

“Best Poster” Award Recipients

The Cancergrid Metadata Registry and Toolset (#21)

CancerGrid has developed a set of caCORE compatible user-oriented tools for the management and consumption of metadata. These tools have been used by the project and its immediate partners to: design and generate case report forms and develop annotated UML models from existing data elements; capture interoperable bioinformatic data in Microsoft Excel 2007; develop metadata standards; and uplift pathology data and tissue metadata into RDF for meta-analysis and the assembly of a tissue banks from multiple sources. This poster introduces key aspects of the work: a personal/workgroup metadata registry application that can be installed in minutes; the (included) Query Service Manager and its associated plug-ins that allow desktop software users to access terminologies, and create and consume metadata elements without leaving their application environment; and the SQIV toolset that uses W3C SAWSDL markup to uplift, transform, validate, cross-tabulate and reason about clinical and bioinformatic data using semantic-web technologies. Examples presented include: importing metadata elements from existing registries such as the NCI caDSR and the NHS Data Dictionary; UML modelling using CDEs in Enterprise Architect; "type a column range from a CDE" in MS Excel; the use of CDE and concept references from caDSR, EVS and workgroup metadata sources for designing and generating case report forms in MS InfoPath; and how run-time transformation, inference and validation between data and metadata can implement privacy policies for trial participants in data warehouses.

AUTHORS AND AFFILIATIONS:

Andrew Tsui MSc.; Charles Crichton BSc; Steve Harris PhD; Jim Davies, DPhil, Oxford University Computing Laboratory

eCRF Designer: Intuitive Dynamic Semantically Interoperable Case Report Form Designer (#32) and Intuitive Graphical User interface for Capturing Clinical Trials Eligibility Criteria Using NCI Enterprise Vocabulary Service (#33)

eCRF Designer: Intuitive Dynamic Semantically Interoperable Case Report Form Designer (#32)

One of the most challenging tasks for running a randomized controlled trial (RCT) is to design data-aware and action-enabled case reports forms (CRF) to follow up and collect participants' data. Traditionally, CRFs are paper-based although recently more electronic-based CRF (eCRF) are being used in clinical trials. Currently simple eCRFs are either created using specific templates based on proprietary software, such as Microsoft Excel, or bespoke systems. However, more complex eCRFs, that can validate data and include metadata annotations, require intelligent programming support, and are usually created by a professional programmer. Various clinical trial management systems started adding tools to create eCRFs but with limited capabilities. The caBIG Form Builder was amongst the first CRF designers to enable the creation of CRF-structures using common data elements (CDE) concepts. Using CDEs enables creating CRF-structures that are more semantically interoperable and re-usable across studies. However, these CRF-structures provide only structural content and lack data awareness, dynamic action invocation and validation, and graphical interface features and capabilities. The ePCRn

eCRF Designer has been created to overcome these limitations. It enables creating eCRFs from ISO-11179 compatible CDEs as its basic elements, which can be newly created or imported from an ISO-11179 store. These CRFs can be enriched by meta-data and graphical user interface attributes. The created forms are then automatically generated and dynamically deployed within the ePCRN clinical management system. These forms can be exported for re-use or imported directly from the caBIG caDSR repository.

AUTHORS AND AFFILIATIONS:

A. Taweel ; Delaney BC; Peterson K; Zhao L; Arvanitis T; Speedie S; Janowiec M.; Sim I; Hobbs FDR, University of Birmingham Edgbaston

Intuitive Graphical User interface for Capturing Clinical Trials Eligibility Criteria Using NCI Enterprise Vocabulary Service (#33)

One of the principal tasks for completing a randomized controlled trial (RCT) successfully is to recruit the required number of participants. For RCTs that will recruit from a pool of prevalent cases, it is possible to conduct searches of individual electronic health records (eHRs) held in clinics, visiting the clinic and running a search on the clinical record system, but this is a laborious process, and results may not be comparable between systems. The electronic Primary Care Research Network (ePCRN) is an NIH Roadmap funded project designed to construct an electronic platform for conducting clinical trials in primary care. In order to link searches with the trial eligibility criteria it is necessary to use a controlled vocabulary, allowing a choice of clinical concepts and target codes. This paper presents a prototype system that was implemented to capture various elements of eligibility criteria in primary care. The system provides an intuitive interface that allows capturing different elements of the eligibility criteria. The interface is driven by an underlying generic research object model for primary care clinical trials. It dynamically links to the Electronic Vocabulary Service (EVS) from CaBIG through a custom developed GUI interface that enables retrieving terminology and the respective available codes and coding information. Once all elements of an eligibility criterion are captured, the system allows generating an XML and/or SQL query based on a CCR structure. This query can then be submitted through the ePCRN research workbench and grid-based infrastructure to retrieve counts of eligible patients that meet the define eligibility criteria.

AUTHORS AND AFFILIATIONS:

A. Taweel ; Delaney BC; Peterson K; Zhao L; Arvanitis T; Speedie S; Janowiec M.; Sim I; Hobbs FDR, University of Birmingham Edgbaston

“Outstanding Poster” Award Recipients

Enabling Data Sharing across Borders with NAACCR and caBIG (#7)

The North American Association of Central Cancer Registries (NAACCR) is a professional organization made up of members from the United States and Canada. It is devoted to the promulgation of uniform data standards for cancer registries. These are used to improve cancer surveillance, cancer control and cancer research. NAACCR, along with a growing number of other institutions and programs, are encoding variables according to the ISO 11179 metadata standard. This standard is implemented in the cancer Data Standards Repository (caDSR) that is maintained by the National Cancer Institute (NCI). The caDSR enables researchers to develop systems that pass standard

NAACCR data elements via NCI's Cancer Biomedical Informatics Grid (caBIG™). The NAACCR standard elements can be accessed using the Common Data Element (CDE) browser at <http://cdebrowser.nci.nih.gov/CDEBrowser/>. The standard is listed on the left side of the portal under the directories of NCI Population Sciences & Cancer Control, Classifications, and the Division of Cancer Control and Population Sciences. Under the NAACCR directory there are 17 objects that cover demographics, tumors, treatment, hospitals, staging and prognostic factors for cancers. Within these objects are 382 CDEs. These CDEs specify definitions, data types and/or enumerated values for the NAACCR variables. Such detailed, internationally accessible, electronically coded metadata enables the sharing of consistent high quality data between any institutions that use this standard to encode their data. Researchers at the Utah Cancer Registry, Huntsman Cancer Institute, Intermountain Healthcare, and Utah Cancer Specialists are developing a distributed data repository using caBIG™ to examine details of clinical care for patients with colorectal cancer across Utah. The NAACCR standard will be the core component for data sharing among the participating institutions.

AUTHORS AND AFFILIATIONS:

Lewis Frey, PhD, Huntsman Cancer Institute & the Biomedical Informatics Department at the University of Utah; Antoinette Stroup, PhD, Utah Cancer Registry and the Huntsman Cancer Institute; Tao He, Huntsman Cancer Institute; Martin Cryer, Biomedical Informatics Department; Stephane Meystre, MD, PhD, Biomedical Informatics Department; Kerry Rowe, PhD, Intermountain Healthcare; Arthur Hartz, MD, PhD, Huntsman Cancer Institute and Intermountain Healthcare

Orchestrating caGrid services in Taverna (#10)

For the empowerment of users from biological or medical domains in creating and executing their workflows efficiently, the caGrid Workflow team, with the ICR working group, has selected the Taverna workbench and successfully created a prototype to orchestrate caGrid Data and Analytical services for ICR workflows. This prototype is the first step towards achieving our goal of providing an easy-to-use workflow authoring and submission tool that will be capable of orchestrating caGrid data and analytical services in executing workflows. Now, we commit ourselves to provide caGrid Workflow builder and Workflow Service as a tool which will eventually support caBIG users across workspaces in creating and executing their domain based workflows.

AUTHORS AND AFFILIATIONS:

Wei Tan¹, Ravi Madduri^{1, 2}, Kiran Keshav³, Baris E. Suzek⁴, and Scott Oster⁵, Ian Foster¹.

Guidelines on Minimum Information Collection for Antibody Therapy Experiments (#30)

Research groups developing antibody therapies are generating diverse data sets, however the value of these individual sets could be greatly compounded by amalgamating all data sets. Analysis to detect interactions between complex sets of parameters may yield further reaching interactions when performed on larger scaled data sets. Valid comparisons between experiments would facilitate the detection of research areas not yet fully explored. In order to achieve functional amalgamation, standards for collecting data from experiments must be defined. According to the NCRI Planning, these standards need to be defined particularly in the areas of pre-clinical and clinical development of

antibody-based drugs. This poster describes the creation of common data elements (CDEs) as guidelines for collecting minimum information from experiments on antibody therapies. The CDEs conform to the ISO11179 standards, concepts used to build the definitions and properties of the CDEs are sourced from the controlled vocabulary provided by NCI Thesaurus. CDEs are created for fields identified on the Guidelines for Information About Antibody Therapy Experiments (GIAATE) tree, developed by the Antibody Society. The CDE value domains link to existing databases identifiers where possible. For example, the protein structure of the antibody target is a field which links into Protein Data Bank through the protein structure identifier. The CDEs are presented through forms, each form containing a field of data to be collected, the field's definitions and the context in which the field is to be used. The forms also present acceptable data values and formats. A server will be hosted to allow researchers to download forms, to propose changes to current CDEs in terms of data values, formats or contextual use and to allow researchers to develop new CDEs, specific to their areas of antibody therapy research.

AUTHORS AND AFFILIATIONS:

Richard Begent, Professor, Cancer Institute University College, London; May Yong, PhD, Cancer Institute University College, London; Igor Toujilov, PhD, Cancer Institute University College, London; Sylvia Nagl, PhD, Cancer Institute University College London

Applying Astrogrid Techniques to the Analysis of Tissue Microarrays (#34)

Antibody hybridization and image scanning of tissue microarrays (TMAs) is a highly automated process, but the subsequent manual scoring of TMAs by a trained pathologist is a major bottleneck in their analysis. We can overcome this bottleneck by applying techniques developed for astronomical imaging data, as part of the global Virtual Observatory initiative, to the analysis of TMA images, and build pipelines which automate the analysis, acquisition and querying of TMA data. We describe an initial test case application for the automated 'scoring' of Estrogen Receptor (ER). ER is an important regulator of mammary growth but is also known to play a role in breast cancer. Assessing ER in the clinical setting enables decisions to be made as to which care programmes should be followed by patients. Pipeline automation, the basis of our 'PathGrid' system, has been implemented by adapting components from the UK's AstroGrid virtual observatory project. The processing pipeline, built from these components, can be executed as a workflow using a plug-in for Taverna ("Astro-Taverna"). The processing algorithms themselves are adapted from those developed to handle deep sky astronomical imaging programs. In our initial pilot study we assess ER in approximately 500 tumors from a large population-based clinical trial (SEARCH; part of the Anglia Breast Cancer Study). The poster describes the workflow covering the login process, file transfer, image conversion, image analysis, generation of results, and storage of resulting images and files on either a local or virtual file system. We also validate our ER scoring algorithm by comparing the results with scores manually assigned by a pathologist. The Pathgrid system should have a significant impact in improving the quantitative analysis of a range of TMA markers, providing increased throughput and objective assessment of expression.

AUTHORS AND AFFILIATIONS:

Nicholas Walton, PhD, University of Cambridge; James Brenton, MD, PhD, University of Cambridge; Carlos Caldas, MD, University of Cambridge; Mike Irwin, PhD,

University of Cambridge; Asif Akram, PhD, University of Cambridge; Peter Macallum, PhD, University of Cambridge; Nikita Makretsov, MD, University of Cambridge; Lorna Morris, PhD, University of Cambridge