Enabling Life Science Research

: BioGrid Australia

Marienne Hibbert, Project Director, BioGrid
Marienne.hibbert@biogrid.org.au
A generic informatics model providing opportunities for beneficial collaboration across organisations and expansion to other research areas.
IF YOU WANTED TO COMPARE

LUNG CANCER PATIENTS,

where would you start?

Data
- Missing
- Incomplete
- Indecipherable
- Somewhere else
- Privacy protected
- Difficult to correct
- “Owned”
BioGrid Project ‘parallel processes’

- **Privacy, Ethics and authorisation**
  - BioGrid infrastructure + processes must have Ethics approval

- **Membership, IP and support**
  - BioGrid membership agreements signed
  - Subscriptions

- **Computer Systems**
  - Install computers
  - Establish secure connectivity

- **Research Collaboration**
BioGrid Australia

- Victorian and Australian Government investment
- First cross-institution, cross-discipline clinical research data integration platform (IBM)
- Nightly upload of data on-site
- No identified health data leaves the site
- No health data stored centrally
- Individual records linked via linkage key
Australia

- Australia’s population - 21 million

- Cancer is the leading cause of death in Australia with 30 per cent of all deaths attributable to cancer. The top four tumours based on mortality are lung, bowel, prostate and breast cancer.

- Australia's Seven National Health Priority Areas are arthritis and musculoskeletal conditions, asthma, cancer control, cardiovascular health, diabetes mellitus, injury prevention and control, and mental health. In 2006, deaths associated with the seven National Health Priority Areas accounted for 77.4% of all underlying causes of death and were either associated with or the underlying cause of 90.8% of deaths.
Why link databases?

- **Research power:**
  - Increase the sample size
  - Increase the potential for research collaborations
- **Link specialist databases covering common diseases:**
  - Screening activity - Genetic predisposition
  - Environmental exposures
  - Genomics, proteomics & epigenetics
  - Co morbidities
  - Quality and audit
  - Treatment strategies
  - Outcomes
Medical Research - the Challenges

- Large amount of data
- Lack of data standards
- Lack of interoperability between databases
- Need a cohesive approach between disciplines
- Ethics, Privacy and Regulation
- Who owns the Intellectual Property
Pilot – then 1 - 2 - 3

Phase 1 (pilot): funded by Victorian Government (A$1.6 M) 2003 - 2005
- 5 hospitals
- 2 Research Institutes
- 3 disease types: Oncology x 1, Diabetes, Epilepsy

Phase 2: funded by Australian Government (A$4.5 M) 2005 - 2007
- 6 hospitals
- 2 Research institutes
- 4 diseases: Oncology x 3, Neuroscience, Diabetes, Respiratory + Images

Phase 3: ACG funded by Victorian Government (A$11.0 M) 2006 - 2009
- 7 hospitals
- 2 Research institutes
- 4 diseases: Oncology x 9, Neuroscience, Diabetes, Respiratory + Images
BioGrid Pilot: Phase 1

- Requirements went to public tender
- IBM design and technology adopted
- DB2 (v8) Enterprise Edition
- DB2 (v8) Information Integrator
- Query Patroller
- Successful demonstration of federated model and technology completed February 2005.
DEST Grant: Phase 2

- But how would researchers find out about BioGrid and its contents?
  - Requirements went to public tender
  - IBM design and technology adopted
  - IBM Rational Data Architect
  - IBM WebSphere DataStage and Metadata Server
  - IBM Business Glossary

- Website launched: now biogrid.org.au
DIIRD Grant: Phase 3

- (Department of Innovation, Industry and Regional Development, State of Victoria)
- Increase in local sites across Australia
- Interest expressed from international institutions
- Expansion of cancer and other disease data
- Increase in development of data management applications and services
- Enhancement and development of architecture and infrastructure

- **New name:** BioGrid Australia
Institute-specific data loaded into institute-specific Local Research Repository nightly.

Public Data Sources
- GenBank
- UniProt
- LocusLink
- PubMed

Authorised researchers query the Federated Data Repository for analysis.
Scope - Collaborating Sites

Victorian Sites
Melbourne
Austin
Western
Peter Mac
Alfred
St Vincent’s
Monash
RCH
Box Hill
Peninsula
Northern
Barwon
Ballarat
Latrobe
Bendigo
Hume

BioGrid

Other Australian states
NSW: St Vincent’s
POW
Tas: RHH/Menzies
Act: Canberra
SA: Flinders
Royal Adelaide
Queen Elizabeth
Lyle McEwan

Sites – International
USA: Moffitt
Vanderbilt
Venter
NZ: Christchurch
UK: St Marks
Brazil: Sao Paulo
Malaysia

De-identified data
Researcher

© Copyright IBM Corporation 2007. All Rights Reserved.
Scope - Clinical Datasets

Cancer:
- Colorectal
- Brain
- Breast
- Lung
- Sarcoma
- ...
- Prostate
- Head & neck
- Upper GI
- Melanoma
- Renal

Neuroscience:
- Epilepsy
- MS

Diabetes
- Type 1
- Type 2

Other
- Cystic Fibrosis
- Crohns
- Well Womens

De-identified data

Researcher
Scope - Analysis

Data:
- Clinical outcome
- Treatment
- Genetic (Microarray, Biomarker, Proteomic)
- Images
- Tissue Banks
- Geographic
- Demographic

Tools
- Bioinformatics
- Statistical
- Drug Discovery
- Image Analysis

BioGrid

De-identified data

Medicare tbc
DHS
ABS tbc
Deaths tbc

High Performance Computing

Researcher
Participating Institution Options

Institution based source data

ETL Tools

Institution based Local Research Repository

Data Source

- MS .net
- ASP
- Access
- SQL Server
- IBM DB2
- Access
- Oracle
- Files:
  - MS Excel
  - Other flat files

- IBM Ascential Datastage
- IBM DB2 Warehouse Center
  - Microsoft SQL Server SSIS

- IBM DB2 UDB
- Microsoft SQL Server 2005
- Oracle 9i +

BioGrid Architecture

IBM Information On Demand 2008

Current BioGrid Architecture

- Participating Institution Options
- Federated Data Integration Server
  - IBM DB2
  - IBM Websphere Information Integrator

- USI Server
  - Oracle 9i
  - USI Generation Propagation

- USIDB
  - Query

- FDIDB
  - Query

- Data Discovery
  - Metadata Discovery Server
    - IBM DB2
    - IBM Websphere Business Glossary

- Data Analytics Server
  - IBM Websphere Information Integrator
  - SAS Enterprise Business Intelligence Server
  - SAS Web Report Studio

- Researcher desktop
  - Microsoft Internet Explorer v6 sp1

- SAS Enterprise Guide
  - SAS Enterprise Business Intelligence Server

- IBM Websphere Business Glossary

BioGrid Australia - Health through Information

Our new image and branding was launched by the Hon Gavin Jennings on March 5, 2008. The new look articulates what we do - linking data thereby creating a ‘grid’ of diverse data for research, planning and other services in health.

We dynamically link data which includes dates of diagnosis, clinical symptoms, pathology, radiology reports, and availability of a biospecimen sample, genomic data, MRI images, treatments: drug therapy, radiotherapy and surgical status in a privacy protected way for authorised users.

We have made steady progress with ethics approvals, collaborative agreements, IT implementation and data linkage. We are proud of the collaborator research outcomes that have resulted from this, and some of the current work includes:

- Data mining – looking at patterns and exceptions
- Data analysis – associations between the genetic and clinical data
- Evaluation of clinical services
- Quality and audit reports of Clinical treatment
### Braintumour Table

This table stores the details of the tumour classification type.

#### Contained Terms:
- Date of diagnosis
- Unique identifier
- Glioblastoma Multiforme
- Anaplastic Astrocytoma
- Anaplastic Oligodendroglioma
- Anaplastic Astrocytoma
- Anaplastic Oligodendroglioma
- Low Grade Glioma
- Medulloblastoma
- Ependymoma
- Germinoma
- Meningioma
- Primary Cerebral Lymphoma
- Brain Other
- Brain Other Type Specified

#### Referenced Terms:
- Glioblastoma Multiforme
- Anaplastic Astrocytoma
- Anaplastic Oligodendroglioma
- Low Grade Glioma
- Medulloblastoma
- Ependymoma
- Germinoma
- Meningioma
- Primary Cerebral Lymphoma
- Brain Other
- Brain Other Type Specified

#### Expanded View:

- Diabetes at StVs Database
- Biomarkers Database
- Gastric Cx Microarray Database
- Cystic Fibrosis at RCH Database
- FAMBIS Database
- Tissue Bank Database
- Accord Database
  - Episode Table
  - CC-Adjuvant Table
  - Breasttumour Table
    - Braintumour Table
  - Epilepsy Clinic Database
Welcome to the MMIM Medical Research Terms Glossary

This glossary contains a list preferred terms for all the data elements stored in the MMIM arranged in a hierarchy of categories that represent the database structure (Database, tables, columns). Each term has a text description and is linked to data to inform you exactly where to find the term in the MMIM databases. i.e; which databases/tables/columns the term is present in. For example, “date of birth” is present in tables throughout all databases but the term is present only once in the glossary with multiple references to it’s location in all databases. There are two way to find a term – by searching or by drilling down the hierarchy.
Project challenges

- **Linking the data**
  - across disparate, clinical and biomedical data sources
  - within and across institutions
  - compliance with security and data ownership constraints
  - each institution to keep their own data models and data control

- **Linking patient/subject records**
  - assigning Unique Subject Identifiers (USI’s) to data
  - patients linked across multiple institutions
  - observing legal, ethical, privacy and data ownership constraints

- **Providing a uniform interface**
  - a consistent model of data, independent of institution

- **Collaboration**
Project challenges

- Governance
  - Infrastructure (Strategic, Maintenance, growth, cost)
  - Use and access
  - Collaborative agreements
  - Commercial use

- Data issues - preservation of privacy
  - Consent and Ethical issues
  - Intellectual property and data ownership
  - Standards and quality
  - Data Types
  - Version control/data stamping/archiving of data extracts
Project challenges

- **Technology**
  - Interoperability - Ensure all the systems can work/link together
  - Meta data - descriptions

- **Tools**
  - for searching what data is there - really user friendly
  - for querying the data itself - really user friendly
  - for analysis
  - to assist in data cleaning and profiling
  - for tracking and audit
  - for record matching
Outcomes – early days

- Expanded collaboration, research capacity and productivity
  - Within disease groups,
  - Across disease groups
  - between clinical researchers, bioinformaticians and IT specialists.
  - between data owners and researchers across academic institutions in Australia
  - between data owners and researchers overseas (US underway and Europe discussions)
International Comparisons

Survival by Adjuvant Chemotherapy

Outcomes achieved in Australian patients, with and without chemotherapy, are as good as anywhere in the world

(P. Gibbs, S. McLaughlin, I. Skinner et al.)
Commercial Opportunities

- Genetic Biomarkers of Drug Response and Non response to Carbamazepine and Valproate Treatment for Epilepsy
Cancer - findings

- Cancer treatment and survival in non-english speaking patients was no different to patients of english speaking background – despite communication and socio-economic differences. (J. Rodrigues, E.Lim, S. McLaughlin et al ANZJS 2006)

- Post-surgical cancer patients who have additional treatment after the initial chemotherapy have much better survival.
  - BioGrid found both patient and physician influence completing further treatment (P.Gibbs, S. McLaughlin, I. Skinner et al JNCI 2006)

- Cancer surgery – increasing the number of lymph nodes improves the pathology diagnosis
  - BioGrid showed an intervention improves the yield of lymph nodes (N. Reiger, FS. Barnett, J. Moore et al. AGITG 2006)
NEXT STEPS

- Growth of the Technology
- Growth in Content / Data types
- Growth in Disease Diversity
- Growth in Participating sites
- Successful Translational Research
- Business Opportunities