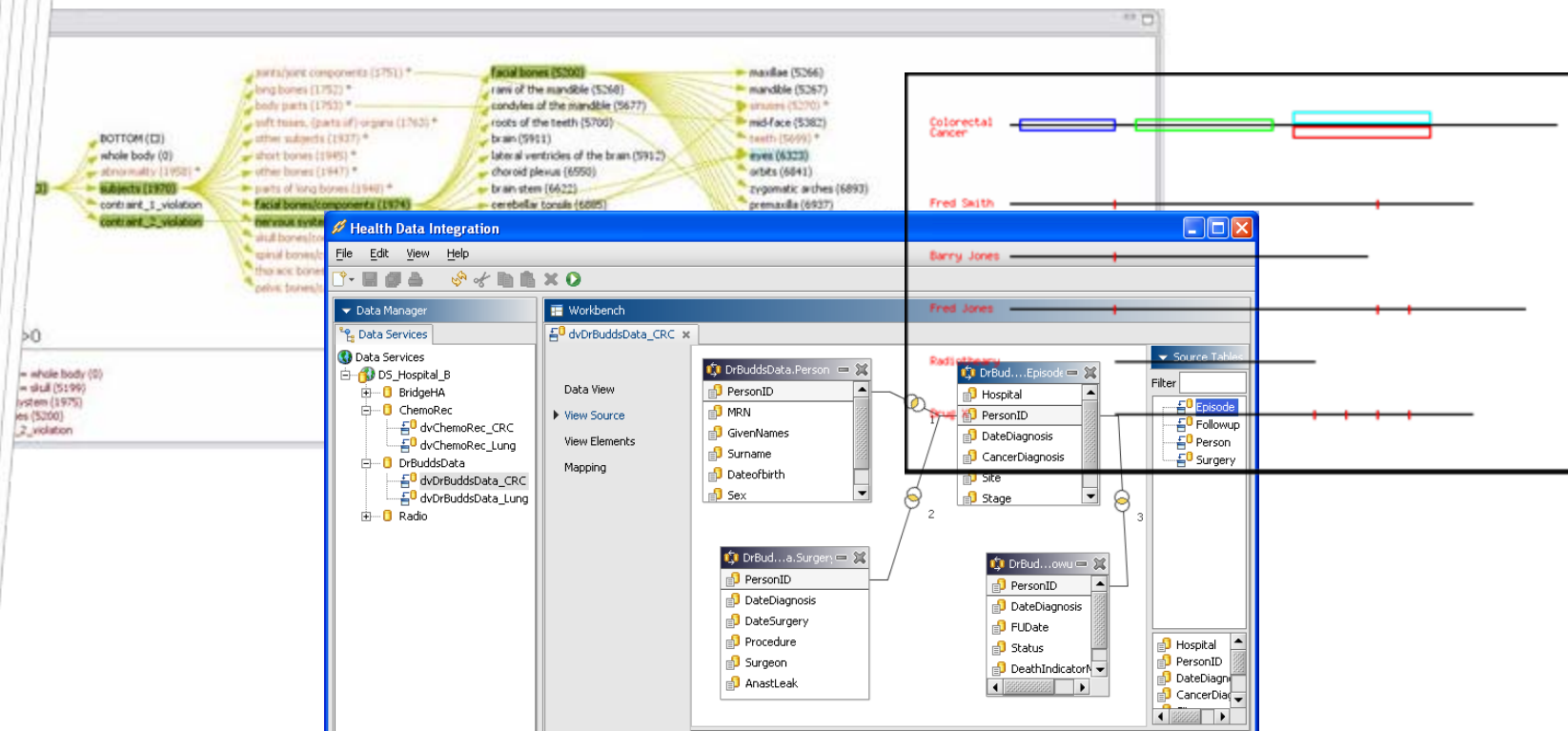
HDI software and services to support cross-organisational clinical research

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.
- More at <http://aehrc.com/> (publications, projects, demos).



Image courtesy the [University of Queensland](http://www.uq.edu.au)

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.

HDI: Building a virtual data set

Q_Budds_and_Chemo							
DATABASEPERS...	DATEOFDIAGN...	PRINCIPALDIAG...	SITE	STAGE	TREATMENTSTA...	SYSTEMICTHER...	ROUTE1
57493.0	1999-09-13 00:...	Colon	Caecum	D	1999-10-25 00:...	5-fluorouracil	Intravenous
57620.0	1995-10-27 00:...	Colon	Ascending Colon	D	1995-12-08 00:...	5-fluorouracil	Intravenous
57780.0	2004-12-01 00:...	Rectum	Rectum-Upper T...	C	2005-01-12 00:...	5-fluorouracil	Intravenous
57841.0	2003-08-06 00:...	Colon	Sigmoid Colon	D	2003-09-17 00:...	5-fluorouracil	Intravenous
57881.0	1998-04-01 00:...	Colon	Hepatic Flexure	C	1998-05-13 00:...	5-fluorouracil	Intravenous
58672.0	1995-12-09 00:...	Colon	Caecum	D	1996-01-20 00:...	Raltitrexed	Intravenous
58680.0	1997-10-15 00:...	Colon	Rectosigmoid	C	1997-11-26 00:...	5-fluorouracil	Intravenous
58751.0	1991-11-07 00:...	Colon	Sigmoid Colon	C	1991-12-19 00:...	5-fluorouracil	Intravenous
58762.0	1998-06-23 00:...	Colon	Hepatic Flexure	C	1998-08-04 00:...	5-fluorouracil	Intravenous
58912.0	2002-11-01 00:...	Colon	Caecum	C	2003-01-24 00:...	5-fluorouracil	Intravenous
59012.0	1997-02-13 00:...	Colon	Transverse Colon	D	1997-03-27 00:...	Raltitrexed	Intravenous
59542.0	2001-04-05 00:...	Colon	Sigmoid Colon	C	2001-05-17 00:...	5-fluorouracil	Intravenous
60002.0	1998-09-23 00:...	Rectum	Rectum-Lower T...	D	1998-11-04 00:...	Raltitrexed	Intravenous
60063.0	1991-09-30 00:...	Colon	Rectosigmoid	C	1991-11-11 00:...	5-fluorouracil	Intravenous

Maximum rows: 1000 Re-execute query

CRC Surgical Data

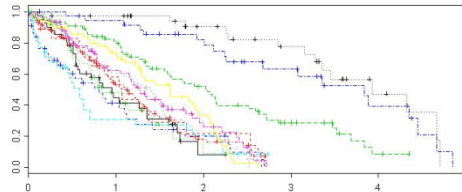
Chemotherapy Record Data

Query results show CRC surgical data and chemotherapy data for each de-identified patient.

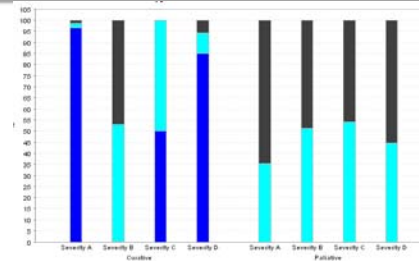
The virtual data set can be exported to your preferred analysis tool.

HDI Logical Platform

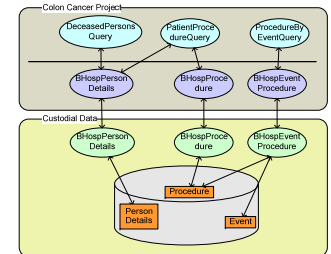
**De-identified
linked data for
analysis**



Statistical Packages
e.g. R, SPSS

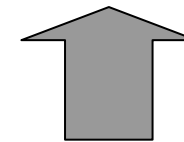


Reporting Tools
e.g. Crystal Reports



Custom Applications

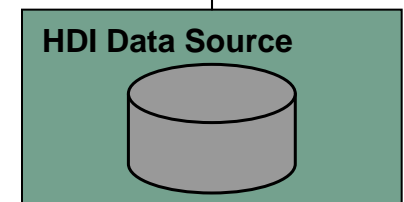
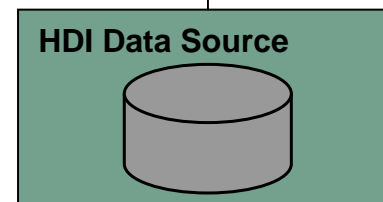
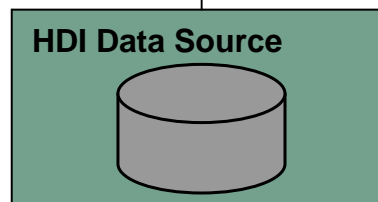
**De-identified virtual
linked data set**



**HDI links and
integrates data**

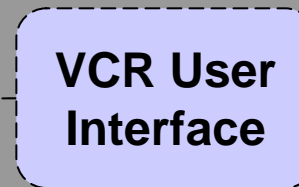


**Custodial
controlled data**

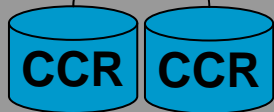
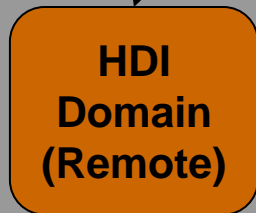


Deployment Scenario – Virtual Cancer Registry

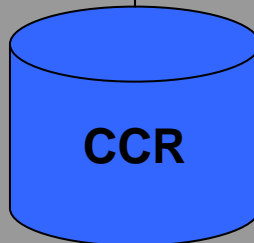
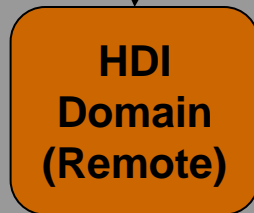
Cancer Institute



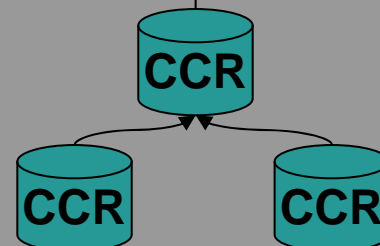
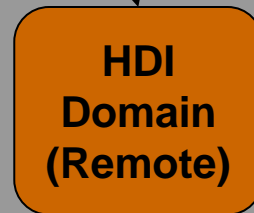
User access to de-identified, virtual cancer registry



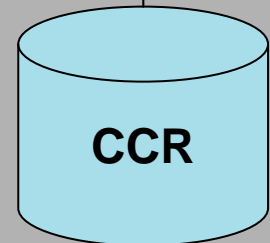
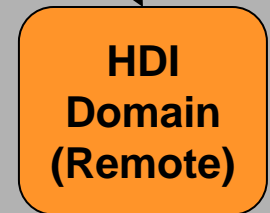
Health Service A



Health Service B



Health Service C



Health Service D

HDI: Process and Roles

Organisation

Project Definition, Ethics and / or Collaboration Agreements. Agree Data Sources and Users

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

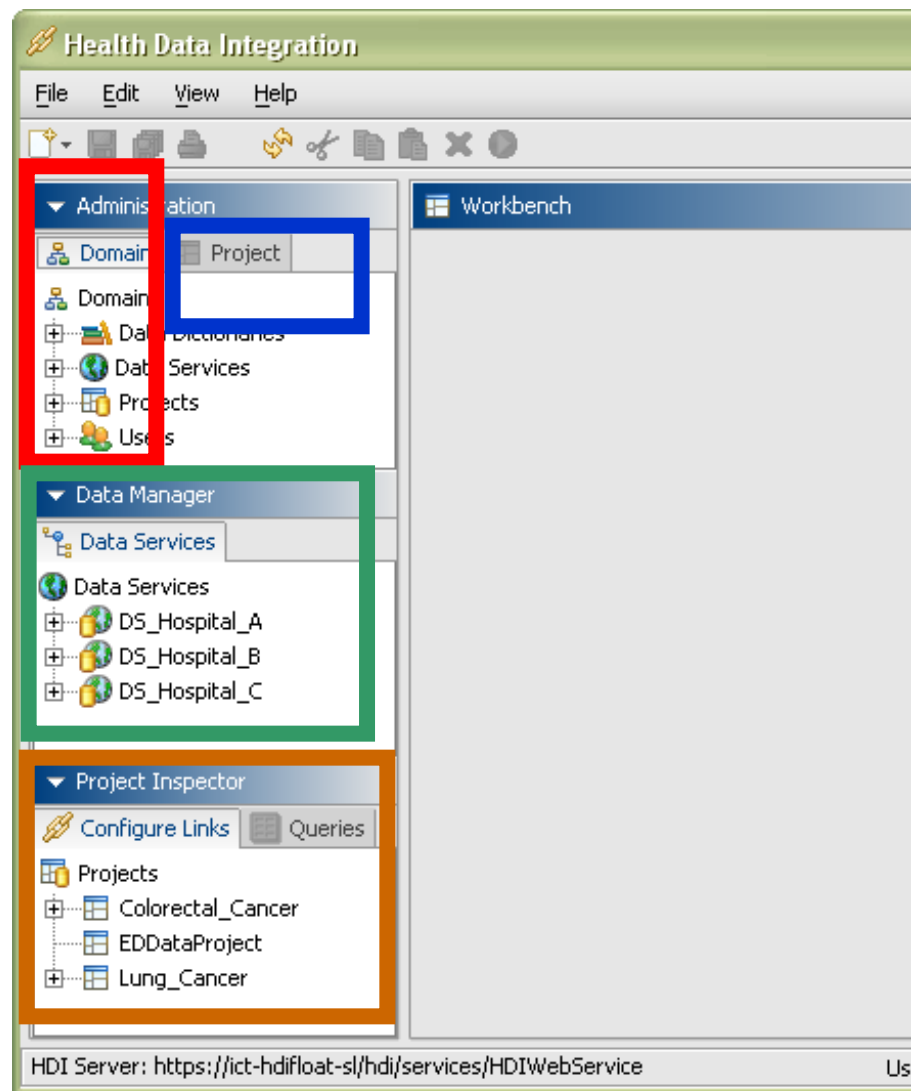
Software supports Delineation of Responsibility

Domain Administrator

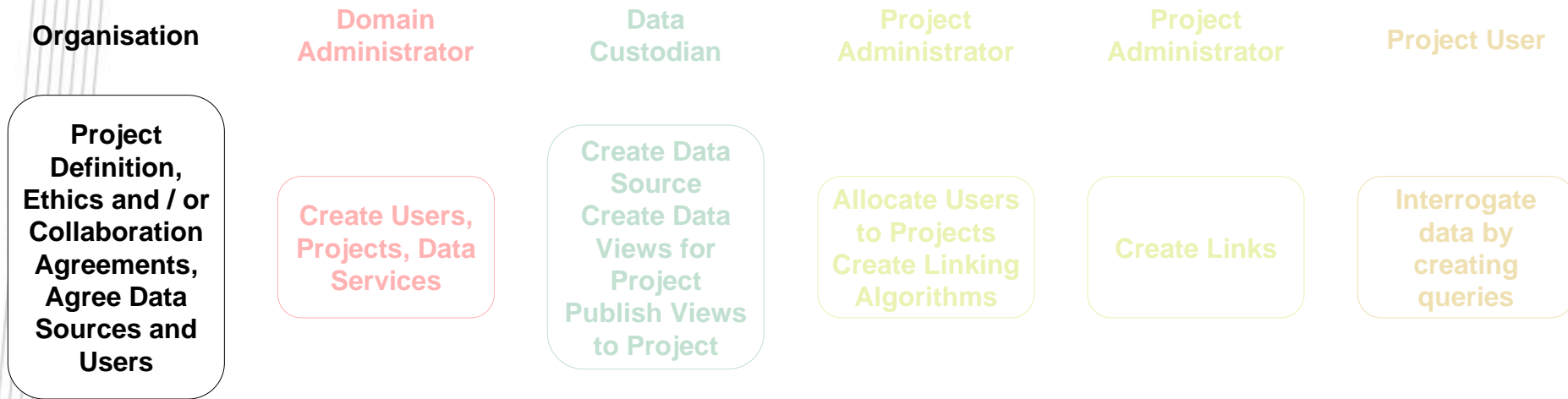
Project Administrator

Data Custodian

Project member



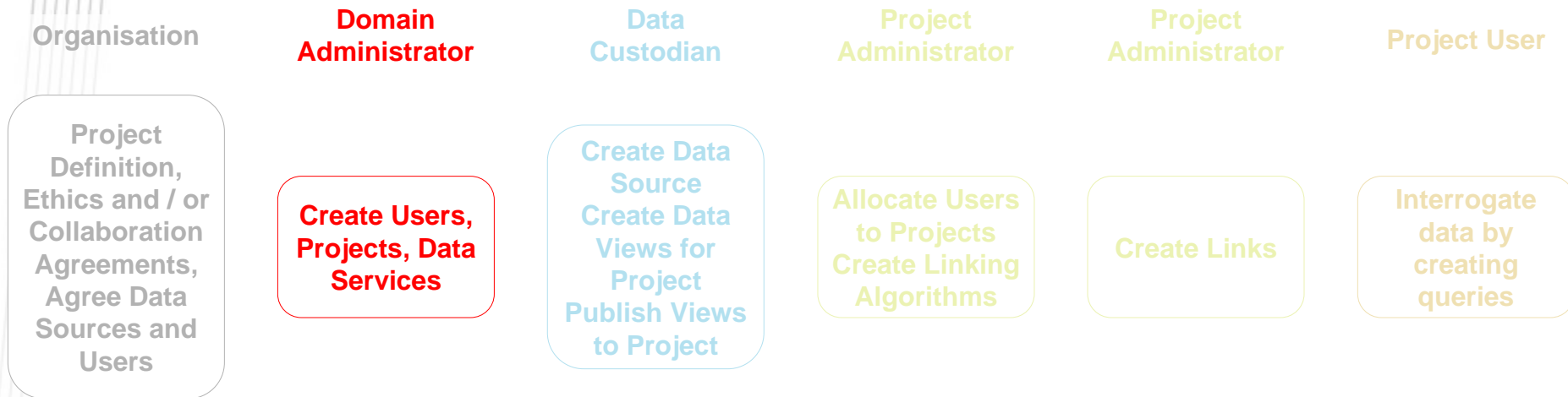
Process – Organisation



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.

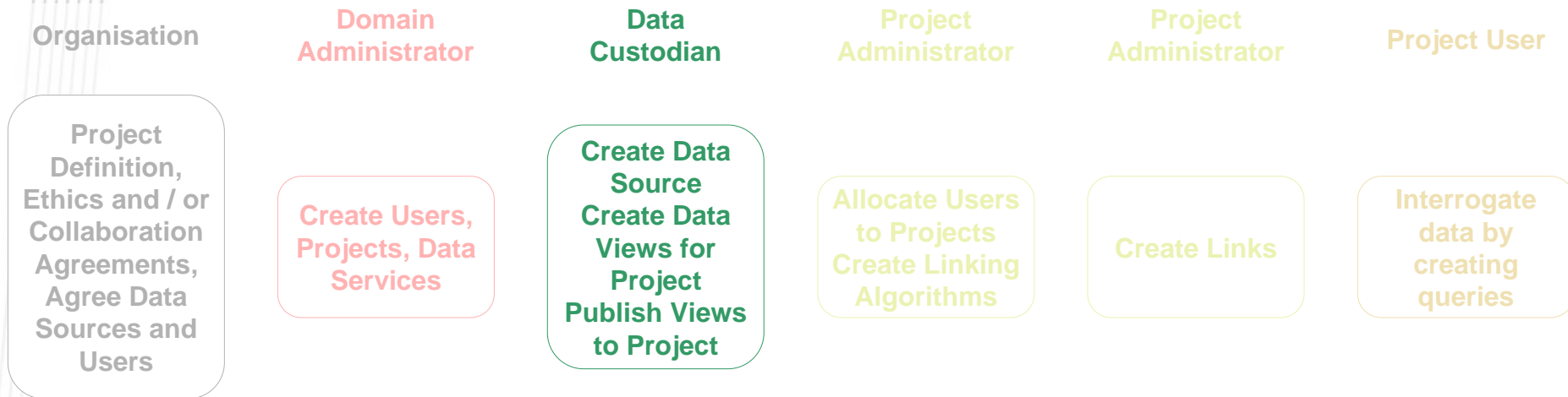
Process – Domain Administrator



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.

Process – Data Custodian

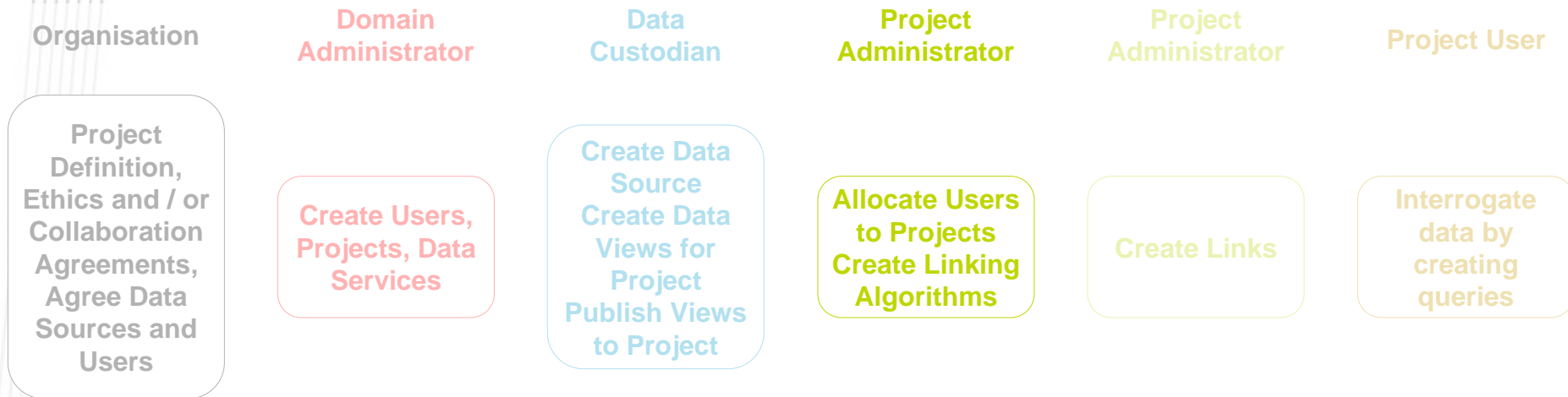


The owners of each data source must then become involved.

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.

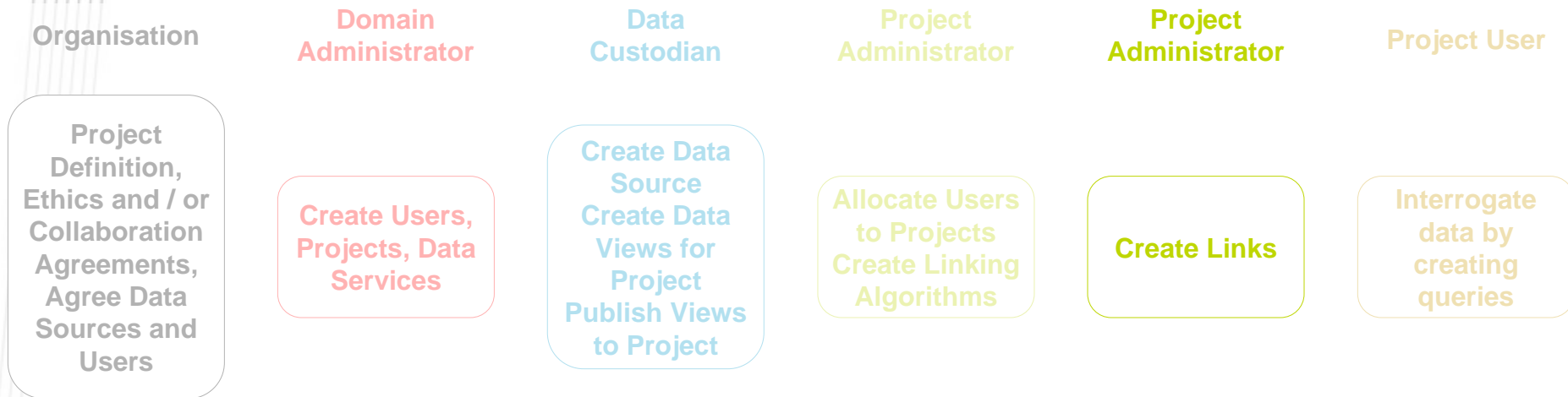
Process – Project Administrator 1



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.

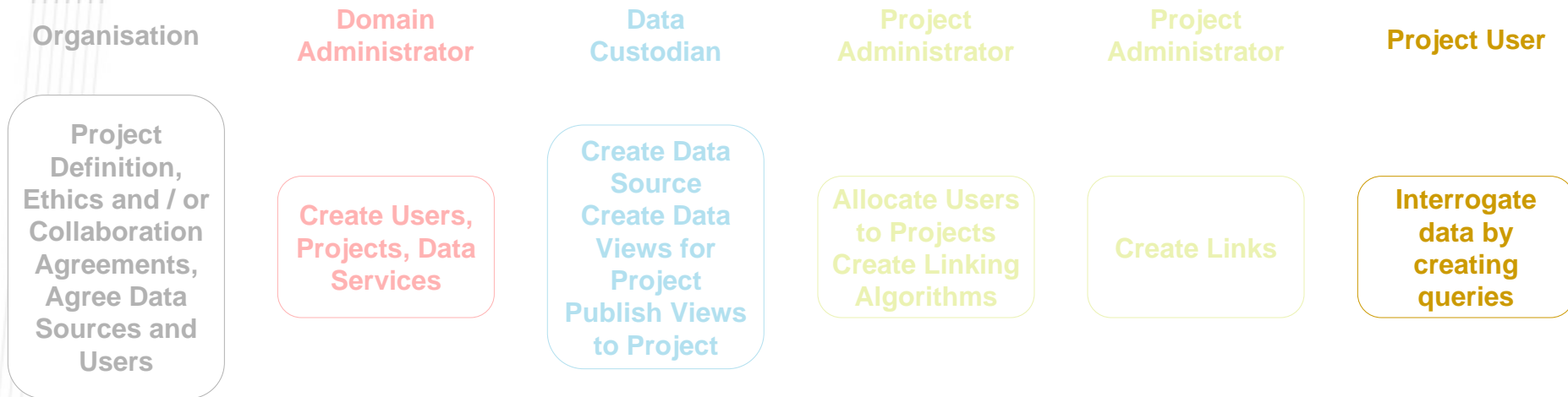
Process – Project Administrator 2



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.

Process – Project User



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.

HDI: Building a virtual data set

DATABASEPERS...	DATEOFDIAGN...	PRINCIPALDIAG...	SITE	STAGE	TREATMENTSTA...	SYSTEMICTHER...	ROUTE1
57493.0	1999-09-13 00:...	Colon	Caecum	D	1999-10-25 00:...	5-fluorouracil	Intravenous
57620.0	1995-10-27 00:...	Colon	Ascending Colon	D	1995-12-08 00:...	5-fluorouracil	Intravenous
57780.0	2004-12-01 00:...	Rectum	Rectum-Upper T...	C	2005-01-12 00:...	5-fluorouracil	Intravenous
57841.0	2003-08-06 00:...	Colon	Sigmoid Colon	D	2003-09-17 00:...	5-fluorouracil	Intravenous
57881.0	1998-04-01 00:...	Colon	Hepatic Flexure	C	1998-05-13 00:...	5-fluorouracil	Intravenous
58672.0	1995-12-09 00:...	Colon	Caecum	D	1996-01-20 00:...	Raltitrexed	Intravenous
58680.0	1997-10-15 00:...	Colon	Rectosigmoid	C	1997-11-26 00:...	5-fluorouracil	Intravenous
58751.0	1991-11-07 00:...	Colon	Sigmoid Colon	C	1991-12-19 00:...	5-fluorouracil	Intravenous
58762.0	1998-06-23 00:...	Colon	Hepatic Flexure	C	1998-08-04 00:...	5-fluorouracil	Intravenous
58912.0	2002-11-01 00:...	Colon	Caecum	C	2003-01-24 00:...	5-fluorouracil	Intravenous
59012.0	1997-02-13 00:...	Colon	Transverse Colon	D	1997-03-27 00:...	Raltitrexed	Intravenous
59542.0	2001-04-05 00:...	Colon	Sigmoid Colon	C	2001-05-17 00:...	5-fluorouracil	Intravenous
60002.0	1998-09-23 00:...	Rectum	Rectum-Lower T...	D	1998-11-04 00:...	Raltitrexed	Intravenous
60063.0	1991-09-30 00:...	Colon	Rectosigmoid	C	1991-11-11 00:...	5-fluorouracil	Intravenous

Maximum rows: 1000 Re-execute query

CRC Surgical
Data

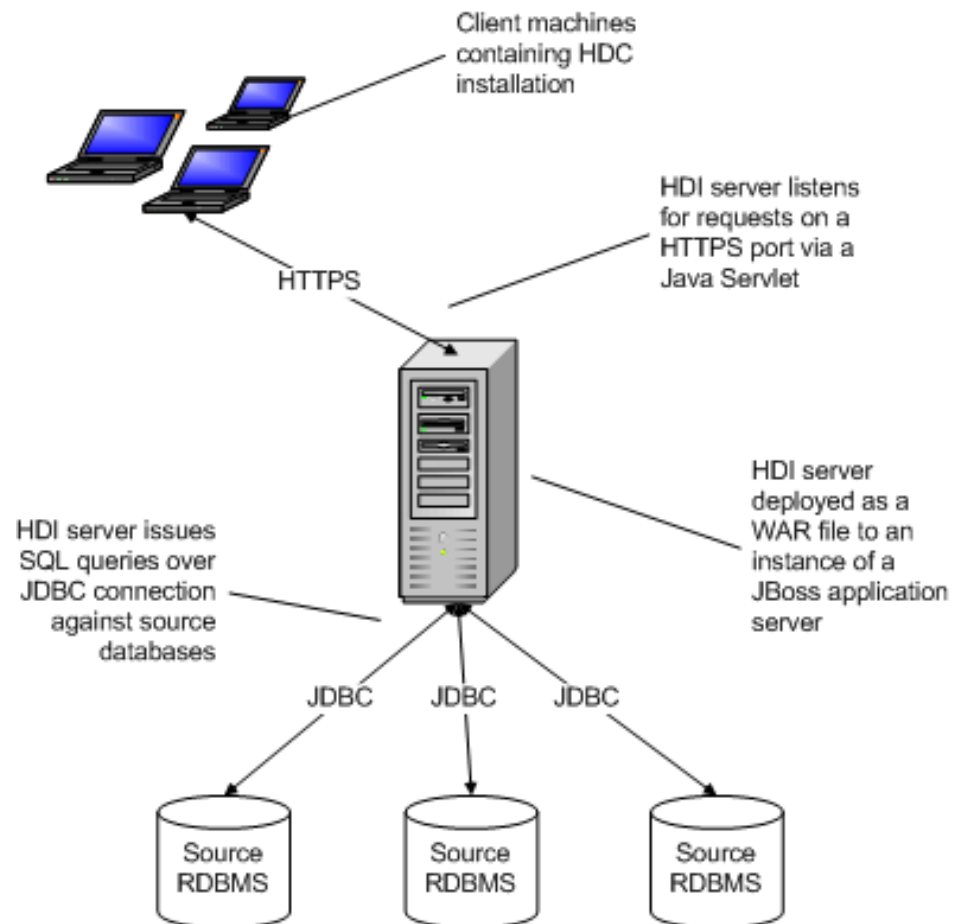
Chemotherapy
Record Data

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.

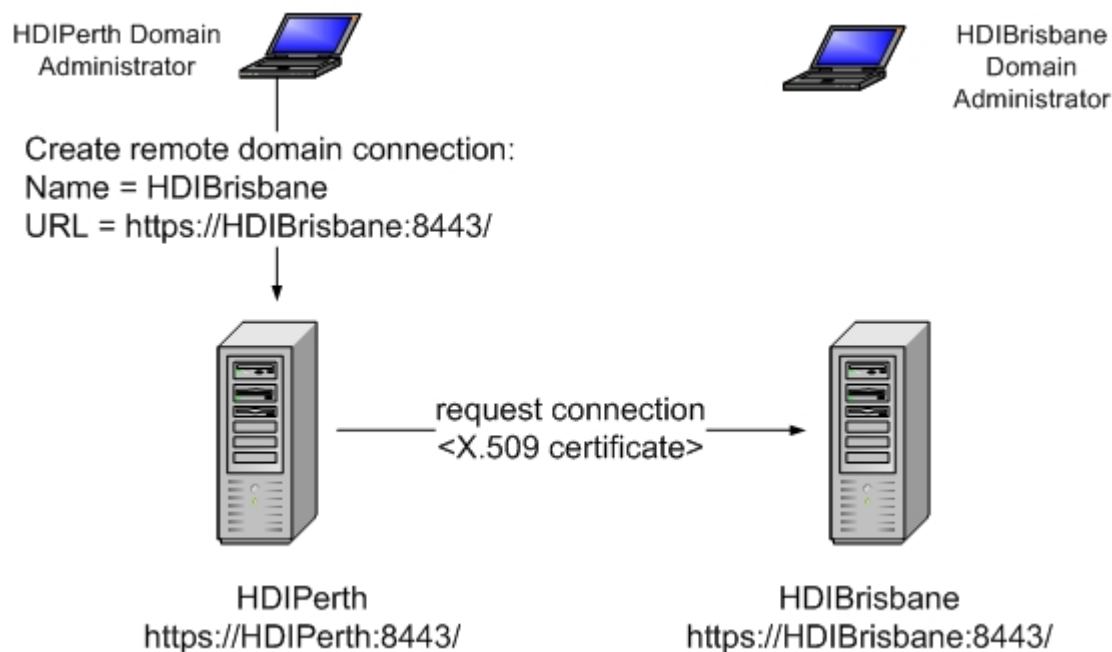
Remote Domains

- **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.**

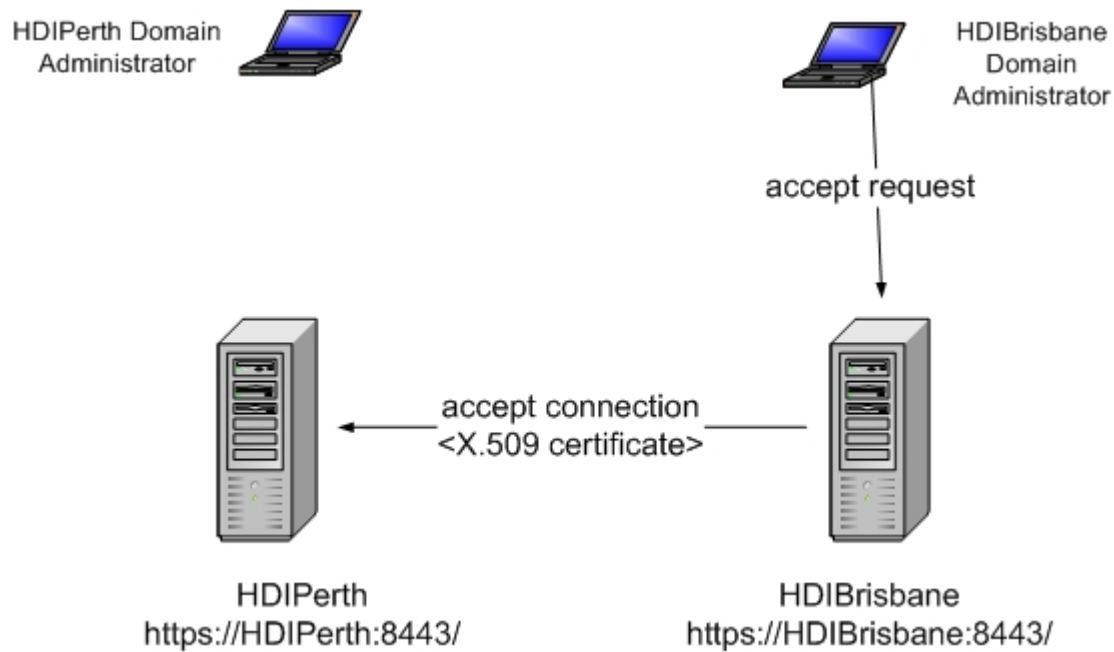
HDI Architecture



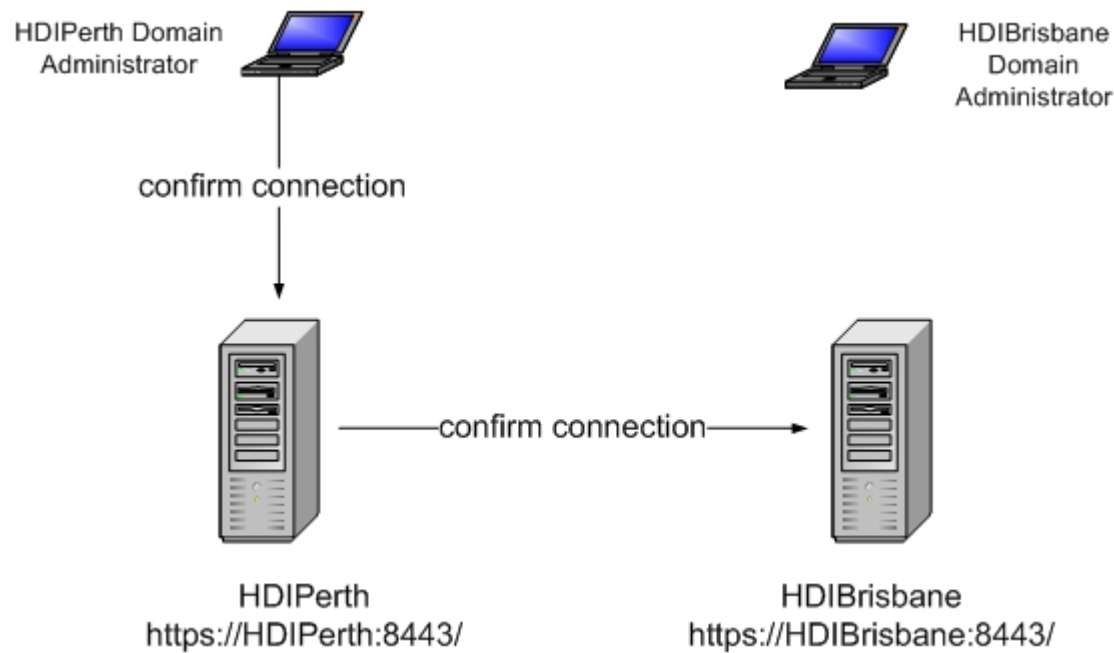
Handshaking Protocol - 1



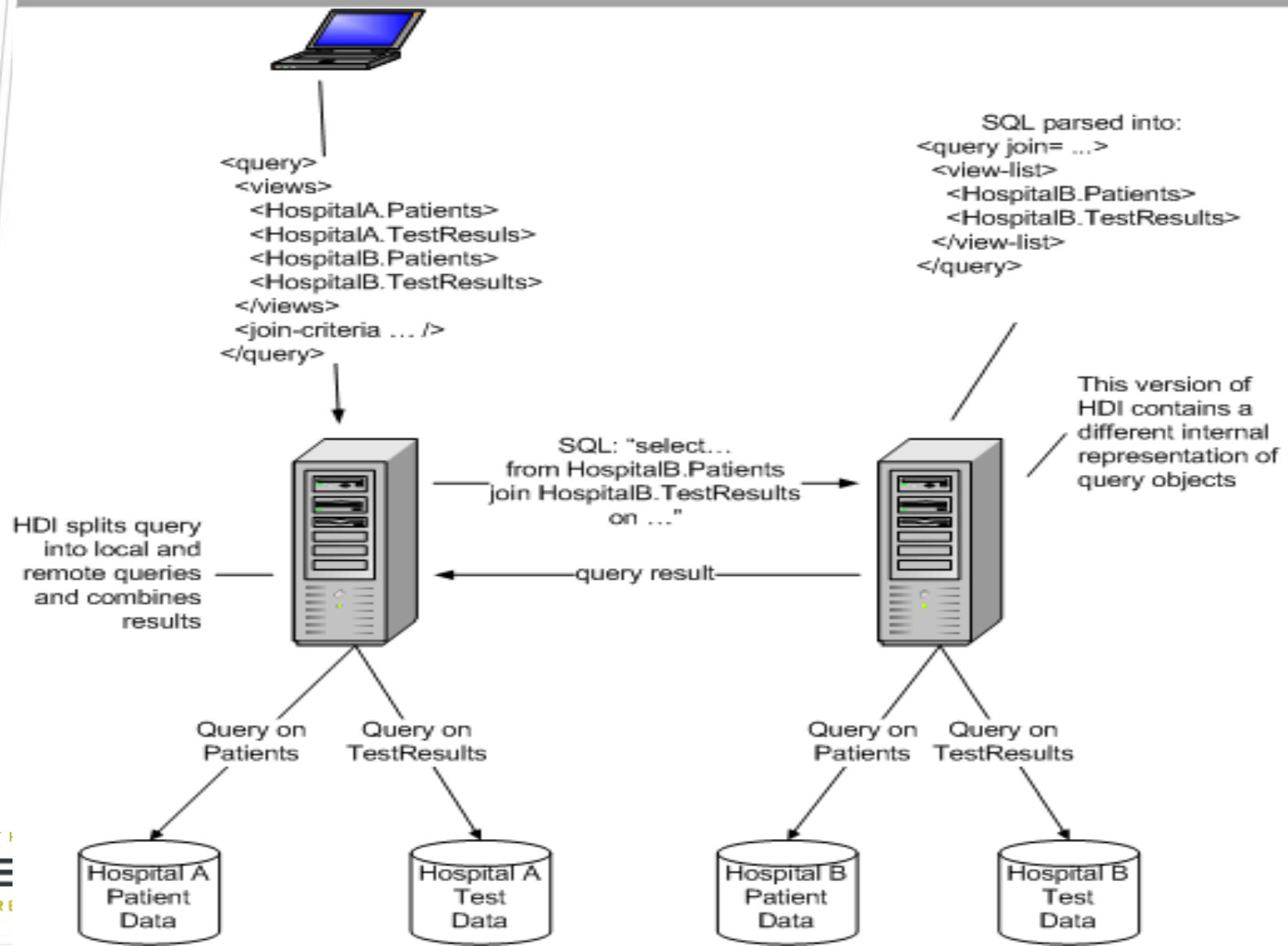
Handshaking Protocol - 2



Handshaking Protocol - 3



Remote Query Execution



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