Developing Applications using the e-Science Central API

Simon Woodman, Hugo Hiden, Paul Watson and Jacek Cala
School of Computing Science, Newcastle University, Newcastle upon Tyne, UK
{s.j.woodman, h.g.hiden, paul.watson, jacek.cala}@ncl.ac.uk

Many scientific research projects face severe challenges in extracting value from the increasingly large volumes of data they generate. From our work on a large number of e-science projects, we have identified four activities that are of prime importance for the scientists we have collaborated with:

- Storing the data from their experiments
- Interactively exploring and analysing the data
- Automating the analysis once a repetitive technique has been devised
- Sharing both the data and analysis services with selected colleagues and organisations

Each of these activities presents its own challenges. Storage can be a problem as experiments can now generate vast quantities of data. Analysing the data can therefore require large processing capacity, and, even if the scientist has access to such resources, the cost of moving the data to them can be a bottleneck. Sharing data and services requires a way to provide controlled, remote access to colleagues irrespective of their location, at all stages of the scientific process, and demands the presence of metadata so that those re-using data understand its format and meaning.

We were unable to find any 'off the shelf' way to meet these needs although there are some components which are able to address some of these issues separately. As a result, over the past four years, we have been pursuing an alternative, which is to design and build e-Science Central– an integrated 'Science as a Service Cloud' [1]. This has allowed us to explore the exploitation of three recent trends in computing to address the needs described earlier: *Software as a Service* - e-Science Central is provided as a website allowing scientists to use it from any browser, upload and analyse data and share data and services; *Cloud Computing* - to provide resources that scale with the number of users, volume of data and complexity of analysis; *Social Networking* - to support sharing of data and services and to facilitate user collaboration.

With e-Science Central, users can therefore write services, upload them into the cloud, combine them in workflows, and use them to analyse uploaded data. The data and

analysis results can then be shared with other users (or groups of users). This can all be done through a browser, so removing the need for users to install and maintain software on their own desktops or servers. In order to support a sophisticated level of functionality, for example in-browser workflow editing, HTML5 and javascript are exploited.

We have also discovered that, as well as a browser based interface for users, many scientific research projects also require a custom user interface. This may take the form of a web, desktop or smartphone/tablet application depending on the nature of the project. A common example of this is the need to provide a simple web page interface that allows users to upload and analyse data.

To address this need, e-Science Central also offers developers the ability to leverage its features via an API (Application Programming Interface). This HTTP/REST based API allows programs to exploit the same functions as are are available to users through the e-Science Central website, including uploading data, attaching metadata, executing workflows, and even creating new services. Tooling such as client libraries and sample projects exist to ease the development of applications in Java, .Net and PHP, allowing the developer to focus on the domain specific functionality. This API allows the full functionality of e-Science Central to be exploited, without it being exposed directly to users.

Given the importance many researchers place on maintaining the security of their data, this was a major consideration in the design of the API. Security is managed in two ways in the API: firstly all applications have a unique identifier and key which are generated by e-Science Central. These are used to sign all the HTTP requests sent from the application to e-Science Central. The requests also contain a random number as part of the request which is used to prevent replay attacks. This random number is regenerated periodically and then incremented for each request. Each request is validated on receipt by e-Science Central to ensure that the identifier is valid, the signatures match and the random number incremented correctly. The second security measure is that users must explicitly grant access to an application in order for it to perform certain functionality. Granting access is not done on a per resource basis, rather on the type of action the application is allowed to perform, e.g. read files, modify files, execute workflows etc.

Over the past year, the API has been used in a wide-ranging set of scientific projects, including medicine, chemistry and the arts [2]. This includes analysing the well-being of older people from accelerometer data, analysing spectrographic images and combining photographs and recorded sound. One particular usage that we believe will become increasingly important is to support data collection and analysis from mobile devices such as smartphones. Smartphones are attractive for the 'in-the-field' capture of data, including sounds, photographs, and video, as well as from internal and external sensors (e.g. accelerometers and heart rate monitors). However, they have limited processing capabilities, and the purely local storage of data precludes co-analysis with data collected from other phones. Therefore, smartphone applications have been developed that use

the e-Science Central API to upload data from the phone to the cloud for storage and analysis.

The paper will give the motivation for developing the e-Science Central API to allow integration with external applications; it will then present the API and discuss the key design decisions behind it. Finally, examples of its use for a set of real-world examples will be presented.

References

[1] Paul Watson, Hugo Hiden, and Simon Woodman. e-Science Central for CARMEN: science as a service. *Concurrency and Computation: Practice and Experience*, 22:2369–2380, December 2010.

[2] Hugo Hiden, Paul Watson, Simon Woodman, and David Leahy. e-Science Central: Cloud-based e-Science and its application to chemical property modelling. Technical report, Newcastle University, 2010.