Integration of heterogeneous data sources for high content screening data exploitation and exchange

Authors: Dmitri Voitsekhovitch, Elton Rexhepaj, Francois Prudhomme, Aurianne Lescure, Elodie Anthony, Sarah Tessier, Philippe Benaroch, Jacques Camonis, Franck Perez, Jean Salamero, Perrine Paul-Gilloteaux, Elaine Del Nery

High Content Screening produces large amounts of data. Images alone can go into the terabytes per month, and other data types must also be considered (e.g. lab experiment reports).

Existing solutions may not be deemed satisfactory either because they may be proprietary, making it difficult to extend its functionalities or because they define restrictive data models, the evolutions of which make component maintenance difficult.

As part of France Bio-Imaging network, we are developing the Real-time Unified Bio-Imaging Exploitation System. RUBIES uses open source tools and widely-known paradigms to create a collaborative platform to visualise and exchange HCS data. Heterogeneous data from HCS is analysed in real-time in order to enrich databases which are exposed through a web service layer.

RUBIES relies on the WebLab integration framework, which is based on Service Oriented Architecture where individual components are encapsulated as web services with common interfaces and communication protocols. It proposes a taxonomy of web service interfaces and a Resource
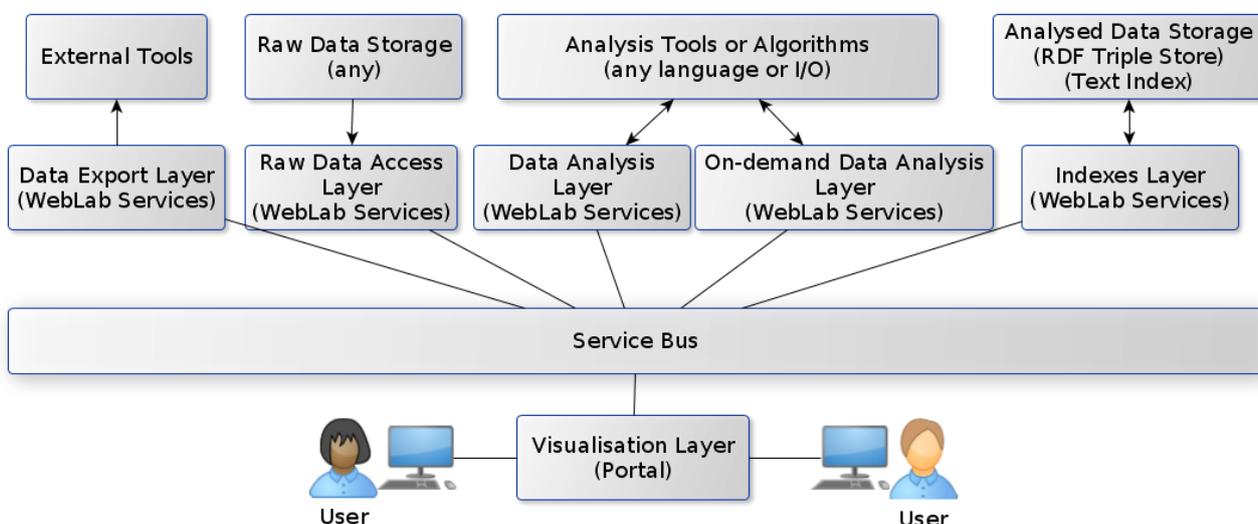
Description Framework (RDF)-based common data exchange format to ease service interoperability.

RDF is a web standard which represents any kind of information statements as collections of (subject, predicate, object) triples.

RUBIES can then interoperate with other data management systems, such as iManage Curie Data base.

Further work will include user-added annotations and exploitation of image and metadata in order to create cell image dictionaries.