Variant marketplace: Facilitating the Publication of Variants from Molecular Genetics Laboratories

David Atlan[1], Owen Lancaster[2], Gudmundur A Thorisson[2], Adam Webb[2] and Anthony J Brookes[2]

1. Phenosystems SA, Belgium *contact@phenosystems.com*
2. University of Leicester, UK *ol8@leicester.ac.uk*

Abstract

Diagnostics laboratories assess DNA samples from many patients with various inherited disorders, and so produce a great wealth of data on the genetic basis of disease. Unfortunately, those data are not usually shared with others. To address this gross deficiency, we are constructing a system that will facilitate the automated transfer of diagnostic laboratory data to the wider community, via an Internet-based Café for Routine Genetic data Exchange (Café RouGE).

Diagnostic laboratories are not reluctant to release their data. Instead, the obstacles are merely practical: First, diagnostic laboratory personnel do not have time nor funding to manually submit data to Internet depositories such as Locus Specific databases (LSDBs). Second, diagnostic laboratories would receive no recognition or reward for releasing their data, giving them little incentive to even try.

The Café RouGE approach takes account of the real-world obstacles and the needs of diverse LSDBs (insights provided by GEN2PHEN: http://www.gen2phen.org).

The core concept involves:

a) Endowing data analysis tools used by diagnostic laboratories with a 'data submission' function that automatically pushes processed data onto the Internet

b) Producing a single Internet depot (café) to receive these data and make them available for download by diverse third parties.

The Gensearch® DNA sequence analysis package (PhenoSystems SA) will soon offer this submission functionality, and other tool providers are encouraged to do likewise. We predict that several hundred mutations per month will be released by users of the Gensearch® tool alone.

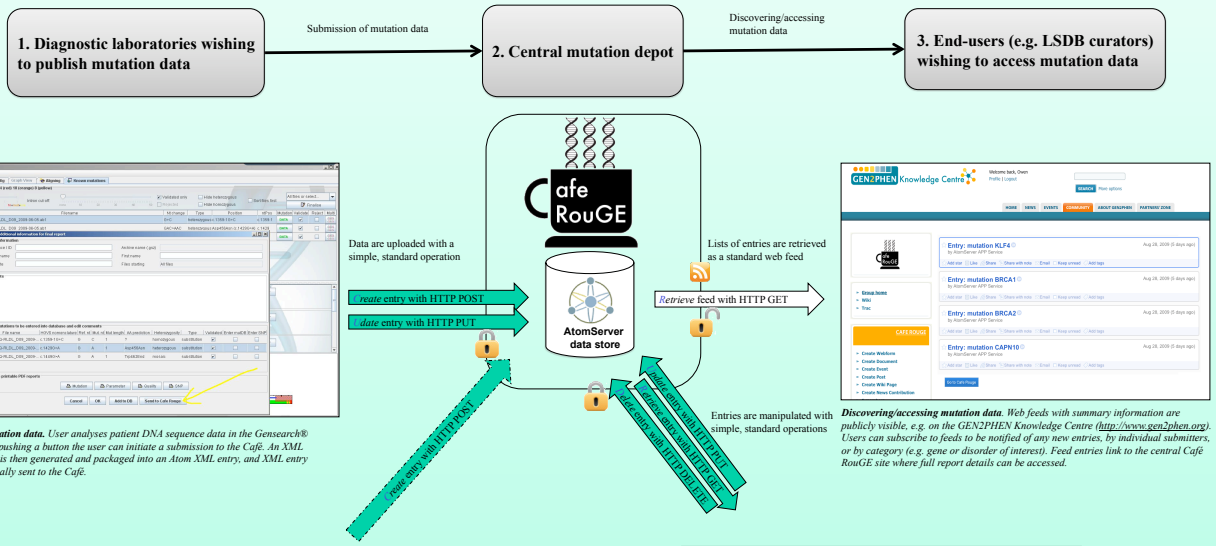
Café RouGE architecture – building on standard Web protocols and open-source components

✓A *web feed* is a simple XML document describing a collection of items, typically blog entries or news headlines. Feeds can be useful in any setting where a list of items is a useful representation. *Atom* is a popular, extensible feed format.

✓*AtomPub* is a simple, standard, RESTful web service protocol for publishing and editing information on the Web, using Atom XML as a *content-neutral data wrapper*.

✓An *Atom store* is a Web publishing platform based on the AtomPub protocol. Google Data APIs (GData) are examples of widely-used Atom stores.

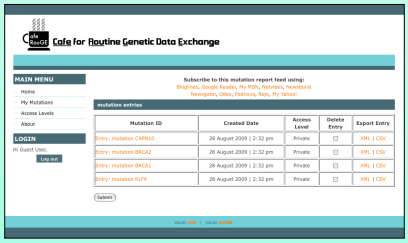
✓*AtomServer* is a generic, off-the-shelf open-source data store modelled after GData. Café RouGE is built around AtomServer.



Submitting mutation data. *User analyses patient DNA sequence data in the Gensearch® application. By pushing a button the user can initiate a submission to the Café. An XML mutation report is then generated and packaged into an Atom XML entry, and XML entry with payload finally sent to the Café.*

Discovering/accessing mutation data. *Web feeds with summary information are publicly visible, e.g. on the GEN2PHEN Knowledge Centre (http://www.gen2phen.org). Users can subscribe to feeds to be notified of any new entries, by individual submitters, or by category (e.g. gene or disorder of interest). Feed entries link to the central Café RouGE site where full report details can be accessed.*

The Café RouGE approach could be applied to distributing other types of biological data, e.g. genotypes



Managing mutation data. *Via the central Café RouGE web application submitters can edit or delete their submitted entries, configure access preferences (open vs controlled) and decline/approve access requests from end users.*
End users can access full report details which may require sending a request for access privileges to submitter.

Café security

- Mutation metadata are fully open-access (via Atom feeds)
- Register with OpenID or username/password to:
  - Request access to, and retrieve, controlled-access mutation data
  - Set up a submitter profile and upload/manage data

Future Developments

- Enable bundling of supporting primary data (e.g. chromatograms) with uploaded reports
- Identity-enabled attribution
  - Link datasets with submitter via unambiguous Contributor ID
  - Support fine-grained contribution statements in submissions
- Extend to other types of biological data where Café data distribution model is appropriate

Find out more

Development website: http://www.gen2phen.org/groups/cafe-rouge-development

Contact E-mail: caferouge@gen2phen.org

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 200754 – the GEN2PHEN project.