



THE UNIVERSITY OF  
WESTERN AUSTRALIA  
*Achieving International Excellence*

# Australian Paediatric Tumour Banking Network

## Databases and Applications

Paul White



Centre for  
Genetic Epidemiology  
& Biostatistics  
THE UNIVERSITY OF WESTERN AUSTRALIA



- 
- I manage an open source software development project, The Ark (formerly WAGER), at the University of Western Australia to provide informatics applications to Australian and international researchers.
  - Background in software development and management consulting
  - Principal Investigator on the OBiBa project



## A few key elements are likely to be required to establish the Paediatric Tumour Banking Network

---

- A database containing de-identified summary tumour data that can be queried over the web, along with the appropriate consent
- Mechanisms for populating the database from data owned by each biobank
- Tools for defining data and protocol standards and disseminating them through the network
- Tools to ensure adherence of summary data provided by each node to quality standards



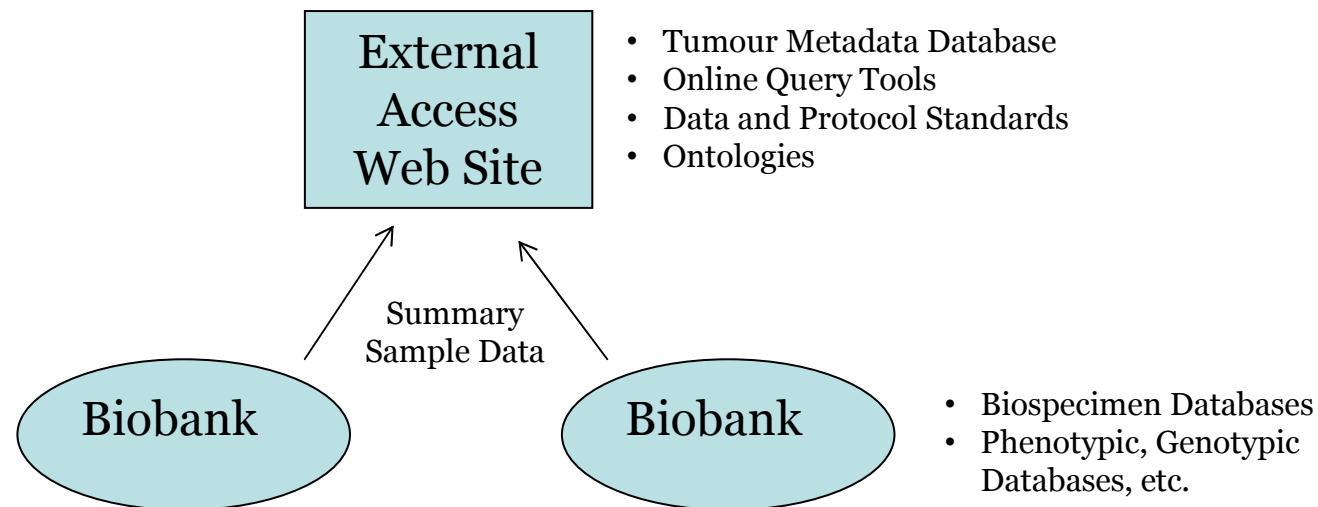
## Beyond this there are other software tools and databases that may be useful/necessary

---

- Researcher request review tools
- Software to support the provision and tracking of samples sent to researchers
- Software to support invoicing and billing
- Databases and software to support the collection and management of biospecimens and other data types, e.g. images, phenotypic and genotypic data
- Software to support the extraction and aggregation of data associated with samples
- Tools to ensure the overall quality control of data and samples provided
- Tools to support the overall management of research studies, including processes such as recruitment, consent management and questionnaire data collection
- Tools and services to support data linkage
- Hospital system integration/data extraction software



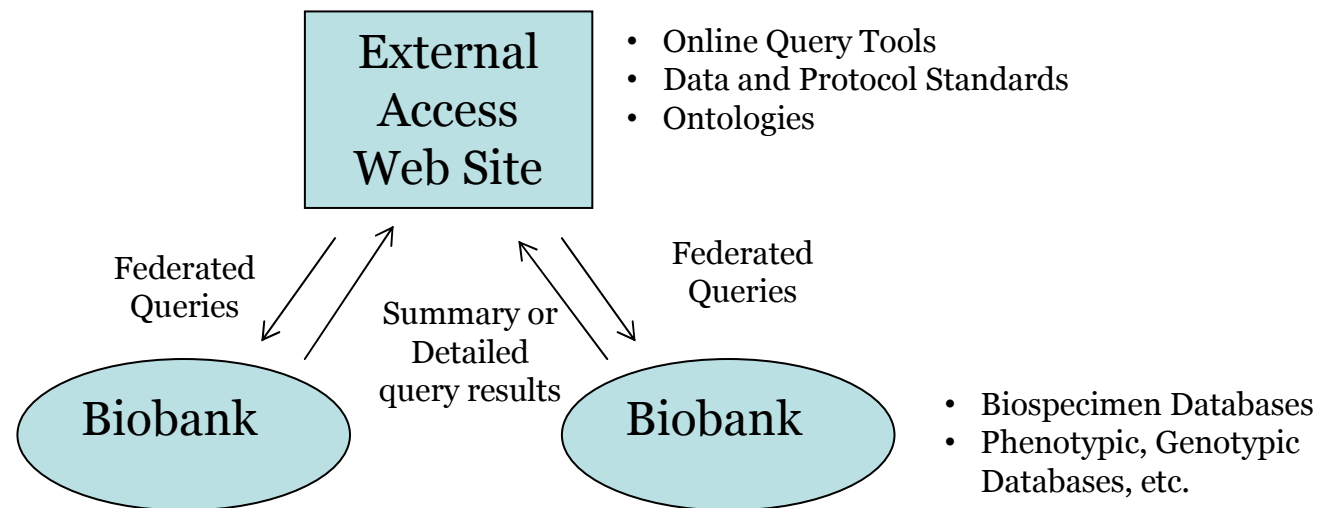
# Database deployment models – Distributed/Decentralised Databases



- Relatively easy and cost-effective to implement
- Model adopted by the Australasian Biospecimen Network
- Data is not typically updated in real-time
- Data security and provision remains the responsibility of the biobank



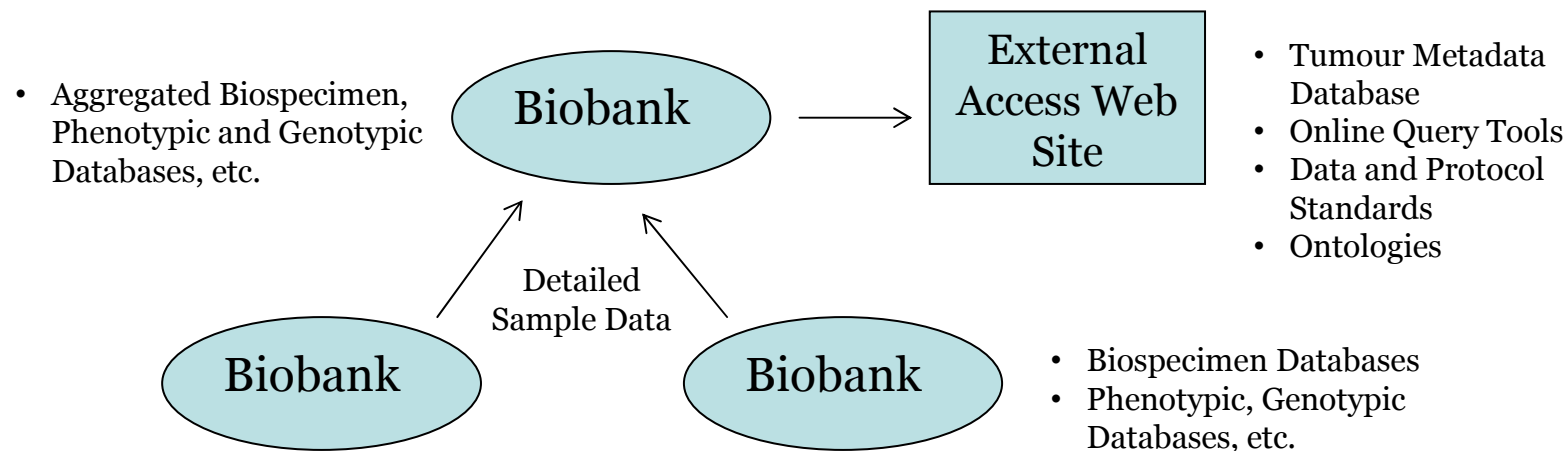
# Database deployment models – Federated Databases



- More complex solution from an IT perspective
- Allows data to be accessed in real-time
- Data may be updated in real-time
- Data security and provision become a shared responsibility
- Model adopted by BioGrid (with replication between clinical and research systems in some locations)



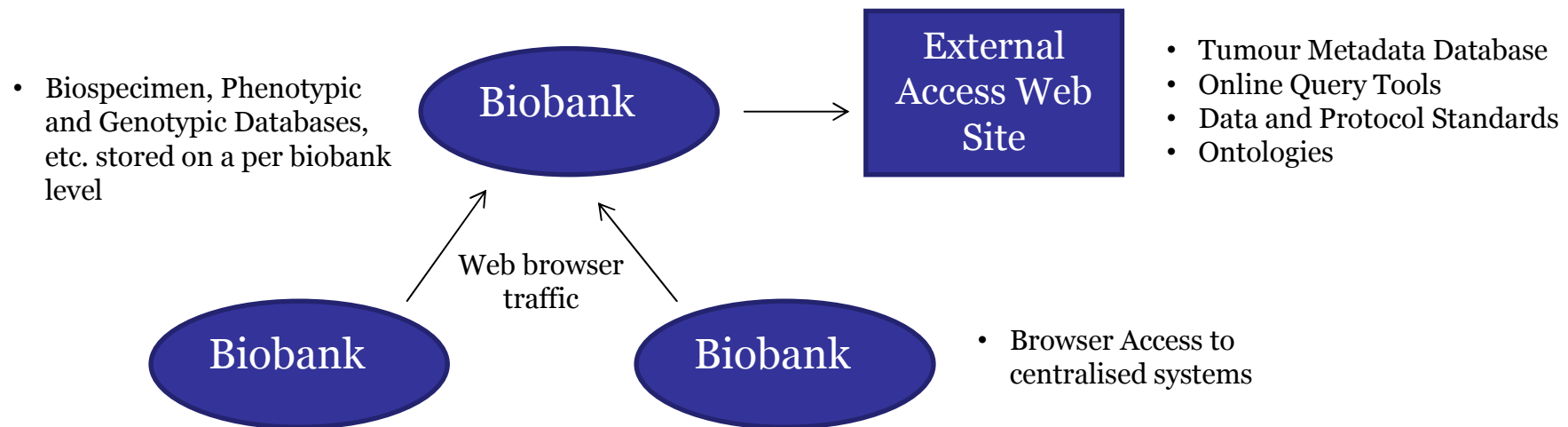
# Database deployment models – Centralised Database – Model 1



- Potentially large infrastructure cost
- Potentially large cost to integrate biobanks with centralised database
- Allows rapid access to detailed and aggregate data
- Data security and provision are a shared responsibility
- Model adopted by CaBIG



# Database deployment models – Centralised Database – Model 2



- Can provide high functionality at relatively low cost – in line with the SaaS or cloud computing model
- Requires adoption of common software for running biobanks
- Allows rapid access to detailed and aggregate data
- Data security and provision can be made a central responsibility





# What is Cloud Computing or Software as a Service?

Software as a service is sometimes referred to as "software on demand" or "cloud computing":

- deployed over the internet
- licensed as a service on demand, through a subscription, in a "pay-as-you-go" model, or (increasingly) at no charge
- part of the utility computing model where all of the technology is in the "cloud" accessed over the Internet as a service
- application delivery typically closer to a one-to-many model (single instance, multi-tenant architecture)
- centralized feature updating, which obviates the need for end-users to download patches and upgrades

SaaS Biobanking providers

- Biotracker™
- The Ark
- Bika LIMS\*
- Core Informatics\*

\* Externally hosted but not necessarily a shared instance



## Characteristics of the chosen network solutions should include:

---

- Web-enabled to promote collaboration and remote access to data
- High degree of integration between different data types – biospecimen, biochemistry, phenotypic, etc.
- Based on existing software, as opposed to being developed from scratch
- Highly configurable with minimal need for customisation
- Well-defined interfaces between systems/databases
- Based on current technology standards
- Incorporation of robust backup and data security features
- Well supported with both technical and user support



## The selection of appropriate solutions will depend on a number of factors

---

- What data standards already exist?
- What solutions are currently being used by network members?
- What funding and technical capability is available centrally and within the nodes?
- What suitable solutions are already available within the marketplace and/or under development?
- What are the target timeframes for development of the network?
- How many samples will each node manage – does this justify the implementation of externally developed software or will tools such as Microsoft Access or Excel suffice?



## Conclusions

---

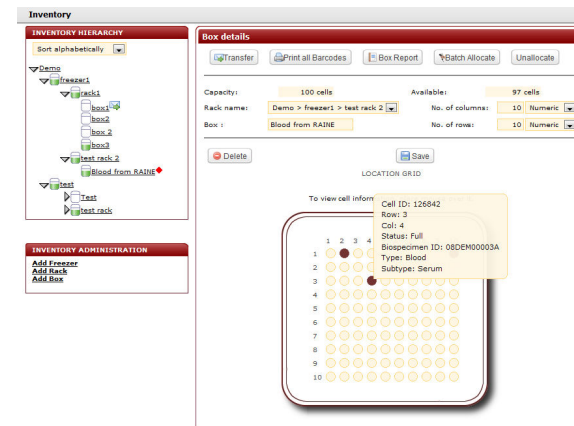
- The selection of database software and applications should be driven by the business requirements and constraints of the tumour bank network
- Different database and software deployment models may be appropriate for different parts of the overall solution
- Overall coordination of data definitions and quality standards will be important
- The implementation burden for each participating node needs to be balanced with the benefit to be derived from participation
- Establishment of a highly functional tumour banking network will require that complex business and implementation issues are addressed in a logical sequence



# Laboratory Information Management Systems (LIMS)

LIMS provide functionality for the tracking of samples, their handling, and their generated test results through the lab. Additional functionality can include:

- Workflow management
- Instrument integration and management
- Inventory management
- Sample shipment
- Cost management and invoicing
- Resource scheduling
- Protocol management



There are many solution providers in this space, both commercial and open-source, e.g. Core Informatics, Biotracker™, LabWare LIMS, Freezerworks, Bika LIMS, Biogenix



# Collection and management of other data types

---

- Online questionnaire solutions with some statistical analysis capability
  - Survey Monkey
- Paper based scannable forms
  - Teleform – Sold by Cardiff
- Online questionnaires with data management capabilities
- OBiBa – Canadian software group developing open-source biobanking software
- RedCap – freely available to research institutions
- The Ark – Australian software group developing open-source biobanking software
- OpenClinica – US-based company



# Brief overview of The Ark

---

- Founded in 2004
- Originally funded by an NHMRC Enabling Grant
- Ongoing funding from The University of Western Australia
- Currently supporting about 20 research groups
- Moving to a cost recovery model
- Transitioning to a full open source Software as a Service deployment model
- Will support separate installations at other institutions (similar to RedCap and OpenClinica)
- Currently manage our own hardware but increasingly leveraging Australian Research Collaboration Service (ARCS) infrastructure
- Application developers in WA, Vic and Montreal

