

Semantic Mashups with BioXMash



Patrick Ziegler¹ and Ela Hunt²

¹Department of Informatics, University of Zurich, Switzerland, pziegler@ifi.unizh.ch
²Department of Computer Science, ETH Zurich, Switzerland, hunt@inf.ethz.ch

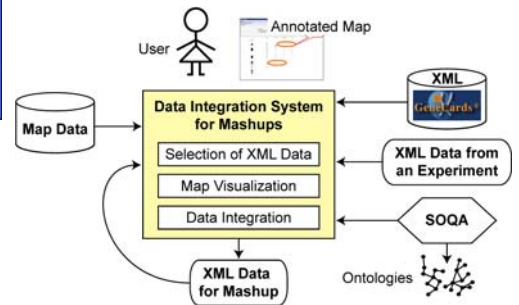


The image displays several software interfaces used in the BioXMash system. On the left is the SOQA Browser showing a GeneOntology hierarchy with categories like 'protein binding' and 'molecular adaptor activity'. In the center is the VisGenome browser showing a detailed view of the CXADR gene (HUMAN21) with its genomic location and associated functions. On the right is the SumGenome browser showing a genome map with various data points overlaid on chromosomes. A smaller window shows a search results list for 'CXADR' from various sources like Ensembl and GeneCards.

Map mashups combine content from multiple sources into a new workspace and visually superimpose various types of data on a map. Mashups provide a rich representation of integrated data gathered from several sources, which is useful in scientific data analysis. Genome maps for mashups can be served by an arbitrary source, such as Ensembl, and data can come from XML sources or files, for instance GeneCards. Mashups can be interactively constructed in several ways, by either adding new kinds of data to the display, adding textual annotations to the existing items, or adding markers or graphics. In this manner, visual data representations are enriched to support the researcher's understanding of data relationships.

Mashups gain additional power from showing data in the context of an existing ontology, such as GO. This allows one to semantically relate data from different sources and fosters correct data interpretation by both humans and computers. Ontologies, in particular, have the potential to add additional semantic perspectives to the complex picture of data relationships. Moreover, combining XML with ontological concepts provides access to semantics-enabled services, such as semantic similarity search [1]. In this respect, our approach extends previous work by utilizing ontological data annotations.

This demo presents a semantic mashup extension of BioXMash [2], an interactive XML data integration tool which collaborates with VisGenome [3] browser and with SOQA [4], a language-independent ontology API. BioXMash integrates data from XML sources, ontologies, and biological maps and shows several facets of the available data in one workspace. We demonstrate how a biologist produces mashups interactively, without programming, by selecting data from a large XML file repository, adding it to a genome map, and producing a mapping mashup presenting integrated data in map context. BioXMash offers textual search in XML paths, path statistics, and ontologies, and supports visual verification of data values. Our demo employs Ensembl maps, 2 GB of data from GeneCards, and local access to GO via SOQA. The SOQA Browser accesses ontologies via mashup maps and VisGenome, triggered by clicking on XML data annotations and GO concept names.



<i>VisGenome maps</i>	<i>SumGenome summaries</i>	<i>SOQA ontologies</i>
<i>data import</i>	<i>XML summarisation</i>	<i>ontology lookup (reasoning)</i>
<i>visualisation</i>	<i>representation</i>	
<i>manipulation</i>	<i>queries</i>	
<i>Data bus</i>		

[1] P. Ziegler et al. Generic Similarity Detection in Ontologies with the SOQA-SimPack Toolkit. SIGMOD 2006, 751-753.
 [2] E. Hunt et al. Defining Mapping Mashups with BioXMash. Journal of Integrative Bioinformatics 4(3), 2007.
 [3] J. Jakubowska et al. VisGenome: Visualization of Single and Comparative Genome Representations. Bioinformatics 23(19), 2641-2642, 2007.
 [4] P. Ziegler et al. The SIRUP Ontology Query API in Action. EDBT 2006, 1172-1175.

