

## **youShare, an online collaboration research environment for sharing data and services.**

Jim Austin, Martyn Fletcher, Tom Jackson, Mark Jessop, Aaron Turner, Michael Weeks.  
Advanced Computer Architectures Group, Dept. of Computer Science, University of York, York, YO10 5DD. UK.  
Email: {austin, martyn.fletcher, tom.jackson, mark.jessop, aaron, michael.weeks}@cs.york.ac.uk

### ***Introduction.***

Increasingly, research collaborations need to share large amounts of data and perform computation intensive analysis tasks quickly and securely. The youShare infrastructure is designed to provide researchers with platform to allow them to run compute-intensive research and allows collaborators to interact with the actual experiment, data, analysis or results. The aim is to promote collaboration by allowing data, programs, and research outputs to be shared flexibly and securely between groups of researchers. This paper overviews the youShare system and its capabilities.

### ***Background and overview.***

Many environments exist that can be used for collaboration, from the simple online sharing of files provided by Dropbox [1] to collaborative environments such as myExperiment [2] where scientists can safely publish their workflows and experiment plans. youShare takes collaboration a step further through combining the sharing of data, metadata, services, workflows together with an execution environment. youShare ([www.youshare.ac.uk](http://www.youshare.ac.uk)) is a development of the Code Analysis, Repository and Modelling for e-Neuroscience (CARMEN [www.carmen.org.uk](http://www.carmen.org.uk)) [3] project that provides a unique virtual neuroscience laboratory: an infrastructure for using and sharing data, tools and services that is now in regular use by neuroscientists. CARMEN is the first fully operational portal system for sharing neurophysiological data and services and is a valuable example of an operational portal for scientific use. The youShare infrastructure extends the CARMEN model and allows researchers any discipline to create collaborative groups through which to work collaboratively using shared data and software accessed through a web portal. The system provides a secure online environment for the storage of experimental or simulation data and metadata, and storage and execution of analysis code (services), together with controlled and secure sharing of data and services with other chosen users or groups of users. The architecture consists of federated compute and data resources at the University of York, which form part of the White Rose Grid computing infrastructure [4] (Fig. 1). User interaction with the system is via the youShare portal (accessed via a web browser) which presents the resources to users in a conceptually centralised manner where data can be uploaded, analysis services executed, etc. youShare plans to see how such aims can be supported, and what barriers there are to operating such a service, and to determine how these barriers can be overcome. The following subsections briefly describe the youShare infrastructure in terms of data storage, services and workflows, the portal and the use of collaborative groups.

### ***Experimental and simulation data storage and metadata.***

The infrastructure allows a user to store data from, for example, an experiment, e.g. time series, optical, etc. and the data outputs produced by the execution of services. The system affords registered users the ability to upload experimental data which can be described with extensive metadata and also to apply sharing policies. Presently, flat file data, including raw experimental recordings, are held in the Storage Request Broker (SRB) [5] installation within the youShare system and metadata are held in a database, providing native search and indexing functions. These resources can be searched, shared with other users, downloaded, visualized and analysed by user supplied code. For each of the resource types in the youShare system a comprehensive metadata definition is documented, for example neuroscience electrophysiology experimental data are described by the Minimum Information about a Neuroscience Investigation (MINI) [6] standard. In order to simplify the metadata entry process, users can create pre-populated templates, which will auto-populate the wizard data entry forms, and allow the user to edit just the fields that change on a regular basis for a particular experimental domain. Auto-population of the metadata schema from external XML or other document structures is possible but is not currently implemented. Data upload is managed through a JAVA applet running within the browser which can handle multiple, parallel streaming across many HTTP ports, to accommodate large file sizes, and which also manages error checking and other server-client transaction issues. Access Control is critical to the management of the data upload process, and before a user can complete a data upload process, they are required to specify the access control lists for both the data and metadata for the file. These are easily and quickly configured, remain under the control of the user completing the upload, and can be modified at any future point in time to add additional users or groups, or to widen or reduce access.

### ***Services and Workflows.***

The youShare infrastructure allows users to create their own services, run these against data held in the system, and to share their services with other youShare users. For the user, this is all achieved via the portal's browser interface, and all data and processing are contained within the youShare system. A youShare service consists of

youShare, an online collaboration research environment for sharing data and services.

a (slightly modified) command-line application that is wrapped inside a Java dynamically-loadable class, and embedded into a JAR file. Python, Matlab (as a compiled executable), Perl, R scripts, or any binary executable can be wrapped into a service. Service wrapping is achieved quickly using simple form-based web pages within the youShare Portal. The wrapping process also generates service metadata that are required to create the browser user interfaces, provide information to the youShare user, deploy services onto a distributed processing node, and to configure service client interfaces. After generation of the JAR file it is stored in the youShare service repository which is held within the SRB, and the service is registered with the youShare service execution engine. Service execution is orchestrated by the user via the portal, though the service actually runs within the distributed processing nodes. The user selects a service, input values and/or files in the portal, and a request is dispatched to an available processing node. The processing node fetches the service (if not already deployed) and the data then configures a generic client using information from the service metadata. The service is then executed, and upon completion, the youShare portal displays the result to the user and stores it in the user's file space. A workflow and execution engine will be integrated into the youShare system in future.

### ***The youShare Portal.***

As described above, the user interacts with the youShare system via the web portal (portal.youshare.ac.uk). The portal presents the systems and its resources to users in a conceptually centralised manner where data can be uploaded (Fig 2), analysis executed, etc. The portal is an AJAX (Asynchronous JavaScript and XML) web application that has been developed and implemented using the open source Google Web Toolkit [7] (GWT). GWT permits the application to be developed and tested in Java using standard software engineering practices and then compiled into cross browser JavaScript for use in any browser. Users can also use desktop tools such as the Signal Data Explorer (SDE) [8] on their desktop to interact with the portal based tools for analysis and visualisation.

### ***Use of Collaborative Groups.***

As stated previously a user can share the various resources with other collaborators. However, youShare also provides users with the ability to create collaborative groups so that resources can be shared only within the group, if desired. The existence of the group can also be made invisible to other users (outside the group) so that the existence of the collaboration is not discernable from youShare system, as desired.

### ***Conclusions and future work.***

The youShare philosophy has many possible applications in other domains and these are being actively pursued. Further third party and legacy tools and services are being integrated into the youShare infrastructure, together with the capability to further develop and execute workflows. In future it will also provide facilities for the development and testing of new services and tools. Through providing an environment where data, tools, services and workflows can be shared and developed youShare reduces the need for re-running of complex and often expensive experiments, and promotes more effective collaboration. Lessons learned will benefit the development of tools for other applications. Ultimately the goal is to provide a cost effective service for UK academics that provides a way to streamline research, collaboration and re-use of software and data to allow UK research to be enhanced and improved. It is intended to link into the shared UK academia compute facilities to provide the resources to run compute jobs on behalf of researchers.

### ***Acknowledgements.***

The funding for youShare is provided by the HEFCE University Modernisation Fund. CARMEN is supported by BBSRC under contract BB/I000984/1.

### ***References.***

- [1] Drop box: <https://www.dropbox.com/> - retrieved 10<sup>th</sup> June 2011.
- [2] myExperiment: <http://www.myexperiment.org/> - retrieved 10<sup>th</sup> June 2011.
- [3] The CARMEN Neuroscience Server. Paul Watson, Tom Jackson, Georgios Pitsilis, Frank Gibson, Jim Austin, Martyn Fletcher, Bojian Liang, Phillip Lord. UK e-Science 2007 All Hands Meeting, Nottingham, September 2007.
- [4] The White Rose Grid: <http://www.wrgrid.org.uk/> - retrieved 10<sup>th</sup> June 2011.
- [5] Storage request Broker (SRB): [http://www.sdsc.edu/srb/index.php/Main\\_Page](http://www.sdsc.edu/srb/index.php/Main_Page) - retrieved 7th May 2011.
- [6] Frank Gibson et al., "Minimum Information about a Neuroscience Investigation (MINI) Electrophysiology," Mar. 2008; <http://hdl.handle.net/10101/npre.2008.1720.1>.
- [7] The Google Web Toolkit <http://code.google.com/webtoolkit/> - retrieved 10<sup>th</sup> June 2011.
- [8] Neural Network Based Pattern Matching and Spike Detection Tools and Services in the CARMEN Neuroinformatics Project. Martyn Fletcher, Bojian Liang, Leslie Smith, Alastair Knowles, Tom Jackson, Mark

youShare, an online collaboration research environment for sharing data and services.

Jessop, Jim Austin, pp. 1076-1084, Neural Networks, Special Issue on Neuroinformatics, Volume 21, Issue 8, published by Elsevier Ltd. October 2008.

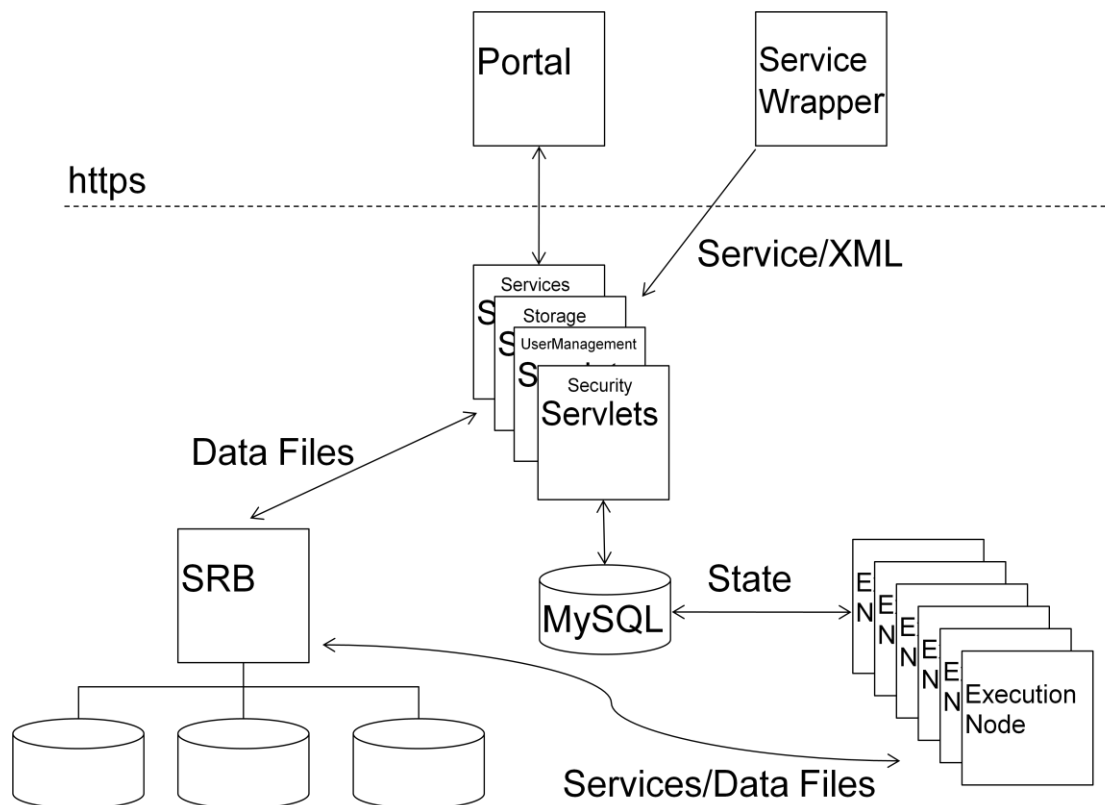


Figure 1 youShare Architecture

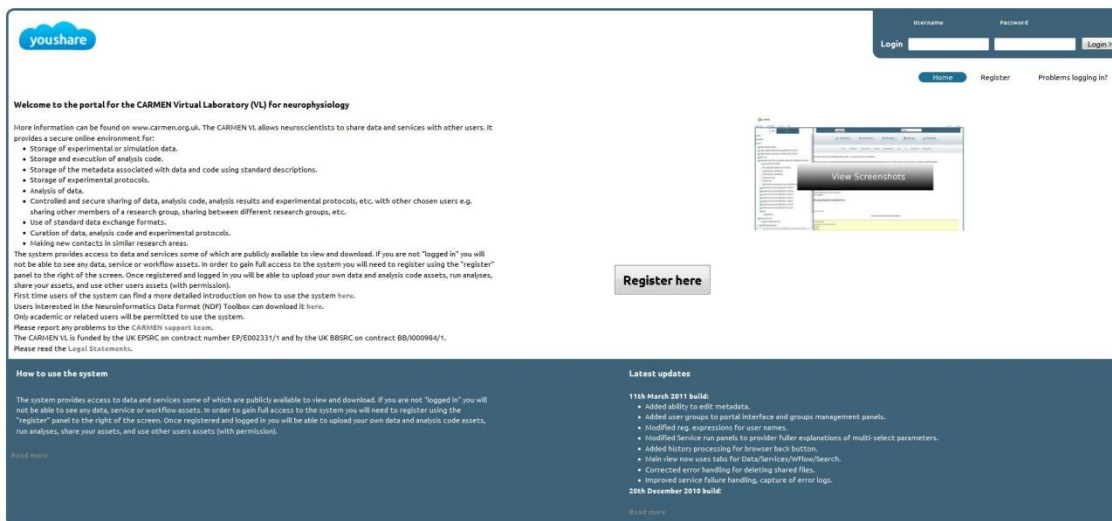


Figure 2 youShare Portal