BACKGROUND

CSIRO is a large, geographically dispersed research organisation with over six-thousand staff at fifty-six sites. Scientists at CSIRO carry out research in various domains and deal with large complex datasets, with the need to collaborate effectively and efficiently.

In response to these issues, geographical separation of staff and increasing data complexity and the need for effective collaboration, an architecture for software and systems that supports eResearch within CSIRO [1] has been developed and is evolving to take advantage of new opportunities to support the work of scientists within the organisation. Several projects have already been completed and several are underway to build the underlying infrastructure and software systems that support eResearch.

ERESEARCH AT CSIRO

A CSIRO-wide Research Data Management Service (RDMS) is being put in place with software and services to enhance the publication, discovery and reuse of data assets. RDMS handles data brought in from external sources, from eLab Notebooks, and from the various science research tools. It provides a data management, search and publication capabilities. An overview of the architecture for RDMS is shown in Figure 1.

As well as research data management service there are other areas that are being supported through the development of eResearch infrastructure capabilities. These include:

- Advanced scientific computing – high performance computing infrastructure for modeling and data analysis;
- Visualisation and imaging – high resolution hardware and advanced software;
- Networking – high speed connections between CSIRO sites, including regional areas;
- Data storage – protecting data and improving its availability;
- Collaboration tools – enabling real-time interactions between scientists in disparate locations;
- Virtual research environments - an integrated set of on-line tools, systems and processes enabled by sensors and sensor networks, instruments and computational, data and visualisation resources;
- Cloud computing services giving scientists cost-effective, scalable and simplified access to data storage, computing power and software.
ERESEARCH ARCHITECTURE

An overall architecture for eResearch is being developed that encompasses the architecture for serveral software systems that support the eResearch vision within CSIRO. The architecture for these system uses the capabilities of the underlying infrastructure and build on these to provide the necessary capabilities to the scientists in CSIRO. Two examples of the types of systems that are being built are shown in Figure 2 and Figure 3.

Figure 2: Architecture for GeoPortal

Geoportal is a system that brings together the metadata describing all of the geospatial data within CSIRO. It is meant to be the single location that a scientists in CSIRO goes to to find geospatial data.

The eTools (Science Workflow) system enables the reuse of workflow models and algorithms across CSIRO. It uses a service and model registry where service components can be deployed on a cloud infrastructure. The cloud infrastructure should make available the underlying compute capabilities to the scientists that that need to run their analysis components.

ARCHITECTURAL CHALLENGES

There are several challenges in developing an architecture for an organisation as large and diverse as CSIRO. Some of these challenges include:

1. Diversity in research domains and scientists needs. Being able to architect solutions for a wide variety of domains (e.g. Mining, eHealth, Marine, Land and Water) and supporting the different needs of the scientists poses significant challenges in building systems and software;
2. Several pieces of the infrastructure and capabilities have not been used in the organisation previously. Introducing new technology and systems is a challenge and working closely and communicating with the scientists to show the benefits of the new technology and systems is important;
3. The solutions need to be scalable and meet the usability requirements of the scientists. Some systems deal with massive volumes of data (terabytes to petabytes) and need to be scalable and meet performance, availability and other non-functional requirements;
4. Integration and interoperability are challenges that are being addressed as part of the architectural solution to eResearch. Scientists may use a wide variety of tools and technology to support their research work so it is important that these tools can work together to support the workflows and processes of research whether it is for research data management (RDMS), capture of results and IP (eLN), or invocation of manipulation, analysis and presentation capabilities (eTools).

REFERENCES

ABOUT THE AUTHOR(S)
Liam O’Brien is the Chief Software Architecture for CSIRO eResearch. At CSIRO he is involved in architecting eResearch solutions based on service oriented architectures and Cloud computing. He is also an Adjunct Research Fellow at School of Computer Science and a Visiting Fellow at the School of Accounting and Business Information Systems at the Australian National University. His research and technical interests also include software and service oriented architecture, reengineering, business transformation, enterprise architectures and cloud computing. He has worked at NICTA (Australia), Lero (the Irish Software Engineering Research Centre) and at the Software Engineering Institute (at Carnegie Mellon University in the US). He holds a PhD and BSc from the University of Limerick, Ireland.