Data Integration, Interoperability and Harmonization

Correlate and synergize otherwise disconnected data or information.

Keywords: data integration interoperability harmonization warehouse schema

Interoperability is the ability of making systems and organizations work together. Our service focuses on making data work together, regardless of what systems they originate from. Data integration involves combining data residing in different sources and providing users with a unified view of these data. This process becomes significant in a variety of situations, like those found in scientific endeavors where combining research results from different bioinformatics repositories are necessary.

Data interoperability can be approached through a variety of technical solutions. The two leading solutions are:

1. **Data warehouse** - Build a new database and load it with data from the original data sources (databases, spreadsheets, documents, etc.).
2. **Mediated schema** - A "virtual database" is assembled and data is referenced in real-time from original data sources (using wrappers, as necessary).

Examples

dkCOIN

| Use Case | The Beta Cell Biology Consortium Coordinating Center was responsible for harmonizing data from multiple NIDDK-funded consortia. |
| Data | Original data was stored in various relational databases (Oracle, MS SQL server, PostgreSQL and MySQL). Direct access to the data was not available. |
| Technology used | A mixed data warehouse/mediated schema was assembled, to store a thin subset of metadata common to all consortia. A SOAP server and clients were developed to allow each data user/provider to interact with the consolidated data and keep it updated. Third-party databases used to enrich data includes PubMed, NCBI Taxonomy and others. |
| URL | Inquire for private access |

Interactive Gene Networks

| Use Case | High-throughput functional genomics experiments (microarray, RNA-Seq, ChIP-Seq) have illuminated millions of gene-gene relationships and interactions. The BCBC warehoused all of this data, but needed an tool to allow scientists to interactively browse though this data in a localized manner. We constructed a local gene network visualization tool that met the needs of the scientists and bioinformaticians. |
| Data | Millions of records described entities (such as genes) and the relationships between these, including data provenance, type of relationship, directionality and more. |
| Technology used | Flash/Flare API, Javascript, HTML5, PHP, PostgreSQL |
Multi-Resolution Microscopy Image Viewer

**Use Case**
A high-content imaging project (Al Powers, Vanderbilt University) required the development of web tools to allow for the annotation and viewing of large-scale and 3D microscopy imagery.

**Data**
Microscopy data in the form of LSM, AFI and SVS files

**Technology used:**
Google Maps API (version 2 and 3), PostgreSQL, PHP, HTML5

**Service Fees**

Hourly rates apply.

Estimates to be provided following consultation.