

Title of Project: Joint Clinical Information System Research Database

Principle Investigator: James D. Brenton, Carlos Caldas, Amanda Taylor

Aim(s) of Project:

To develop a flexible database model to link clinical data with associated biological samples. The model will be

- Affordable and dependable — using free software packages based on Linux
- Open and extensible — GNU license for non-proprietary components
- Database and browser agnostic
- Scalable and configurable

Please provide a summary of the project proposal (max 1 page):

The University of Cambridge Department of Oncology has developed an open, extensible clinical information system called JCIS (Joint Clinical Information System) in partnership with Addenbrooke's NHS Trust and the West Anglia Cancer Network. The pilot project has focused on breast and gynaecological oncology clinical systems and was completed in April 2003 and externally reviewed by an independent healthcare consultancy company (SD Partners). This review was extremely favourable and confirmed that the JCIS pilot project had successfully demonstrated the utility of using a rapid application toolkit for the implementation of a eGIF-compliant clinical information system. In addition, the project also showed that it was feasible to connect multiple, simultaneous users to the JCIS system from three NHS Trusts in a secure manner across NHSnet. This was achieved with full integration of the patient administration systems and pathology systems at each Trust with JCIS.

This proposal seeks to extend the JCIS project to provide a parallel, pseudonymized clinical database system that facilitates translational research by allowing clinical and non-clinical researchers to identify tissue samples on the basis of high quality clinical and outcome data (figure). The JCIS research database will be wholly separate from the JCIS clinical information system and accessible across the internet via web-based interface screens. This JCIS breast databases already contains detailed information on more than 800 cases of primary invasive breast cancer. The Addenbrooke's Tissue Bank holds breast cancer samples for approximately 400 of these cases.

Research questions to be addressed by this project include

1. What is the best method to pseudonymize clinical data to securely protect patient confidentiality? We currently favour a model in which the only personal data item that is exported from the JCIS clinical system is the NHS number and this is encrypted using public key cryptography (PKI). Development of this requires not only software development but also the exploration of relevant

- ethical, legal and social issues resulting in a set of standard operating procedures and administrative frameworks that protect patients.
2. How should tissue access across NTRAC be implemented. How can the JCIS research database interface with third-party tools to track and check out tissue samples (tissue locator).
 3. What are the best web-based tools to interrogate the tissue locator and clinical data repository?
 4. Is the JCIS database schema (which is based on the National Minimal Cancer Data Set) sufficient for meaningful research? Should other clinical ontologies be used as the framework for data model design?

Please list 3 deliverables that the project will contribute to the UK and/or international cancer informatics community

1. A research data mirror of the JCIS clinical database accessible to NTRAC collaborators
2. Free, open source software for a standardized tissue bank database and web-based query tool
3. An XML transfer definition for the Minimal National Cancer Data Set

Please describe how the project will incorporate and/or reuse existing informatics infrastructure and/or resources. If the project will not use any existing infrastructure or resources (e.g. data standards or ontologies) please explain why this is the case.

This project will extensively reuse technology developed in the JCIS project. This includes the Royal College of Pathologists minimum clinical datasets for cancer, user interface design and database security. We will work with our collaborators at UCL (Begent, Nagl) to continue to evaluate the best ontology for clinical systems.

Please describe the plans for the sharing of data and dissemination of knowledge that arise from the project:

All software developed in this project will be available under the GNU Public license and therefore freely available. Data held within the JCIS research database will be accessible to all NTRAC collaborators (including the TUBAFROST EU consortium). It is expected that clinical data will be deposited in the public domain on publication although no framework for this currently exists.