

Daniel Bornman, David Rigsby, Manjula Kasoji, Jennifer Sawchuk, Steve Gomori, Pamela Hartford, and Steve Rust

Battelle Memorial Institute, Statistics and Information Analysis, 505 King Ave, Columbus, OH 43201

ABSTRACT

We are currently applying a systems biology approach for the investigation into the pathogenesis among various strains of Influenza. This approach is supported by an internal informatics platform that has been engineered to 1) manage and organize high-throughput data in a contextual way, 2) integrate correlated data from disparate technologies and data types, and 3) manage the multitude of analytical applications and platforms utilized within a research group. **We have developed an application that serves as a central "portal" providing access to multiple sources of internal and external research data and common bioinformatics software.** Research data is available to the user for analysis using integrated analytical applications identified as critical to the research and discovery pipeline. This research effort utilizes a number of omics technologies as well as traditional assays for elucidating the underlying biology of Influenza infection and requires stable methods for organizing data of this diversity and magnitude. At the core of the Battelle Informatics Portal is a **metadata-centric** database and file archive system. This metadata database provides context to each dataset and provides the research analyst necessary information on experimental conditions, design, and technologies. To ensure that sufficient and accurate information is maintained, we also enforced the use of **standard ontologies** for describing metadata attributes at the Study, Assay, and Sample levels. By implementing established standard nomenclature within the metadata system, the internal data management component of the Battelle Informatics Portal facilitates the **integration of multiple data types** during analysis. Once the necessary data is acquired, the Battelle Informatics Portal provides methods for uploading data into **integrated analytical bioinformatics software.** We are currently utilizing the integration of the Gaggle systems biology application network (The Institute for Systems Biology), Genespring Workgroup 11 (Agilent Technologies), and the Genedata Expressionist Suite (Genedata). The large multi-disciplinary, collaborative research program that we have utilized this application for has benefited in several ways by utilizing this application: Research data and analytical applications have been centralized for direct access to all stages of the pipeline. A strong **collaborative environment** has been fostered as staff with diverse roles (lab scientists, analysts, and project managers) interact with one another through active data exchange. This approach has significantly increased our ability to collaborate among our own geographically dispersed research staff as well as with our partners and service providers. Lastly, **standard analytical procedures** and standardized data elements are enforced across the system that increases efficiency and improves the overall accuracy and consistency of information.

TECHNICAL APPROACH

The Battelle Informatics Portal utilizes the **Netbeans Platform** as the underlying application architecture. A few of the advantages to utilizing the Netbeans platform include 1) a Java-standard application architecture, 2) **stability**, 3) a robust architecture, and 4) **modularity**. The concept of modularity is strictly enforced within the design to promote application **extensibility** overall and within each component module. The extensibility of the platform is critical to our ability to quickly and efficiently respond to changing requirements, to evolving scientific directions and technologies, and to our ability to respond to new business opportunities. Additionally, the current modular approach allows our bioinformatics core team to update features dynamically without redeploying the entire application. Modules may be installed individually by users with instant **plug and play** functionality; users may also choose to leave out certain modules that do not pertain to them or if a module requires a license purchase as in the case of our Genespring connection module.

BATTELLE INFORMATICS PORTAL

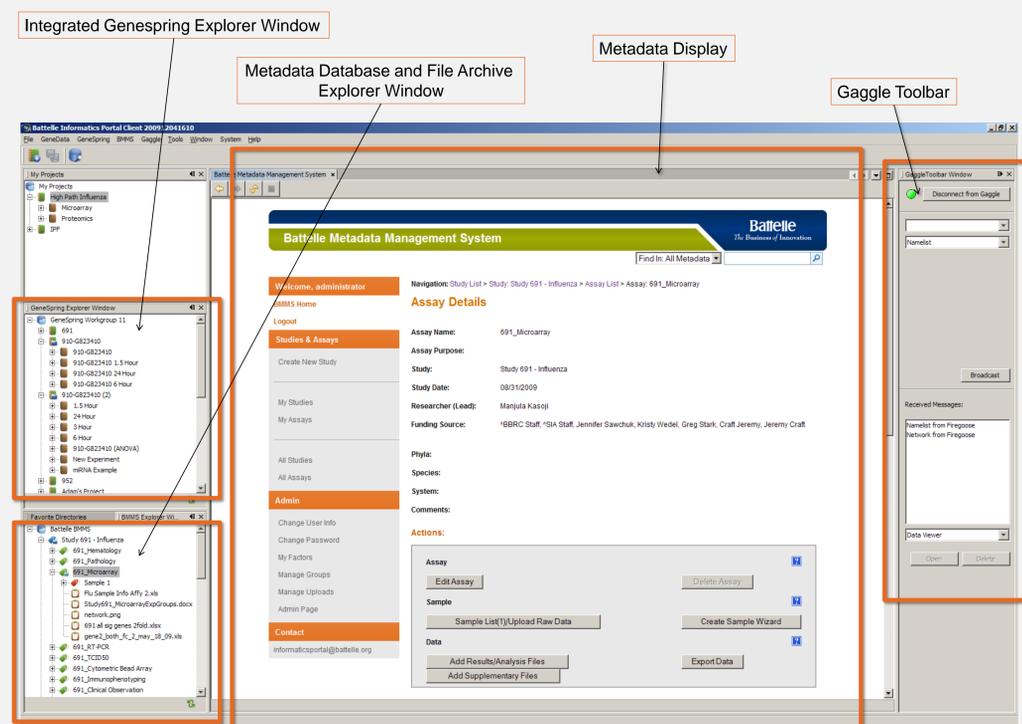


Figure 1: Screenshot of the Battelle Informatics Portal. This platform application provides the user with a dashboard view of the various integrated data sources and analytical applications that are available. The user may choose from a selection of explorer windows each giving a specific view into an integrated data source. The main window shows the metadata manager within the Battelle Informatics Portal and provides a means for users to either input metadata or query the database for projects and assays with matching attributes. The right-hand panel shows the custom Gaggle toolbar interface for broadcasting datasets within the Portal to various Gaggle applications for analysis.

INFORMATICS PIPELINE

As experimental data is produced for the current multi-disciplinary study, each data set is annotated for integration with other data types as it is moved through the informatics pipeline.

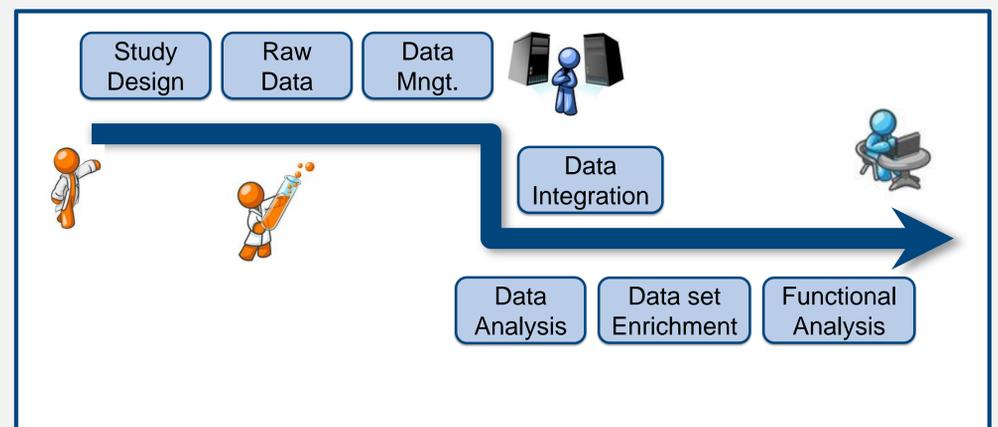


Figure 2: Schematic diagram of the major pipeline components supported by the Battelle Informatics Portal. Staff contributing to a study may access data at various points as it develops along the data pipeline.

STUDY DESIGN

Groups	Assay	6	12	24	72	96
	Pathology (lung, liver, spleen, brain, thymus)	X	X	X	X	X
	IP (lung)			X		X
	mRNA (lung)	X	X	X	X	X
	microRNA (lung)	X	X	X	X	X
1-4	Proteomics (lung)	X	X	X	X	X
	RT-PCR (lung)	X	X	X	X	X
	TCID50 (lung)	X	X	X	X	X
	CBA (lung)	X	X	X	X	X
	Guava (lung, spleen, thymus, bone marrow)	X	X	X	X	X

Table 1: Design description of the current Influenza challenge study. Each assay is being performed at each indicated time point following Influenza challenge. Groups 1-4 represent strains of Influenza: Group 1, A/Texas/36/91; Group 2, A/Vietnam/1203/04; Group 3, Cell culture vehicle control; Group 4, Egg culture vehicle control.