

**Generalizing B-Fabric towards an Infrastructure for
Collaborative Research in Switzerland**

Deliverable No D2:

***Coupling of Some Selected EAWAG
Resources with B-Fabric***

January 2010

Document Information

Project	
Project Title	Generalizing B-Fabric towards an Infrastructure for Collaborative Research in Switzerland
Project Start	01.06.2009
Project Sponsor	SWITCH (AAA/SWITCH Project)
Project Number	UZH.5

Document	
Title	Coupling of Some Selected EAWAG Resources with B-Fabric
Date	
Author(s)	Fuat Akal, Can Türker
Total Number of Pages	10
File Name	D2.docx
Key words	Ad-hoc coupling, external data resource, data provider, workflow, application

Table of Contents

Generalizing B-Fabric towards an Infrastructure for Collaborative Research in Switzerland.....	1
1 Architecture for Coupling an EAWAG Data Resource with B-Fabric	5
2 Concrete Steps of the Coupling.....	6
3 Conclusion	10

Summary

B-Fabric allows ad-hoc coupling of external data resources. The detailed design for coupling external data resources with B-Fabric was given in the deliverable D1. This document explains how a selected EAWAG data resource was coupled with B-Fabric.

1 Architecture for Coupling an EAWAG Data Resource with B-Fabric

The architecture for realizing the coupling EAWAG data resources is depicted in Figure 1. At EAWAG, access to a selected data resource is granted via an SSH server. At FGCZ, an instance of B-Fabric is running which is used to perform the coupling and allow access to the coupled external data. Note that the data after the coupling still resides at the EAWAG data store.

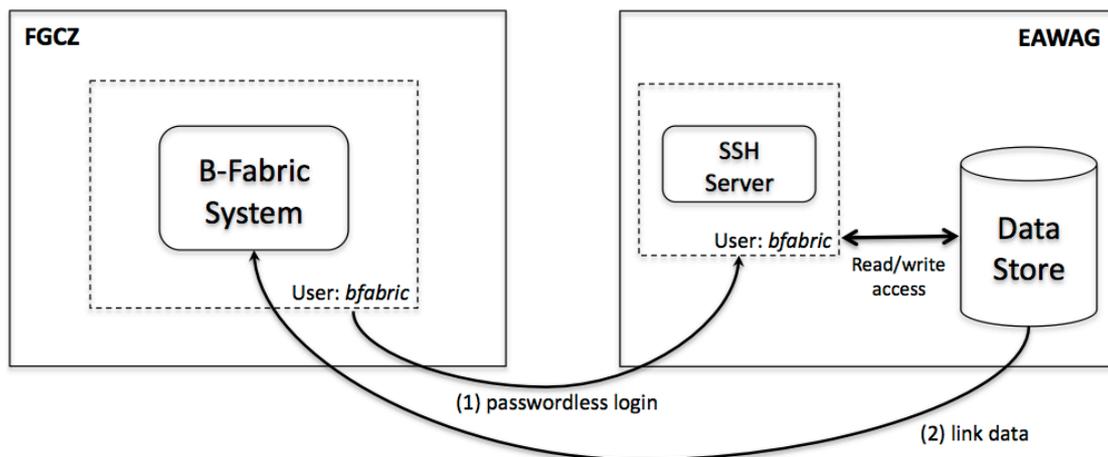


Figure 1: Architecture for Coupling EAWAG Resources

The coupling of the external data resource is realized in two steps.

First, B-Fabric accesses the data provider on EAWAG and is granted access to data. The data provider provides an interface between the data store and B-Fabric (please see the deliverable D1 for details). In order to couple a new data store with B-Fabric, the system deployer has to choose the proper data provider for the data store and configure it. As the data provider, an *SSH Server* is chosen. EAWAG provides an SSH server which allows the user *bfabric* of FGCZ to login without a password. This user has read/write access to the data to be coupled with B-Fabric.

Second, the system deployer chooses an appropriate workflow, which handles the accessing and processing of the import data. Note that external data resources can be coupled with B-Fabric via its application registration concept. Each registered application is associated within an internal B-Fabric workflow. This workflow first collects the input parameters for the invocation of the external application (if any), then invokes the application with these parameter, and finally stores the returning result of the application. Currently, B-Fabric supports different types of data import workflows. In case of the coupling of the EAWAG data, a workflow

based on linking the external data to B-Fabric is chosen since the data is already in a safe repository and the owner of data desired its data to be physically accessed only when required to save its internet bandwidth. With this linkage, the virtually imported external data is transparently accessible via B-Fabric.

2 Concrete Steps of the Coupling

First, a data import application must be registered in B-Fabric to allow an import of the selected EAWAG data resource. Figure 2 shows the corresponding screen of B-Fabric. Using the "Create/Edit Application" screen, a user provides the properties of such a data import application. In the current version of B-Fabric according to the FGCZ policies, only eligible users, i.e. FGCZ employees, can register applications. In a further generalization step of B-Fabric, this policy could also be changed in favor of a more flexible policy allowing an ad-hoc granting the right for registering an application to a given user depending on certain restrictions.

In Figure 2, an employee creates an application named *eawag_import* and specifies its properties. Since this application shall represent a data import application, the type of the application, i.e., the associated B-Fabric workflow, must be set to "import". B-Fabric support several types of import workflows. In the EAWAG scenario, the application uses the *generallinkimport* workflow, i.e., the imported data will physically stay on the coupled external data resource while B-Fabric maintains links and thereby transparent access to that data.

The external data resource is hosted at *uranus.emp-eaw.ch* and located under */export/bfabric/lc_ms/Orbitrap/Utox/Marc/Data/* directory. This directory will recursively be accessed by using *SCP*, which is an SSH-based copying program. Out of all files residing under the specified directory, only the files with the extension "RAW" (case insensitive) will be provided and accessed, respectively, via this registered data import application of B-Fabric.

Edit Application

Name • ?

Technologies

- Sequencing
- Metabolomics
- Genomics
- Proteomics
- Transcriptomics

Hidden ?

Type • ?

Workflow • ?

Protocol • ?

Host • ?

Base Path • ?

Project Separated ?

Path Extension ?

Recursive Listing ?

List directories ?

Directories Recursive ?

Include Pattern ?

Exclude Pattern ?

Description ?

SAVE

 CANCEL

Figure 2: Registration of an EAWAG Data Import Application

Once this data import application is registered with B-Fabric, it can be used to import data from the directory specified by the host and base path. The result of such a data import application is always a workunit. A workunit is the B-Fabric concept for representing a set of logically related files.

With the registration of the data import application, B-Fabric dynamically extends the list of available applications on the "Import Data" screen (see Figure 4). As the EAWAG data import application was not restricted to a specific technology (Sequencing, Metabolomics, Genomics, Proteomics, Transcriptomics), the link to the corresponding data import screen appear under "General".

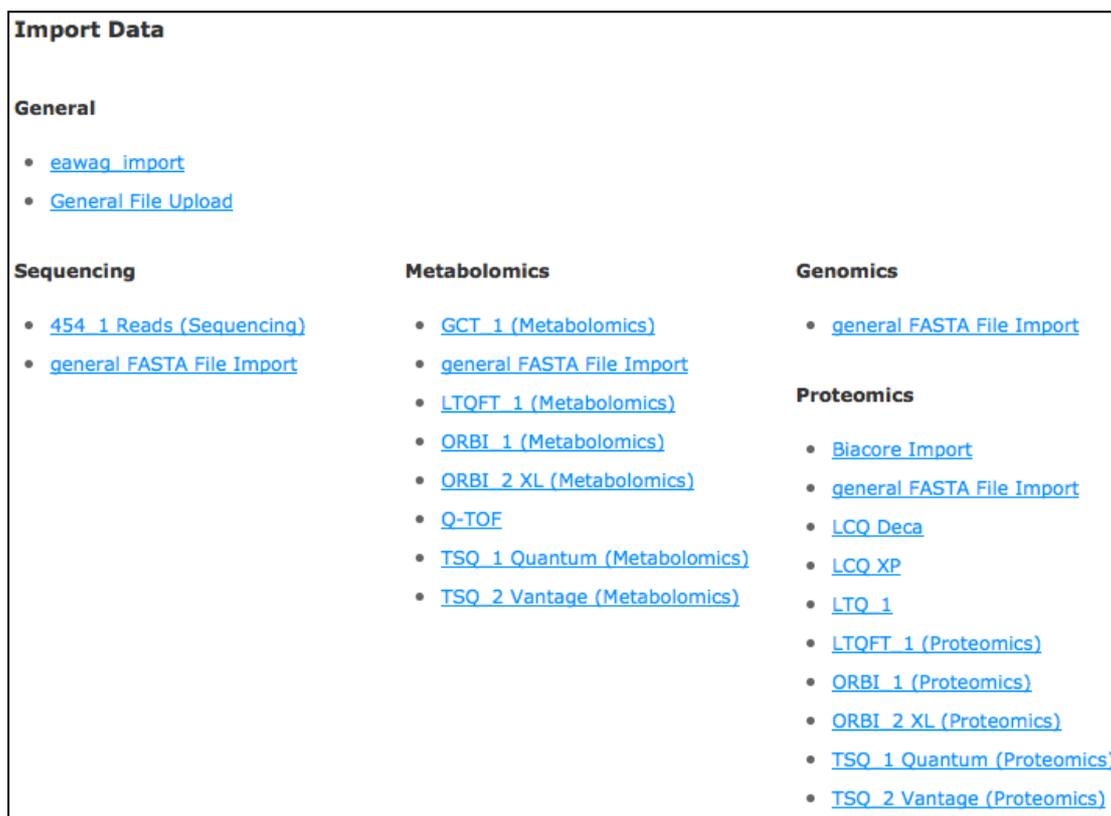


Figure 3 Data Import Screen with List of Available Applications

After clicking on the link “eawag_import”, the user will be directed to the screen depicted in Figure 4. This “Create Workunit” screen is used to specify the parameters for the execution of the EAWAG data.

By using the information of the registration of the application, B-Fabric accesses the data that comply with the access pattern (RAW files) in the application registration and provides a list of all files satisfying the specified conditions. Using the project association, this list is restricted to files that are created within the selected project. This restriction to the potentially relevant data import files is very useful feature of B-Fabric that can save the user a lot of time, especially in case of large projects where ten thousands of files are generated. From this list of files, the user then selects the files he wants to import together and gives a meaningful name for the resulting workunit. In the example, the user chooses three files and creates a workunit named as *eawag_workunit*.

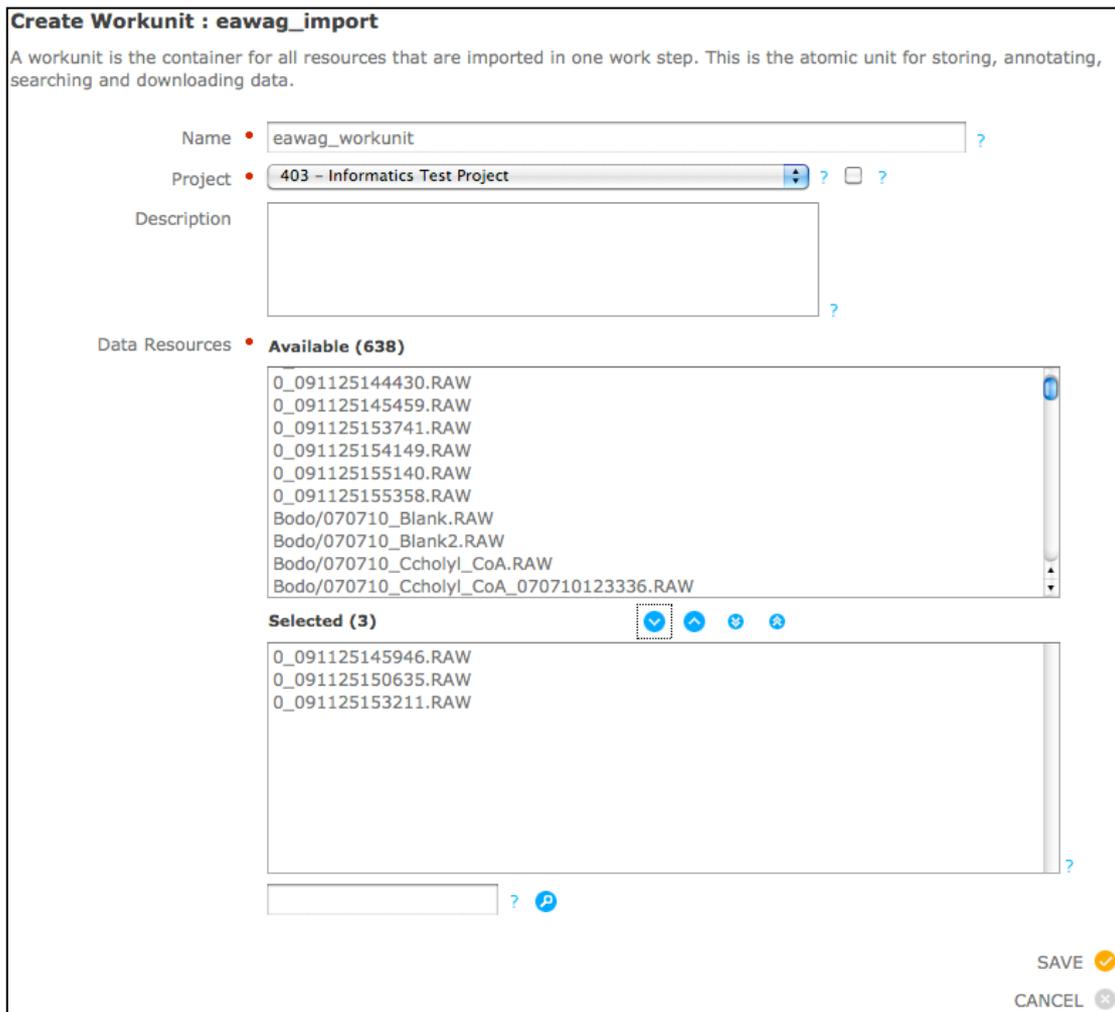


Figure 4: Execute EAWAG Data Import Application

When the workunit is saved, the screen in Figure 5 appears. This screen basically shows that 1) the data of EAWAG is successfully imported into B-Fabric and 2) B-Fabric is waiting for the user to perform the next step of the workflow underlying the EAWAG data import application. In this example, the next step is to assign the imported RAW files to previously defined extracts (descriptions of the biological sources) in B-Fabric.

Going beyond this point in the example is beyond the scope of this document. Note that the location of data is completely transparent to its users, because imported data, either physically or logically, are used in the same way by B-Fabric.



Figure 5: Link External Data to Biological Information

3 Conclusion

With this delivery we have proven the principle feasibility of B-Fabric’s ad-hoc coupling capability. Using the generic feature of registered applications, B-Fabric allows the transparent provision of application functionality to users. This provision requires no programming or rebuilding efforts on B-Fabric. It requires only logging on to B-Fabric and filling a web form to create an application definition. Of course, the registering user must carefully enter the application properties such that the registered application works smoothly. With this delivery we have demonstrated the usefulness of this approach by showing how a concrete data source can be coupled with B-Fabric.