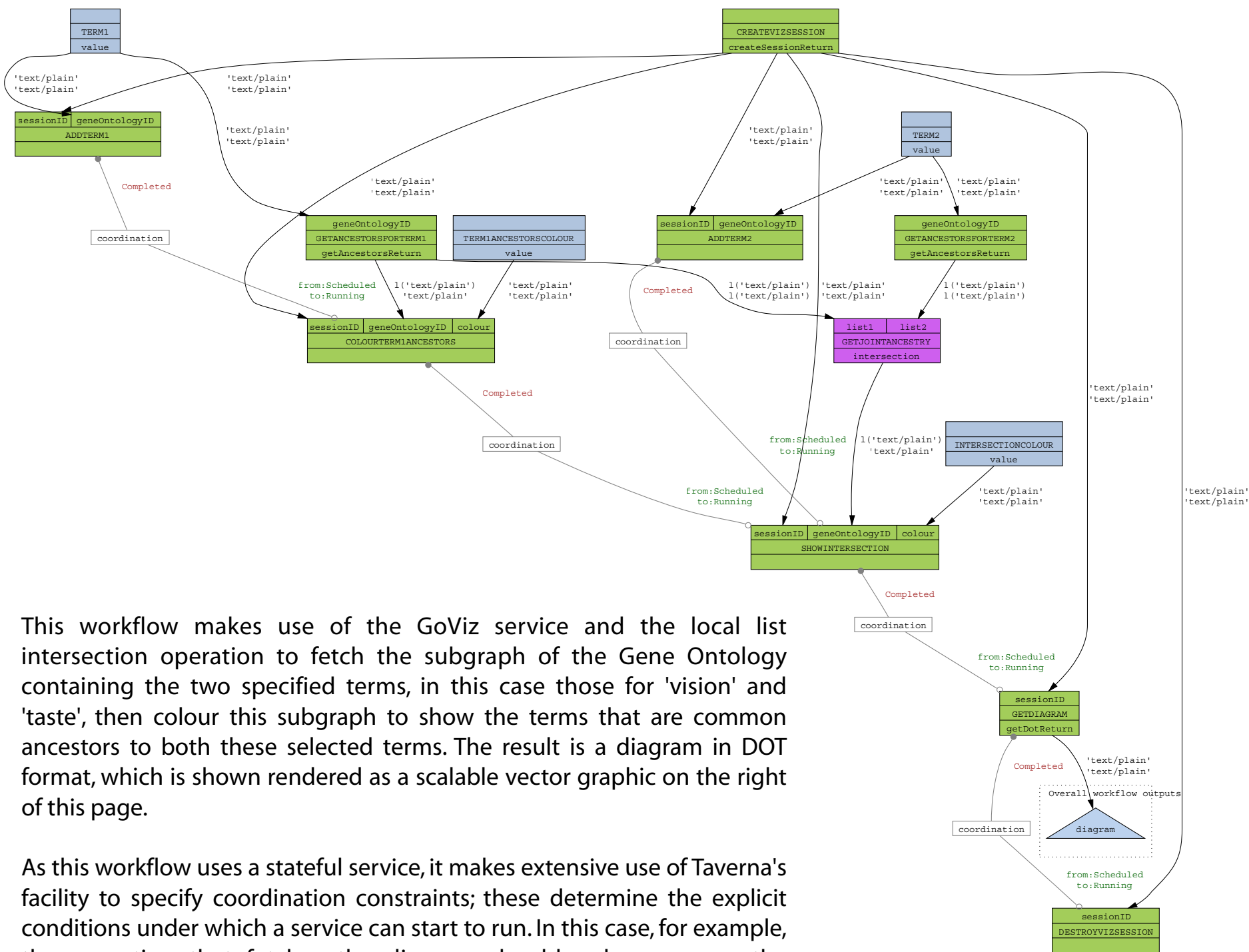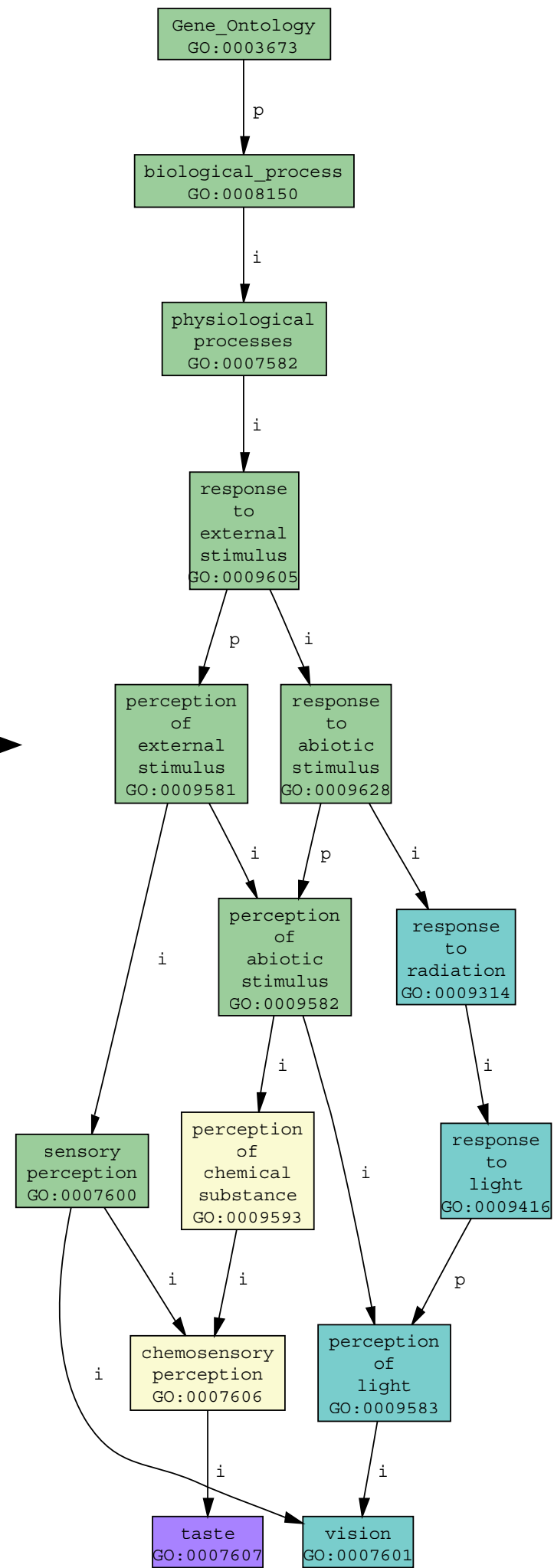# Taverna Beta7

Example Workflow showing Gene Ontology Visualisation

tmo@ebi.ac.uk

This workflow makes use of the GoViz service and the local list intersection operation to fetch the subgraph of the Gene Ontology containing the two specified terms, in this case those for 'vision' and 'taste', then colour this subgraph to show the terms that are common ancestors to both these selected terms. The result is a diagram in DOT format, which is shown rendered as a scalable vector graphic on the right of this page.

As this workflow uses a stateful service, it makes extensive use of Taverna's facility to specify coordination constraints; these determine the explicit conditions under which a service can start to run. In this case, for example, the operation that fetches the diagram should only run once the ancestors have been coloured correctly; there is no direct data link between these operations so the workflow designer must make this constraint explicit. Coordination constraints are shown in the workflow diagram as gray lines annotated with the exact gate condition required.
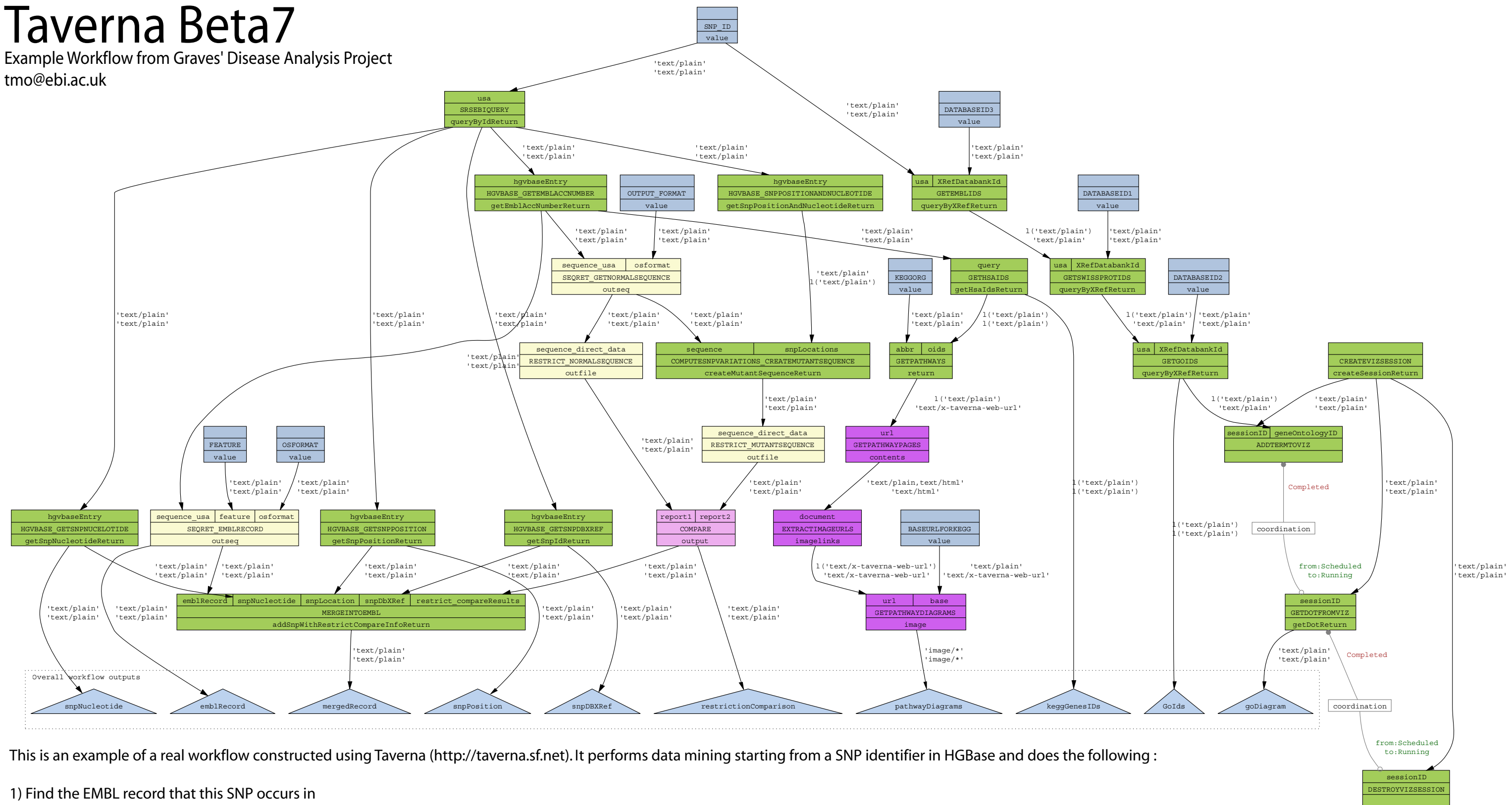
The GoViz service used here sits on top of our Oracle servers and the accompanying Java APIs, and is an example therefore of a facade layer on top of an existing system, the service was about four days work to write using facilities in Apache Axis for simple deployment.

Taverna Beta7

Example Workflow from Graves' Disease Analysis Project

tmo@ebi.ac.uk

This is an example of a real workflow constructed using Taverna (http://taverna.sf.net). It performs data mining starting from a SNP identifier in HGBase and does the following :

1) Find the EMBL record that this SNP occurs in
2) Use this record to fetch the sequence, and create the mutant sequence including the SNP
3) Run the RESTRICT program from the EMBOSS suite on both sequences and compare the results, adding new features to the EMBL record corresponding to any RFLPs predicted
4) If possible, cross reference to the KEGG GENES database from the EMBL identifier and fetch any pathway diagrams, annotated with the mapped gene.
5) Fetch any terms from the Gene Ontology (GO) cross referenced from this EMBL record and construct the subgraph of the GO containing these terms, building a graphical representation of this.

In order to perform these tasks, Taverna's workflow system accesses resources located at Kyoto in Japan (for KEGG), Newcastle University for some mapping services, the European Bioinformatics Institute for the GO service and EMBOSS via Martin Senger's Soaplab toolkit, and some simple local scripting operations running in the address space of the enactment engine. It shows how distributed compute and GRID technology can be used to perform real world scientific tasks in the life science domain; this workflow is part of a collaboration with the Newcastle Centre for Life, especially Clare Jennings and Simon Pierce, the workflow being constructed by Peter Li and Tom Oinn in consultation with these two biologists.