The R hackathon

The session started on Friday July 10, at 11:30

Attendees: Javad Chamanara,Claas-Thido Pfaff , Pawandeep Kaur, Nafiseh Navabpour, Viktor Senderov.

The challenge is to retrieve a designated dataset from a BExIS instance and use in an R script for further analysis.

BExIS is running as a web application under IIS on Windows servers, is developed using C# and does not allow direct access to its database. Also, it is an authentication and authorization mechanism that protects not only the functionality provided by the system, but also the access to individual datasets. It is also remarkable to mention that datasets in BExIS get versioned upon any changes, so it is possible to request a specific version.

In order to tackle the challenge and provide a hack style solution, the team decided to reduce the scope to a simple scenario, in that an R package is developed. The R package has a single function that accepts a dataset ID and requests the latest version of that dataset from a designated BExIS instance. All authentication, authorization and versioning are bypassed.

On the BExIS side the team provided an easy RESTfull API that by accepting a dataset ID, returns the dataset (latest version) in CSV format. The R package then was able to call the API using the curl package.

The curl package was used to issue an HTTP GET request on the exposed API, so the result was a CSV file created on the fly. The result of the curl request then was fed to the R’s csv.read function which resulted in a usual data frame, ready for further processing.

Having the basic scenario up and running, the team created an R package named rBExIS and published it on github (accessible on: <https://github.com/cpfaff/rBExIS>).

After publishing and polishing the package, the team decided to capture more detailed requirements for the package. The result was to have functions to get a specific version, a minimalistic metadata, the full metadata and a subset of the data structure of the dataset requested.

The functions and the API are decided to be stateless, which means no information including credentials is preserved between the calls. For these purpose the functions take the URL and the credentials as parameters but in order to reduce the amount of parameters needed to be passed to the functions, the R package has the BExIS instance’s base URL as well as user credentials as options, hence, the package user does not need to enter them for each function call.

By now requirements are documented, some functions are available in the package, and the API in place, but all need a serious attendance to make the whole integration work in a useful way.