

# Life Sciences Data and Application Integration with B-Fabric

Can Türker, Fuat Akal, Ralph Schlapbach

Functional Genomics Center Zurich (FGCZ), UZH / ETH Zurich,  
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

<http://www.fgcz.ethz.ch>

[\(tuerker|akal|schlapbach\)@fgcz.ethz.ch](mailto:(tuerker|akal|schlapbach)@fgcz.ethz.ch)

## 1 Demonstration

In this demonstration, we sketch some central features of the B-Fabric life sciences data management system [1, 2]. As an example scenario, we use a scientist who is working on a plant named *Arabidopsis Thaliana* with the goal to figure out the effect of certain genes and the effect on light on it. For this purpose, he registers his samples and extracts with B-Fabric, loads his data into B-Fabric and defines his experiment. Afterwards, he runs his experiment and stores the results in B-Fabric. To complement this scenario, retrieval and administrative issues are being demonstrated as well. Main topics that will be presented in the demo are as follows:

**Register Samples/Extracts:** Users register their samples and extracts through intuitively designed forms. Data entering is facilitated by providing as many drop-down menus as possible to select annotations from the system vocabularies and by dynamically drawing forms according to selected annotation values. In addition, users typically register several samples and extracts where only a few attributes differ. In order to further ease the registration of them, cloning as well as batch registration of samples and extracts are supported.

**Annotation Management:** B-Fabric provides extensible vocabularies for the different annotations. If a user does not find a needed annotation in the corresponding drop-down list, the user can create a new one. All annotations created by users must be reviewed by an expert (in our case by an FGCZ employee). The expert checks the annotation and releases it if it is correct.

Annotation reviewing can be a tedious task due to similarly written versions of the same annotation. In such cases, B-Fabric automatically detects similar annotations and recommends merging them. B-Fabric allows similar annotations to be merged easily to maintain the annotation consistency system-wide. If the expert decides to merge two annotations, B-Fabric provides a form where he can easily select the attributes of the resulting merged annotation. When the two annotations are merged, B-Fabric automatically associates the samples which were previously associated with the misspelled annotation.

**Task Orientation:** B-Fabric is a task-oriented system that reminds its users about open tasks, awaiting to be performed next. For instance, as soon as a new annotation is added to the vocabulary, a new task to release this annotation appears in the task list of the corresponding expert.

**Data Import:** B-Fabric supports two ways to import data: 1) physically copying and 2) linking data files. To import data from a data source, a proper data provider must be configured. The B-Fabric deployment at FGCZ allows importing data files from local file systems as well as several instruments available at FGCZ. New data providers can be added to the system easily.

With the configuration of a data provider the selection of the data files in corresponding data stores can be restricted to the ones that are potentially relevant for the user. This is a crucial feature since the number of the data files can be huge. An import results in a workunit. A workunit thus represents a unit of logically related data files. B-Fabric implements the data import via workflows. The workflow-driven approach of B-Fabric is very useful in practice to reduce human mistakes and avoid skipping steps.

**Application Integration:** Integration of external functionality into B-Fabric is done via *application registration*. First, a connector is written for a certain type of application, e.g., for running *R* scripts on an *Rserve* system. Then, a small interface is defined to describe how the application gets its input. Finally, the scientist writes the application in any language. This on-the-fly coupling of external applications is a crucial feature of B-Fabric, which allows fast evolution of the system. Once an application is registered, an experiment can be created to run this application. Defining an experiment consists of selecting data resources, samples, extracts, and an arbitrary number of attributes (e.g. species and treatment in the example) that will be used as input for the application. The demo will show how easily a previously registered application (Two Group Analysis) can be invoked to conduct the desired experiment. Once the experiment is started, a corresponding workflow is initiated. The graphic presentation of the workflow is also used to show what is happening underneath in the system. B-Fabric supports arbitrarily complex workflows based on its underlying workflow engine. When the experiment is done, the scientist can view the experiment results. The results of the experiment is also presented to the user as a zip file so that they can easily be transferred to another medium.

**Full-text Search:** B-Fabric provides full-text search capabilities. A search may vary from certain attributes of certain objects to the content of readable attachments and data resources. The system provides quick search boxes on the main screen as well as more refined advanced search form. Searches done by the user are kept in the search history during his session and can be executed easily by selecting a search query from the search history. A query can also be saved for future reuse. A later invocation of such a saved query will of course include all objects satisfying the query at run-time. Another important feature of B-Fabric is that search results can be exported into files.

**Miscellaneous Functions:** In addition to all major aspects mentioned above, B-Fabric provides some additional functionality. Especially, B-Fabric supports a view on the main data objects in a networked fashion. Users can simply browse bidirectionally through all objects linked together. In addition, all data manipulation operations (create/update/delete) are logged in the system such that the user can remember what he did in the past and the system can be monitored. Last but not least, B-Fabric provides a bunch of administrative functions to manage objects, workflows, errors, and to maintain the system.

## References

- [1] Türker, C., Stolte, E., Joho, D., Schlapbach, R.: B-Fabric: A Data and Application Integration Framework for Life Sciences Research. In: Data Integration in the Life Sciences, DILS 2007, LNCS 4544, pp. 37–47. Springer-Verlag, 2007.
- [2] B-Fabric. <http://www.bfabric.org/>