



NCRI
INFORMATICS
INITIATIVE



NCRI
National
Cancer
Research
Institute

Newsletter

WINTER 2008 ISSUE 12

Welcome to Issue 12 of the NCRI Informatics Initiative Newsletter

2009 promises to be a busy year for the Informatics Coordination Unit, between the first launch of a production version of ONIX, a period of extended user acceptance testing, and the wider, main launch of ONIX. As well as all of this, the 3rd annual Joint NCRI/caBIG[®] conference is already being planned, and will be held in London over 10-11 of September, so please make space in your diaries.

The Platform Development team have been working hard going through another round of User Group testing and testing of the pre-production version of ONIX. The first production version of ONIX will be available in January 2009. This first live version will be made available to a controlled number of users initially, while it undergoes an extended round of 'user acceptance testing'. This period of

user acceptance testing will allow ONIX to be refined in light of feedback from a selection of real-world end-users. The ICU has selected a sub-group from the funders of the Initiative, and asked them to identify suitable researchers and physicians to act as this test group. Following this community testing phase, ONIX will be made generally available to the cancer research community around the middle of 2009.

Recently the ICU presented a progress update to the Informatics Management Board (IMB), as part of its regular 'Gate Review' process. The IMB agreed that the ICU had passed all of the requirements of the Gate Review and all ONIX development tasks were agreed as being on-time and on-budget.

We hope you enjoy this issue!

INSIDE THIS ISSUE:



SPECIAL FEATURE
[Enterprise Support Network](#)



UNIT ACTIVITIES
[ONIX v1.0](#)



MEETING UPDATE
[CODATA conference](#)



UNIT NEWS
[Data sharing Debate](#)

If you have any queries or comments on any article in our newsletter, or would like to contribute to the next issue, please contact us at:

NCRI Informatics Initiative
PO Box 123
61 Lincoln's Inn Fields
London, WC2A 3PX

Email us at:
info@cancerinformatics.org.uk

SPECIAL FEATURE

caBIG® Enterprise Support Network

Connecting the Biomedical Research Community

The cancer Biomedical Informatics Grid® (caBIG®), developed under the coordinating direction of the US National Cancer Institute's Center for Bioinformatics and Information Technology (CBIT), is a set of interoperable open source tools, standards and infrastructure designed to speed basic and clinical cancer research through collaboration and connectivity. The caBIG® community includes over 1,000 individuals representing more than 200 organisations from academia, government and industry, both national and international. To support broader adoption of caBIG® tools, standards, and infrastructure, caBIG® has launched the **Enterprise Support Network (ESN)**, a major expansion of support resources to serve the needs of the growing caBIG® community. The ESN consists of two primary components: **Knowledge Centers** and **Support Service Providers**.

Knowledge Centres

caBIG® Knowledge Centers (KCs) are based at major research institutions and organisations, and are staffed by IT specialists, bioinformaticists, and basic and clinical researchers with extensive expertise in a particular domain, the tools that support that domain, and the application of those tools to address research questions. caBIG® KCs provide

a centralised knowledge repository, offer web-based support, and provide expert guidance to end users, IT staff and senior decision makers interested in adopting caBIG® standards, tools, and infrastructure. The caBIG® KCs can be accessed at https://cabig.nci.nih.gov/esn/knowledge_centers and include:

caGrid KC, led by The Ohio State University and The Ohio State Comprehensive Cancer Center, with University of Chicago and the Argonne National Laboratory, provides support to organizations wishing to implement caGrid, the underlying network architecture that provides connectivity among institutions, allowing researchers to share data and services.

Clinical Trials Management Systems KC, led by Duke University Comprehensive Cancer Center, with Northwestern University, Cancer and Leukemia Group B–InformationSystems (CALGB-IS), and SemanticBits, provides support to organizations using caBIG® tools for the management, conduct, and reporting of clinical trials.

Molecular Analysis Tools KC, led by Columbia University Herbert Irving Comprehensive Cancer Center with the Broad Institute of MIT and Harvard provides support to institutions and individuals using caBIG® tools for the management and analysis of genomic data.

Tissue/Biospecimen Banking and Technology Tools KC, led by Siteman Cancer Center, Washington University at St. Louis, provides support to organizations using caBIG® tools for the management, annotation, and tracking of biospecimens.

Vocabulary KC, led by Mayo Clinic with SemanticBits, provides support to individuals and organisations using the distributed vocabulary services infrastructure to code biomedical information in order to enable the integration of disparate data sources.

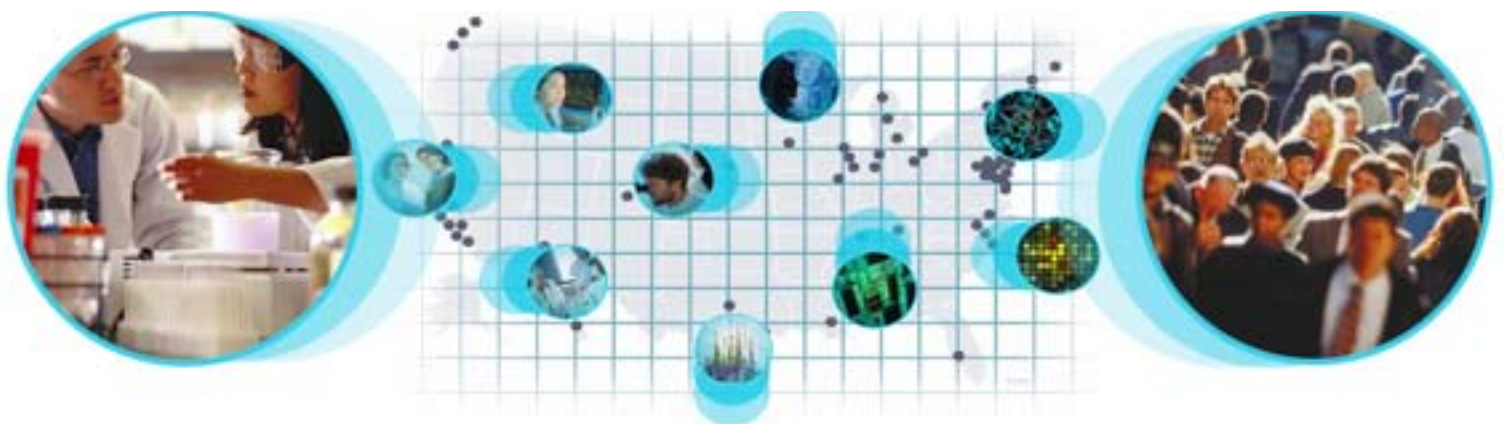
Data Sharing and Intellectual Capital KC, led by Michigan, provides a centralised, authoritative repository of processes, model agreements, and other resources to encourage and facilitate data sharing to advance scientific discovery, consistent with applicable legal, regulatory, ethical and contractual requirements.

Support Service Providers

caBIG® licensed Support Service Providers (SSPs) comprise a second component of the ESN. These are third-party organisations that deliver comprehensive technical and end-user support. Each Support Service Provider has been evaluated based on their biomedical domain expertise and technical capabilities. Service contracts are negotiated between Support Service Providers and their clients. Specific categories of services offered include:

- Help Desk Support
- Adaptation and Improvement of caBIG® Compatible Software Applications
- Deployment Support for caBIG® Software Applications
- Documentation and Training Materials and Services

Additional information about the caBIG® Support Service Providers can be found at https://cabig.nci.nih.gov/esn/service_providers. ■



Gateway to the cancer Biomedical Informatics Grid.

UNIT ACTIVITIES



ONIX v1.0

A voluntary community network enabling access and integration of cancer information

Since its inception, the NCRI Informatics Initiative Co-ordination Unit (ICU) has been working on implementing a voluntary network or “grid” environment (the **Informatics Platform**) that will increase access to and integration of data and resources generated through cancer research. This requires the co-ordination of both cultural change and development of technical solutions.

On the technical implementation of the Informatics Platform, the ICU has been gathering pace for the first production release of the **NCRI ONcology Information eXchange (ONIX)** scheduled for the end of January 2009. ONIX is the internet portal visible and accessible to users of an underlying grid infrastructure. Its aim is to connect heterogeneous and geographically distributed resources useful to cancer researchers and provide seamless and secure exchange of data with other grid-enabled resources.

The demand for such an infrastructure is timelier than ever; scientific breakthroughs in technology have made high-throughput generation of heterogeneous research data a reality while the implementation of e-health solutions has increased the availability of patient-related data in electronic format. Data access and integration is therefore the key to unlock the much wanted, belated reality of systems

biology and translational medicine.

For the community, by the community

ONIX has been developed in a step-wise fashion with active community involvement at every step through the NCRI Informatics Task Force and the ONIX User Group that provided valuable feedback on functionality and scientific requirements. All this major input into the strategic development direction of ONIX has been included in this first release that aims to provide a testing bed to allow experience from initial work to benefit later versions and activities.

This first release sees the establishment of a core functionality, initially consisting of a portal, user interface (UI) and catalogue of sources and services (the **Resource Catalogue**), coupled to a framework of standards and tools for community use. A software tool will also be available to facilitate the linking of resources into ONIX.

The Resource Catalogue

The main feature of ONIX v1.0 is the Resource Catalogue, which constitutes a central collection of registered resources to allow both information to be known about a resource, even if not physically connected to the grid itself, and the rapid filtering of potentially useful resources a specific user might be interested in interacting with. Resources comprise anything from biomedical data, data analysis applications, supporting tools, to new research technologies and projects developing community data standards.

The Resource Catalogue is populated through a collaborative community-based

effort that all resource owners are welcome to participate in. It is an international effort that seeks to unite not only UK resources, but leverage assets on a global scale by including resources from our strategic partners, EBI and NCI’s caBIG®.

Minimal Information for Resource Registration

Resource registration entails the entering of metadata, a minimum set of information that is sufficient to provide a basic description of that resource to the wider members of the cancer research community.

Registered Resources vs. Connected Resources

Upon initial submission there is a formal review step undertaken by the ICU, particularly to ensure consistency in the categorisation of the resource. Once the ICU has assessed and then added the information to the resource catalogue the resource is granted certification level 4. This is the lowest level of ONIX interoperability (certification) and is driven by the fact that the resource is not physically attached to the ONIX grid infrastructure in a way that allows any form of searching of its contents to be undertaken.

Throughout 2009, there will be a steady increase in both resource registration and grid-connectivity. The unit will provide the necessary training, information and support for the cancer research community to build adapters that will enable them to link their data and analysis tools to the ONIX grid. ■



The ONcology Information eXchange (ONIX) homepage.

MEETING UPDATE

21st International CODATA Conference

5 - 8 October, 2008
Kyiv, Ukraine



The Committee on Data for Science and Technology (CODATA) is an interdisciplinary scientific committee of the International Council for Science (ICSU), was established 40 years ago with the fundamental aim of fostering worldwide cooperation in scientific and technical data. One of its primary activities is an international biennial conference on data, which attracts approximately 300 data specialists from around the world.

The NCRI Informatics Initiative made a joint presentation with NCI's caBIG[®] given by Dr Leslie Derr, (NCI CBIIT) entitled "Collaborative data-exchange mechanisms: a transatlantic example". Our presentation showcased this extremely successful and fruitful collaboration between the two initiatives to establish an information network without borders to connect the cancer research community and enable the sharing of tools and data, and the meaningful integration of data in order to accelerate the discovery, development and delivery of personalised molecular medicine.

One of the success stories demonstrating the sharing of data and services across the Atlantic is the collaboration between NCRI Imaging and Pathology project and caBIG[®] Imaging Work Space. By using caBIG[®] and NCRI technologies the two teams have established a system to support simultaneous review of multiple, integrated radiology and pathology images to improve quality of surgery.

For more information on CODATA, please visit: <http://www.codata.org/> while, browsable images from the NCRI Imaging and Pathology project and other relevant information can be found at the University of Leeds Virtual Pathology website (<http://www.virtualpathology.leeds.ac.uk/>). ■

UNIT NEWS

NCRI Cancer Conference Data sharing Debate

6 October, 2008
Birmingham, UK



The purpose of the debate hosted by the NCRI Informatics Initiative was to revisit the benefits and drawbacks of sharing research and healthcare data and ask the community what are the major barriers that need to be overcome for a wider adoption of a data sharing culture.

The panel consisted of experts with a diverse background from academia (Professor Sylvia Nagl, UCL), industry (Dr Tim French, AstraZeneca and Kevin Johnson, PanGenetics), scientific publishing (Iain Hrynaszkiwicz) and funding bodies (Peter Dukes, MRC). The aim was to provide the audience an opportunity to hear viewpoints they might not otherwise consider and initiate a floor discussion around the issues raised.

When the chairman, Professor Richard Begent, asked the participants to vote, the majority voted 'yes' to data sharing!

Although the audience showed willingness to share, it seemed apprehensive recognising that there are still cultural and technical challenges that need to be addressed and requested: better education, training and support; more examples of the 'real' benefit; recognition incentives; and a mechanism indicating data quality and provenance.

More specifically, for the academic community, although it is the one reaping the immediate benefits of data sharing

through new hypothesis generation and additional publications through collaborations, it was obvious that some members still have two major concerns: the threat to intellectual property and commercialisation potential; and the critique and misinterpretation of their data.

This debate provided reassurance that the community is supportive of data sharing and that the issues raised can be addressed successfully through better education and new funding mechanisms.

For more information please visit: <http://www.cancerinformatics.org.uk/datasharing.html>. ■

Cyberinfrastructure wiki

In the September 2008 issue of Nature Reviews Genetics, Lincoln Stein published an excellent article entitled "Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges".

This article not only is an excellent review of the current status of the various components, international efforts and challenges to deploy such an infrastructure but it also comes with a companion wiki page. The intention is to continue keeping the community up to date with current and pending implementations.

We would like to applaud this effort and we have added the NCRI Informatics Initiative as one of the "Cyberinfrastructure projects".

The article and wiki page can be accessed at: <http://nrgwiki.nature.com/cyberinfrastructureforbiology/show/HomePage>. ■



Hands up for data sharing!