

B-Fabric: The Swiss-Army Knife for Research Facilities

A core research facility such as the FGCZ has to cope with huge amounts of heterogeneous data produced by/for many users. Besides the management of the data itself, FGCZ must handle and keep track of the entire life cycles of userlab projects and service orders, which are the access-secured containers for the produced data.



The FGCZ solution to the core facility data management problem is called *B-Fabric*, an in-house developed information infrastructure that covers all aspects of a productive information management system, including data capturing and generation, data annotation and meta-data management, data search and browsing, transparent access and standard access protocols, communication and notification, task-orientation and monitoring, logging, reliability, performance, authentication, authorization, and many more issues required in the daily business of a researcher and a research facility. The design of *B-Fabric* is driven by the following observations:

- Flexible infrastructure: be ready to integrate any data and applications
- Annotated data: desired, essential but not easy to get
- Duplicate merge: avoid unnecessary communication problems
- Combined search/browse: provide easy access to interesting data
- Security: central issue due to high competition
- Data sharing: desired but in one direction
- Data quality: users are often not able to provide correct/expected data
- Data storage: do not store ALL ever produced data

Functionally, *B-Fabric* particularly tackles the following goals:

- Track Entire Project/Order Life Cycle
- Capture/Manage/Provide Data
- Allow Access-Controlled Data Sharing
- Manage/Track/Charge Resources
- Dynamically Add/Revise Services/Functionality
- Generate Statistics/Reports
- Reduce Time/Costs for Project/Order Application/Management
- Improve Service/Data Quality

The following list names the major *B-Fabric* features:

- Register/Manage User Accounts
- Submit/Review/Coach/Publish Projects
- Manage Project Members
- Import/Annotate (Research) Data
- One-click Access to My/All Accessible Data
- Browse Data Networks
- Export/Download Data
- Create/Run External Applications
- Manage Annotations
- Manage Services
- Submit/Process/Track Project Orders
- Charge/Invoice Projects/Orders
- Reserve/Charge Instruments
- Manage/Control (User Access) Roles
- Synchronize with Computer Access Rights
- Log/Browse Emails
- Detect/Merge Duplicates (Users, Institutes, Services etc.)
- Detect/Eliminate Orphans
- Send Messages to Selected Users
- Request Access to Lab

In the following, we sketch a few of them.

Data Import and Annotation. Research data is captured and annotated together with its semantic context in form of workunits. A workunit represents a unit of logically related data together with its meta-data. The data import and annotation workflow consists of the following steps:

1. Register samples and extracts, i.e., provide the meta-data about the biological sources used for creating the research data
2. Select data import application, i.e., choose the data basis from where you want import the research data
3. Select resources for import, i.e., decide which resources are related together and thus form a workunit
4. Assign extracts to workunit resources, i.e., annotate the workunit by providing the missing link between workunit resources and their corresponding extracts
5. Describe resulting workunit, i.e., provide some details on the workunit as basis for later use

Register Samples/Extracts. Users register their samples and extracts using intuitively designed forms. Data entering is facilitated by providing as much drop-down menus as possible to select annotations from the extensible system vocabularies.

Extract : 14894 - 20121114_Hela_PhosphoEnrichment

Details	ID	14894
Comments	Name	20121114_Hela_PhosphoEnrichment
Workunits	Project	781
Resources	Parent Sample	9610 - 20120627_Hela
	Type	Proteomics
	Separation Techniques	C18 ZipTip
		TI02
		SCX
	Enzymes	Chymotrypsin
		Trypsin
	Chemical Modifications	Carbamidomethyl (C) ; Iodoacetamide
	Description	Phospho Enrichment of about 900ug proteins
	Created By	Dr. Paolo Nanni (paolo) 2012-11-14 09:37
	Modified By	Dr. Paolo Nanni (paolo) 2012-11-14 09:37

Typically, users 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.

Import Research Data. From a user's point of view, importing data to *B-Fabric* is quite easy. The user runs the corresponding import application which in return provides a list of all resources that can be imported.

Applications

Run	Name	Type	Technology	Description	Supervisor	Annotation Required
Run	QTRAP_1 Import	Import	Proteomics	Import QSTAR_1 wiff...	Dr. Christian Panse	true
Run	ORBI_2 XL Import	Import	Proteomics	Import proteomics f...	B-Fabric Administrator	true
Run	Local File Import	Import	General	Import files from yo...	B-Fabric Administrator	true
Run	Local File Import (no annotations)	Import	General	This is an applica...	Dr. Can Türker	false

From this list, the user selects the resources he wants to import as a workunit. A workunit can reflect the result of an analysis, a search result, or whatever the user wants to store as a logically grouped unit.

Run QTRAP_1 Import : Select Resources

Available Resources (31)	Relative Path	File Size	File Date
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_047_schedMRM_Mut_inf_3h_C.mzML	4.611 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_046_schedMRM_VT_inf_3h_C.mzML	4.612 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_044_schedMRM_Mut_inf_2h_C.mzML	4.608 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_043_schedMRM_VT_inf_2h_C.mzML	4.600 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_040_schedMRM_Mut_inf_1h_C.mzML	4.612 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_039_schedMRM_VT_inf_1h_C.mzML	4.610 MB	2014-03-06 11:24
<input checked="" type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_037_schedMRM_Mut_inf_3h_C.mzML	4.610 MB	2014-03-06 11:24
<input checked="" type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_036_schedMRM_VT_inf_3h_C.mzML	4.608 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_033_schedMRM_Mut_inf_3h_B.mzML	4.606 MB	2014-03-06 11:24
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_032_schedMRM_VT_inf_3h_B.mzML	4.608 MB	2014-03-06 11:24

Annotate Research Data. Research data is linked to biological meta-data by assigning extracts to resources. This step is absolutely essential for the later research data reuse. Without this connection the interpretation of the resource content is hardly possible afterwards.

Run QTRAP_1 Import : Assign Extracts

Selected Resources (2)	Relative Path	File Size	File Date	Extracts	Propose	Clear
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_037_schedMRM_Mut_inf_3h_C.mzML	4.610 MB	2014-03-06 11:24	15100 - 20121118_Hela_Pho	Select Item	Select Item
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_036_schedMRM_VT_inf_3h_C.mzML	4.608 MB	2014-03-06 11:24	15099 - 20121118_Hela_Pho	Select Item	Select Item

Assigning extracts to resources also comes with some intelligence. The user gets proposed best matches between data resources and extract names. Often, the user just needs to press the save button and continue with the description of the resulting workunit.

Run QTRAP_1 Import : Create Resulting Workunit

Name : QTRAP_1 Import

Project : 781 - BallBEX Proteomics - Manipulating the fight between human host cells an

Description

Save Edit Resource Selection Cancel

Selected Resources (2)	Relative Path	File Size	Extracts
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_037_schedMRM_Mut_inf_3h_C.mzML	4.610 MB	15100 - 20121118_Hela_PhosphoEnrichment_Washing
<input type="checkbox"/>	p781ProteomicsQTRAP_1f1asa_20110923mdml_pr_1f1asa_20110923_036_schedMRM_VT_inf_3h_C.mzML	4.608 MB	15099 - 20121118_Hela_PhosphoEnrichment_FlowThrough

B-Fabric allows importing data files from local file systems as well as several instruments available. New data import applications can be added to the system easily. With the configuration of a data import application the selection of the data files in the corresponding data storages can be restricted to those ones that are potentially relevant for the user. This is a crucial feature since the number of the data files can be huge.

Annotation Management. *B-Fabric* provides extensible vocabularies for different annotations. If a user does not find a needed annotation in the corresponding drop-down list, the user can create a new one easily. All annotations can be reviewed and released by an expert. Annotation reviewing can be tedious due to similarly written versions of the same annotation. In such cases, *B-Fabric* automatically detects similar annotations and recommends merging them. If asked so, similar annotations can be merged easily to maintain the annotation consistency system-wide.

Application Integration. *B-Fabric* can easily be coupled with external functionality. It is possible to register external applications written in any language. The applications can perform data import, analysis, or whatever the user wants to realize. The on-the-fly registration, wrapping, and running of external applications together with the feedback and linkage of

the results are crucial features of *B-Fabric*. The resulting dynamic evolution of the system adds a tremendous value to the users since they can straight-forwardly run new or improved applications on all accessible data.

Edit Application : 10 - ORBI_1 Import

Name : ORBI_1 Import

Supervisor : Barkow-Desterreicher, Simon

Type : Import

Pageflow : Import

Annotation Required :

Import Resources Required :

Technology : Proteomics

Help :

Description : Import proteomics files from the Orbitrap 1 instrument

Hidden :

For Employees Only :

Succeeding Applications

Add	Remove Selected	Application Id	Name	Type	Technology	Description
<input type="checkbox"/>		64	Download via Torrent	No Filter	General	This application gen...
<input type="checkbox"/>		88	Exclusion list	Analysis	Proteomics	Given reproducibile c...
<input type="checkbox"/>		168	MaxQuant_Scaffold_LFO_fast	Analysis	Proteomics	Input is either a co...
<input type="checkbox"/>		164	MaxQuant_Scaffold_LFO_slow	Analysis	Proteomics	Input is either a co...

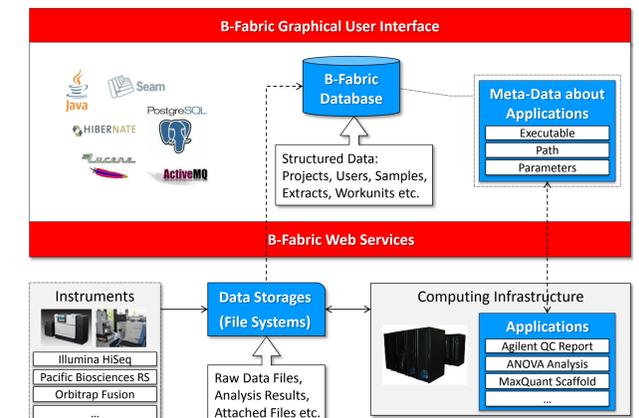
Total: 4 Rows XLS CSV

Data Search. *B-Fabric* forms a huge browsable data network with personal views for each user depending on the user's access rights. Above that, *B-Fabric* provides full-text search capabilities. A search may vary from certain attributes of specific entities to the content of readable attachments and data resources. The system provides quick search boxes on the main screen as well as advanced search forms. Search queries can be saved for future reuse or sharing with other users. A later invocation of such a saved query will of course include all the entities satisfying the query at run-time. Another important feature of *B-Fabric* is that search results can be exported into files.

Project Management. *B-Fabric* supports the entire life cycle of a research project from its application, reviewing and coaching towards data production and publishing. *B-Fabric* could also be used by any professor, even outside the life sciences, for instance, to manage PhD projects. All primary and secondary data produced by a PhD student could completely be captured, documented, and archived. As a desired consequence, transparency and reusability of research results would be achieved with no additional efforts for the professor. Besides, *B-Fabric* provides an agenda feature which allows to easily handle employee absences and generates signature-ready vacation reports.

Order Management. *B-Fabric* above that supports the entire life cycle of a service order starting with its submission, processing, data provision, charging, invoicing, and ending with the generation of ready-to-send bills. Using the comment feature of *B-fabric*, the entire communication between users and/or FGCZ staff is tracked such that absences of FGCZ staff members must not automatically lead to delays since somebody else can take over the order processing. Note that comments can also be used to share information and documents within an order and/or a project.

Technical Background. *B-Fabric* is implemented based on a synthesis of a number of open source technologies. The implementation is done on Java EE and the Seam Framework. The application itself runs in a GlassFish application server. The communication between the components of *B-Fabric* relies on ActiveMQ. All structured data is stored in a PostgreSQL database. The object-relational mapping between the Java application objects and the database tables is done using Hibernate. Quick and advanced data search over *B-Fabric* data is based on Apache Lucene. Besides the Web-based GUI, *B-Fabric* provides a Web Services interface to the major entities of the system.



Final Remarks. *B-Fabric* is running in daily business at FGCZ since 2007. Here are some figures as of December 2014:

Users	3480	Institutes	406	Samples	19305	Workunits	85535
Projects	1683	Orders	5892	Extracts	24265	Data Resources	163109

Without *B-Fabric*, such a number of projects could not be handled at the service quality of FGCZ. To conclude, besides all its obvious research data management benefits, *B-Fabric* increases transparency while reducing significantly the work of all involved people (scientists, secretaries, lab heads, and IT admins). Last but not least, on the one hand it provides the basis for future reuse and validation of research data and on the other hand it is flexible for performing the continuously required adaptation of the inherently evolving and unforeseeable requirements of the research field.

Contact: Dr. Can Türker, tuerker@fgcz.ethz.ch