# The Ark and St John of God Hospital

## Document Purpose

This goal of this document is to describe our (The Ark team’s) understanding of the overall objectives of the WARTN research group and the likely longer term objectives for research within SJOG.

## Background

WARTN has been using the WAGER system for a number of users to manage biospecimens within the tissue bank. The short term goal of The Ark team has been to develop The Ark to a point where the data and functionality required to operate the biobank is at a point where the existing users and data can be moved over to the new system so that WAGER can be turned off.

The longer term goal of The Ark team is to provide an open source software platform that can support the core data management functions for research within SJOG. At this point it is important to draw the distinction between systems for managing operational data within the hospital, systems for collecting research data and the system(s) for managing and querying research data.

### Operational Systems

Operational systems are developed and deployed to support the ongoing diagnosis and treatments of patients within the hospital. These systems are accessed by administrators and clinicians on a daily basis to undertake their patient care responsibilities.

### Research Data Collection

Research data can be collected from multiple sources. Some of this data is collected explicitly for research purposes only and is never used for operational purposes while other data are collected to support the operational aspects of patient care but are also extremely valuable for research.

At the moment a significant amount of the operational data that is being gathered for research purposes is being transcribed from written patient records. Data is also being extracted electronically from some existing hospital systems.

A lot of work has recently gone into developing custom applications (using FileMaker Pro) that can be used to build a set of disease-specific clinical research data databases. These applications support a mix of data entry and data extraction approaches to create these research data sets.

Each of these data sets is stand-alone but other tools, such as Excel, can be used to answer queries that span datasets. Common ids, such as subject and biospecimen ids are used to link these disparate datasets. This approach works but is somewhat cumbersome and complicated.

## Future Opportunities

There are a number of approaches that can be adopted to allow WARTN and SJOG to move to a point where, not only is research more tightly embedded within the SJOG environment, but the research functions are much better supported in terms of research data collection, research data management, research data extraction and research data analysis.

### A consolidated view of the research data

One of the primary objectives of the research groups is to have a system(s) where a holistic view of the research data exists for a given disease and/or subject. This research data should include the following types of data:

* Core subject data, such as Name, DOB, etc (as required but limited by ethical considerations)
* Consent data at the different levels required
* Collection and biospecimen data
* Clinical research data (both data collected for research and appropriate data that has been collected in an operational setting)
* Genotypic and other ‘omic data
* Pedigree data
* Etc.

### Data collection at the point of clinical practice

Rather than having to implement new, custom solutions to support the extraction of data from existing clinical systems and to support data entry, a better long-term solution would be to ensure that the data required to support the operational needs of the hospital is all collected in an electronic format and then mechanisms are put in place to support the extraction of this data directly into a research databases where it is immediately available for analysis purposes.

### Data Collection for research-only purposes

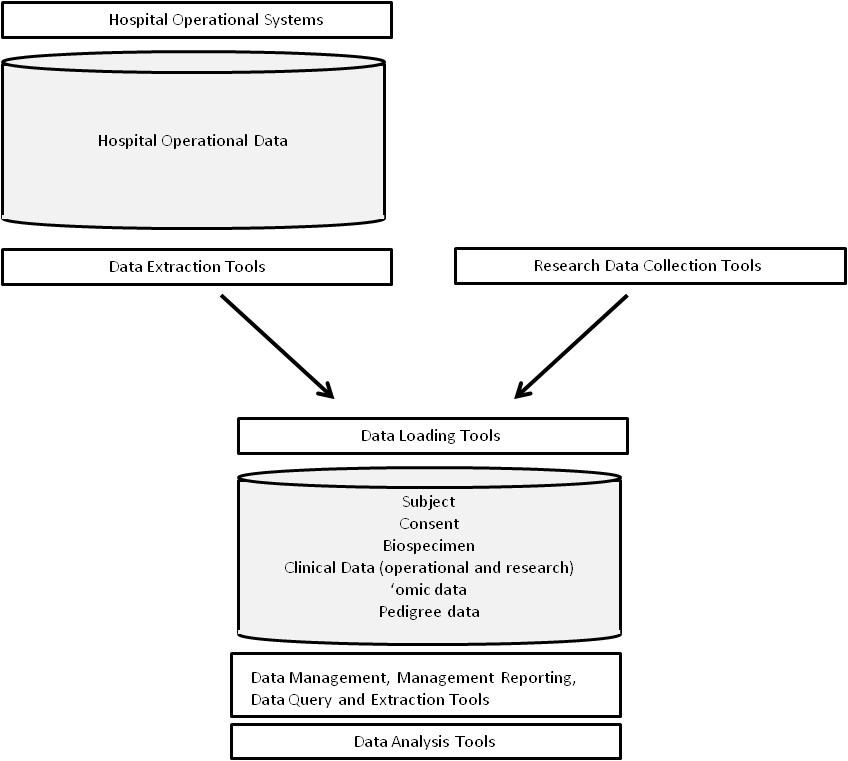
At the moment there is not a lot of data collected within SJOG specifically for research but, as SJOG moves to embrace research further then there will be a greater need for tools to support research data collection.

The sort of tools that will be required are likely to include:

* Software to support the rapid development and deployment of online patient questionnaires
* Tools to support the easy transcribing of paper-based forms into electronic formats
* Tools to support the linking of SJOG datasets with external, e.g. DLU, datasets

## A possible future architecture

It is unlikely that there will ever be a single system that can support all of the functionality required to support the research environment within SJOG. The best approach is probably to develop a commonly accepted model for integrating operations with research and then adopting an overall architecture that can support this model. We propose the following model:



### Comments

* The Current FileMaker Pro applications can be viewed as fitting into the Research Data Collection Tools category in the diagram above
* The Ark has some tools for clinical data entry but it is not currently intended as a full-featured questionnaire design tool. Existing tools, such Lime Survey, are already very well developed and supported and can be readily used by non-technical staff to develop very user-friendly data capture forms.
* Tools used for collecting research data via electronic forms should not require programming staff for their development. Suitably trained non-technical staff should be able to do this work. Again Lime Survey seems to meet this requirement, as will a number of other web-based tools.