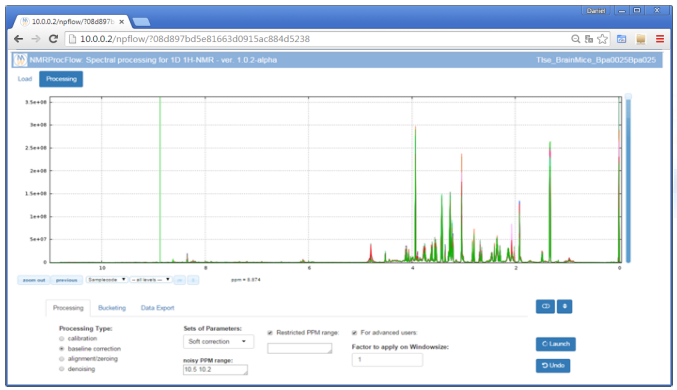
NMRProcFlow



An efficient GUI tool
for spectra processing
from 1D 1H-NMR metabolomics data

http://nmrprocflow.org/

PhenoMeNal
Large-Scale Computing for Medical Metabolomics

Galaxy

http://phenomenal-h2020.eu/home/

combined with the computing power of
a Cloud Research Environment

Daniel Jacob

Two major metabolomics approaches

Metabolic Fingerprinting

The complex data are directly and initially used for global multivariate statistical analysis.

Subsequently, **metabolite features that distinguish sample classes are identified** and then the structures of distinguishing metabolic features are established

Targeted Metabolomics

Quantitative approach wherein **a set of known metabolites are quantitated**.

The identities of metabolites were initially established based on the available databases and using standard compounds.

The identified metabolite peaks are then quantified based on internal or external reference compounds.

Standard Operating Procedures (SOP)

Metabolic Fingerprinting

Harvest / Storage

Sample Preparation

Analytical Sample Preparation

Analytical

Targeted Metabolomics

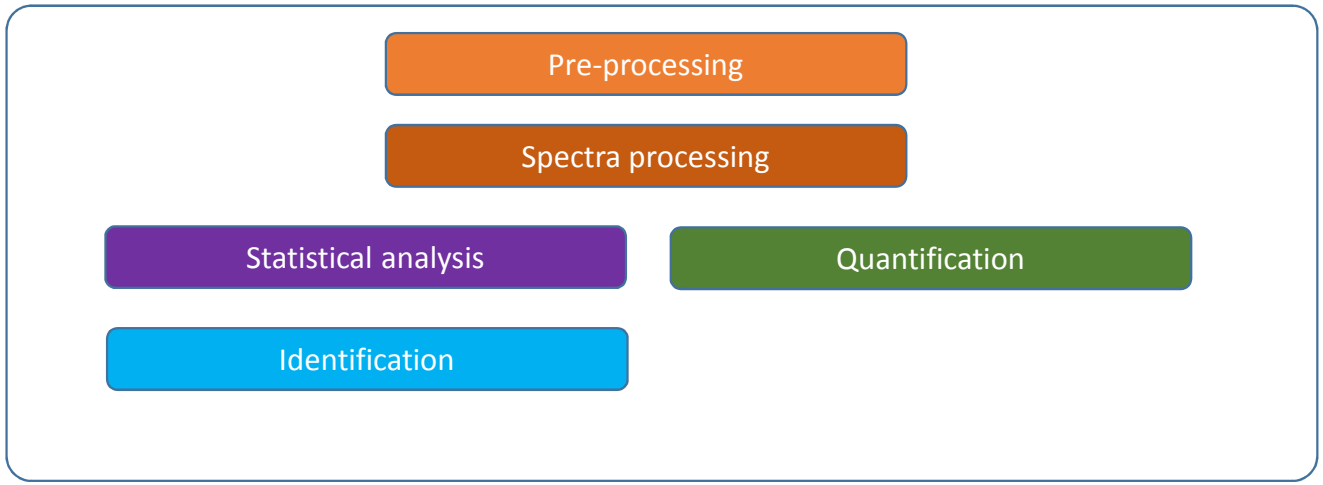
Efficiency,
Repeatability

⇒

Need of
Traceability

⇒

Evidence,
Confidence



Science ⇒ Data ⇒ Experiments ⇒ Good practices

Galaxy – Weakness

"Standard workflow"

Given the nature of the 1D NMR spectra and due to the diversity of problems encountered during the various stages of processing:

- *baseline correction,*
- *calibration ppm,*
- *removal of solvents and other contaminants,*
- *re-alignment of areas having high variations in chemical shifts between spectra, ...*

and depending on:

- *the biological context (humans, plants, micro-organisms),*
- *the type of tissue (plasma, urine, extract of fruit flesh ...),*
- *the analytical protocol (choice of the NMR sequence, additives for calibration and / or quantification, the buffer solution to stabilize the pH in the tube, etc ...).*

It is essential to process this type of data, with an interactive interface that enables the spectra visualization.

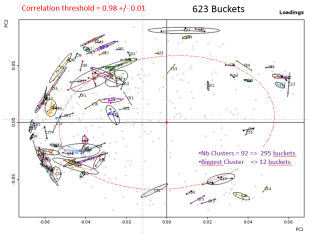
Galaxy – Recommendation

Confidence ? Only if **1** workflow <=> **1** use case => Upstream SOP

We consider the implementation of NMR spectra processing workflows executed in *batch mode* as relevant provided that we want to process in this way very well-mastered and very reproducible use cases, i.e. by applying the same Standard Operating Procedures (SOP).

Galaxy – Strength

Raw Data
1D NMR

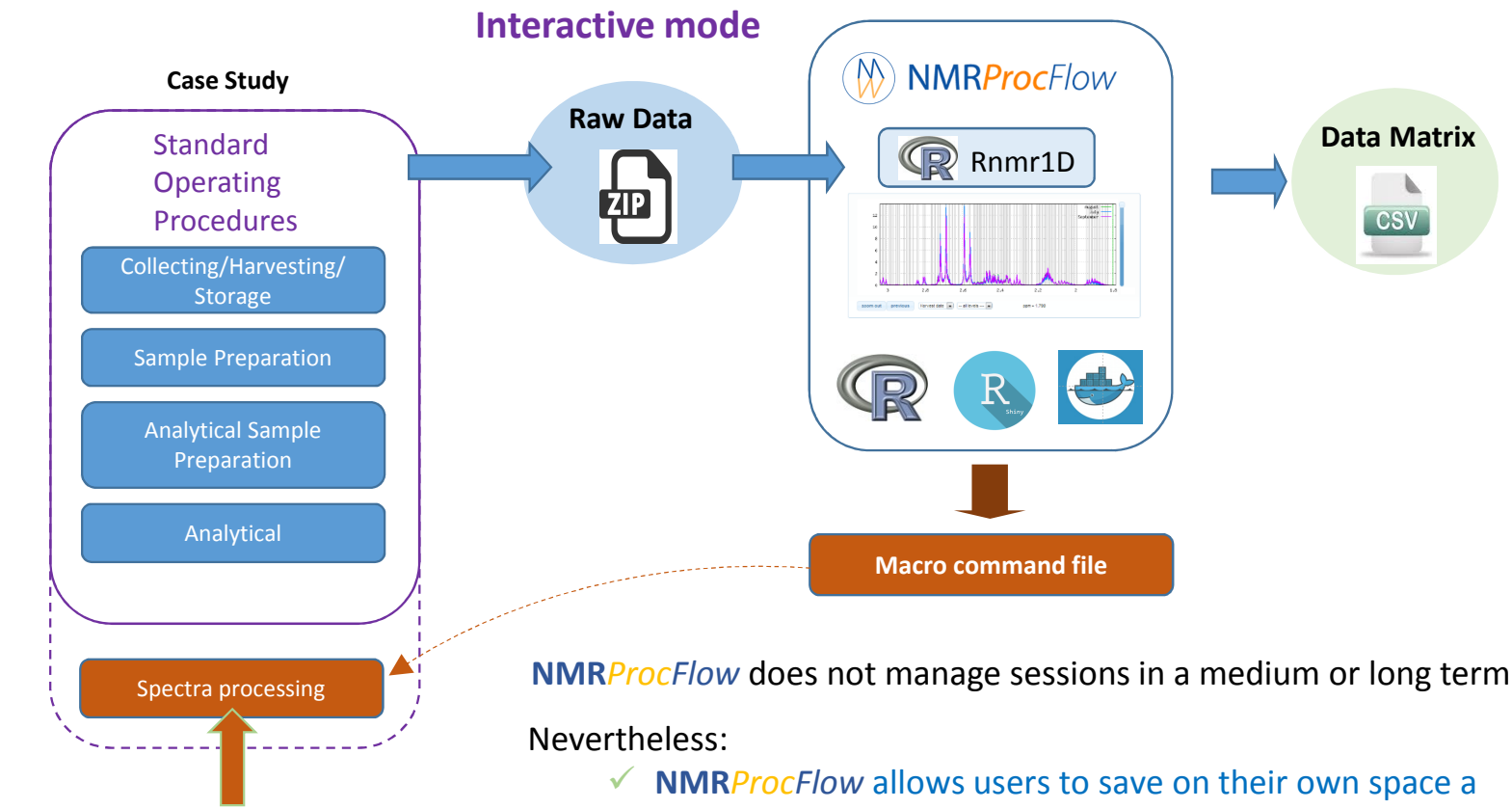


By taking into account the previous consideration, the strenghtness of Galaxy is to :

- Allows to process large dataset sets (ie> 300)
- **Allows non-experts to handle their data**

NMRProcFlow

NMRProcFlow - Interactive mode execution



NMR*ProcFlow* does not manage sessions in a medium or long term period.

Nevertheless:

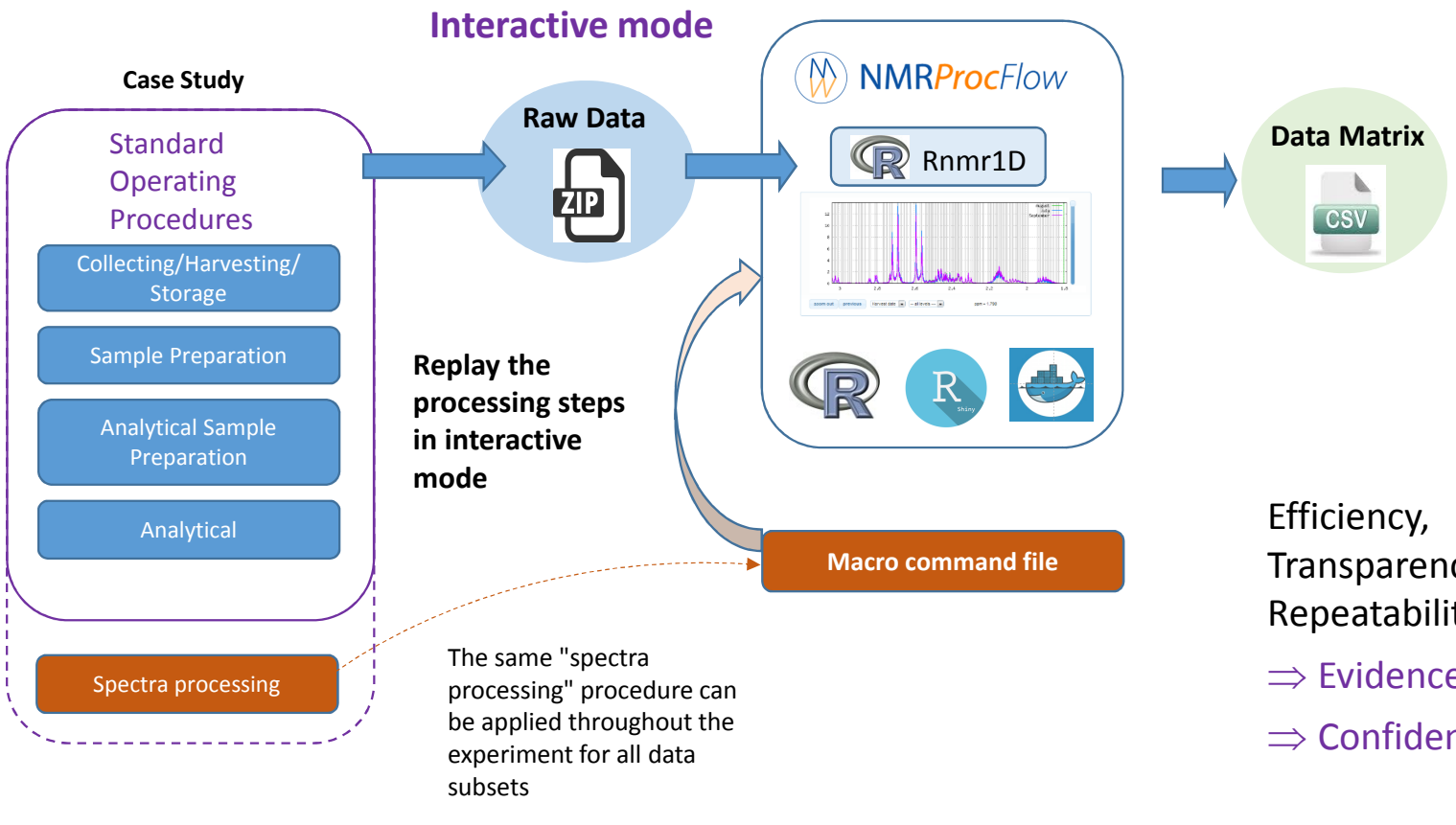
- ✓ **NMR*ProcFlow*** allows users to save on their own space a minimal set of small files in order to recover / replay their session

Completes the SOP list

- by an identical "spectra processing" procedure throughout the experiment for all data subsets

NMRProcFlow

NMRProcFlow - Interactive mode execution

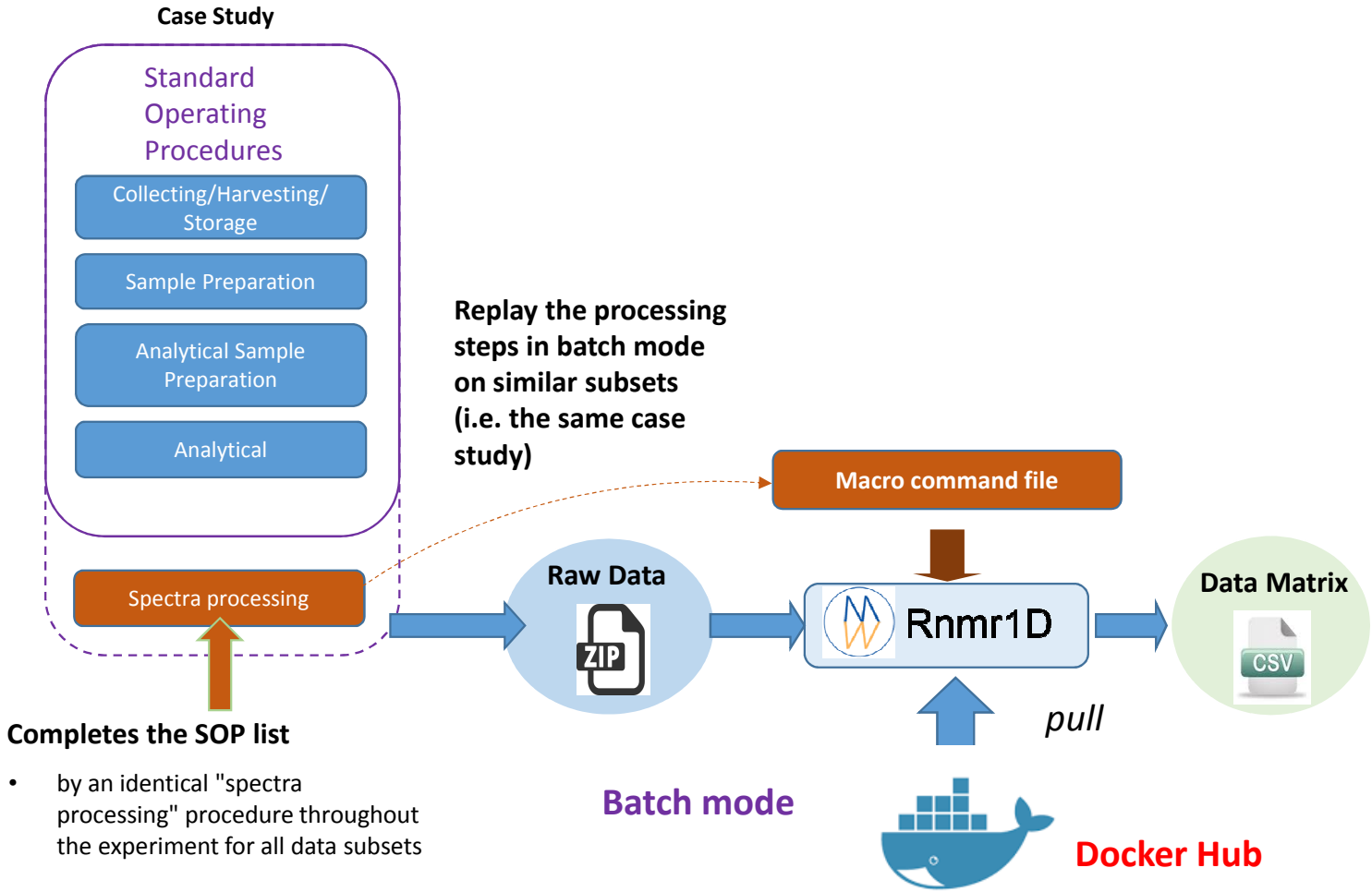


Efficiency,
Transparency,
Repeatability

⇒ Evidence

⇒ Confidence

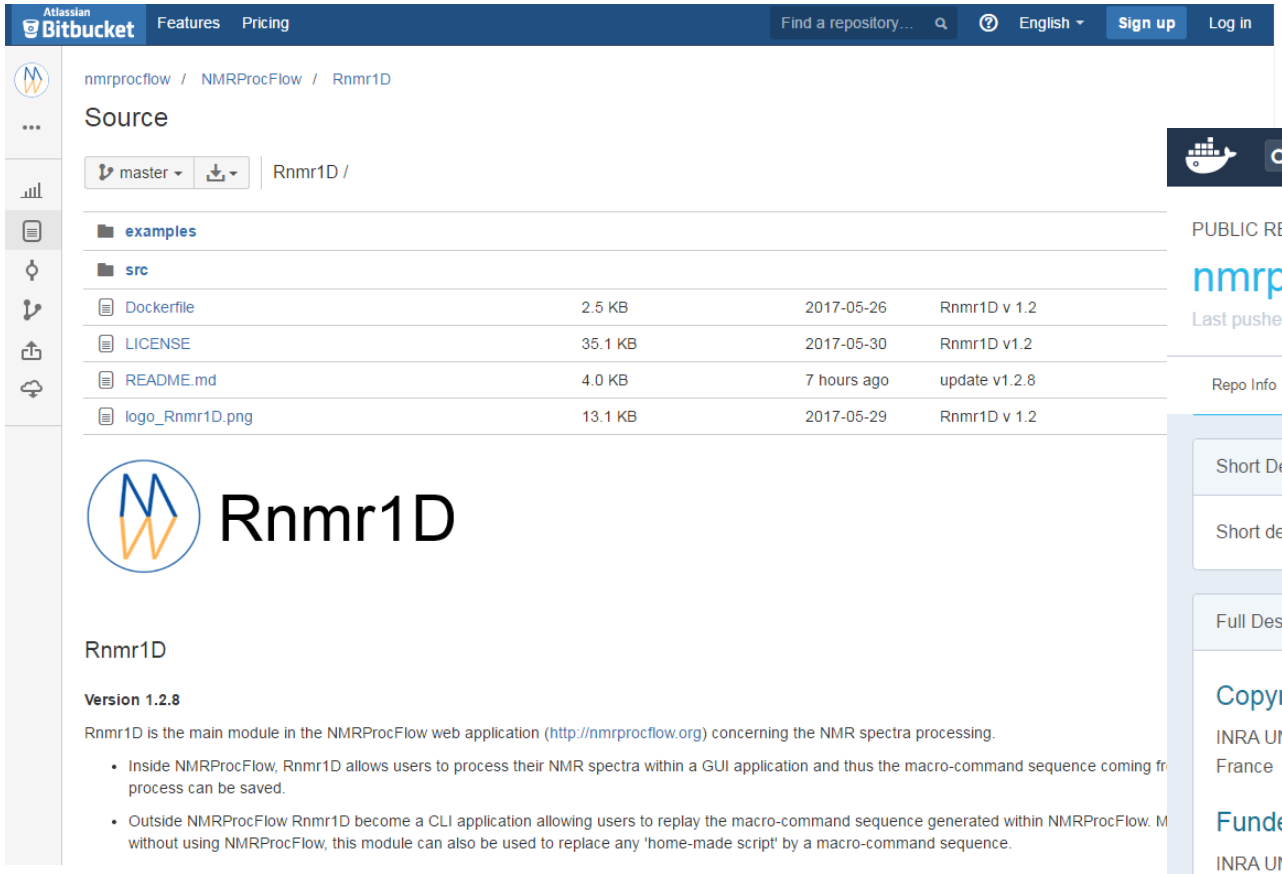
NMRProcFlow - Batch mode execution



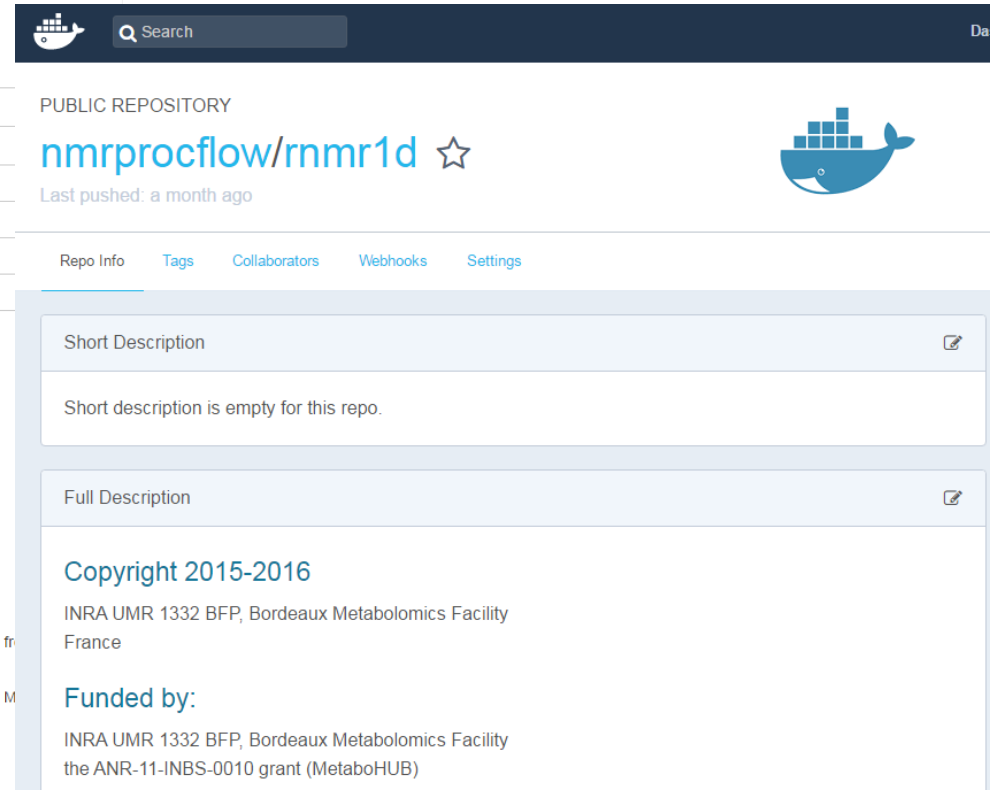
Completes the SOP list

- by an identical "spectra processing" procedure throughout the experiment for all data subsets

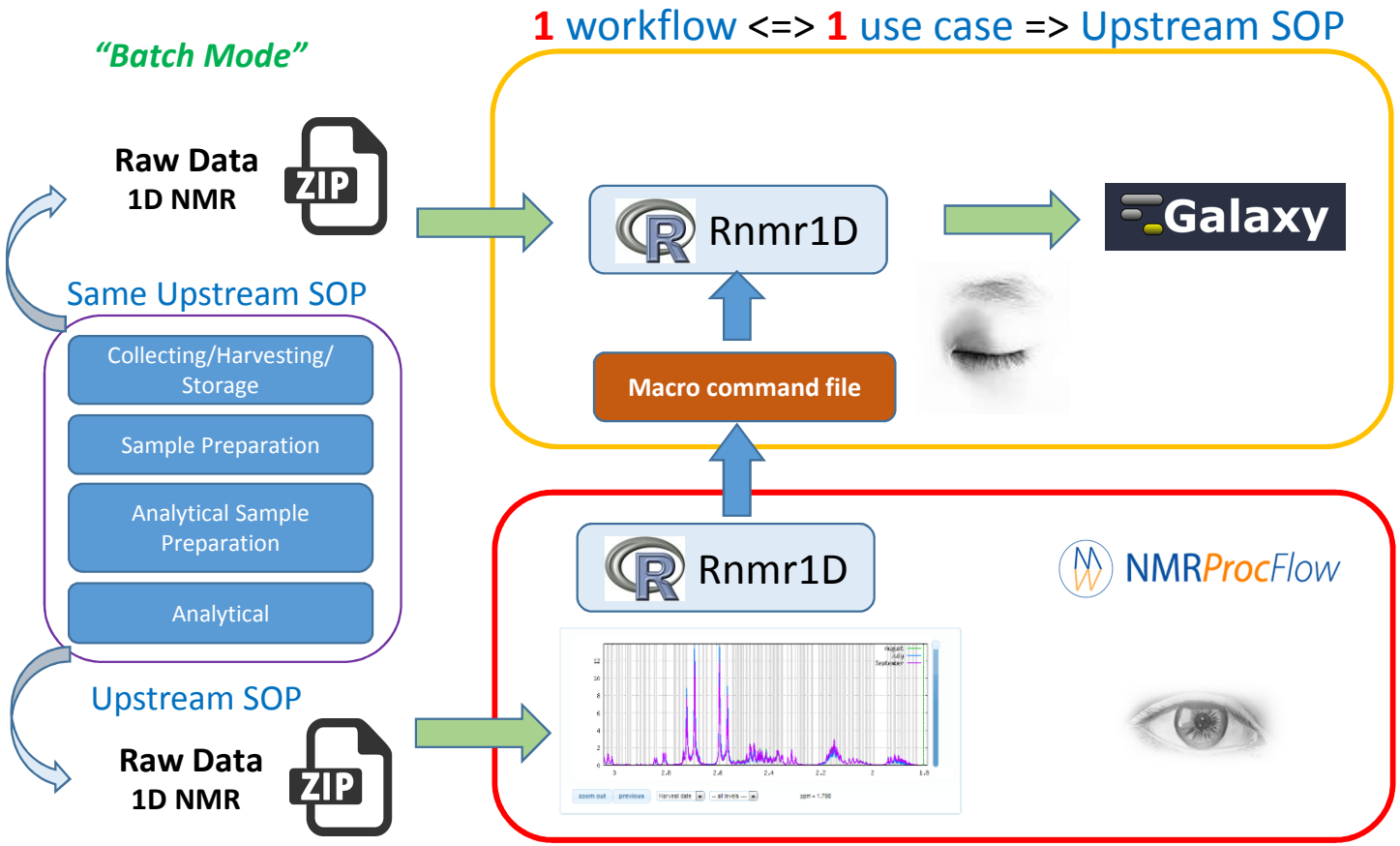
https://bitbucket.org/nmrprocflow/rnmr1d



https://hub.docker.com/r/nmrprocflow/rnmr1d/



NMRProcFlow & Galaxy – Combine strength of each one

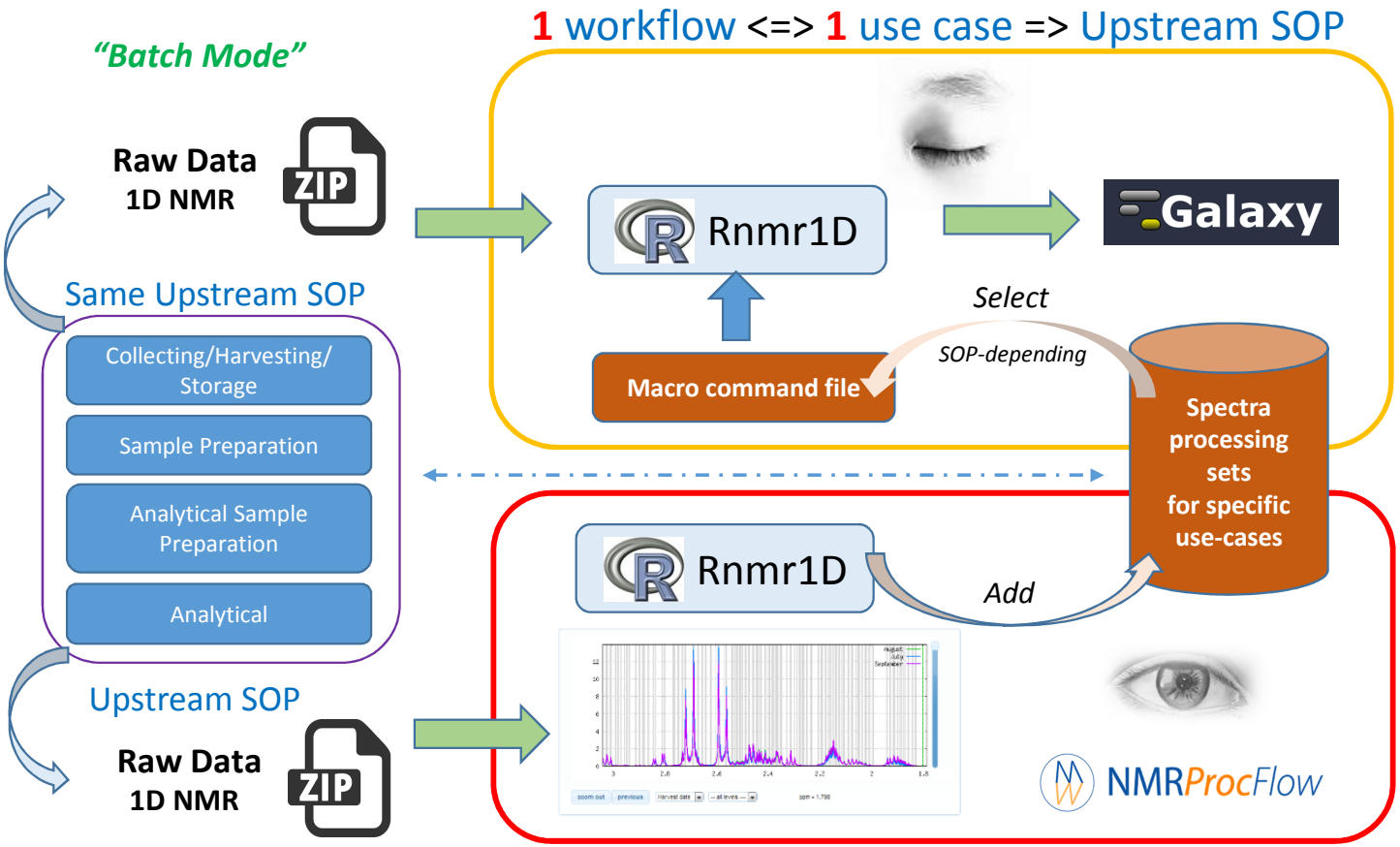


Many pitfalls to overcome involving specific settings (case-depending)

Motivation for using Galaxy (in that way)

- Tuning of the spectra processing applied on a subset of the NMR spectra (~ 100) within NMRProcFlow,
- Then applying the processing upon the whole set (i.e.> 300) within Galaxy

NMRProcFlow & Galaxy – Combine strength of each one



Many pitfalls to overcome involving specific settings (case-depending)

Motivation for using Galaxy (in that way)

- Expert settings checked and validated for each corresponding use case
- Allow non-experts to handle their data **with confidence and efficiency**

PhenoMeNal
Large-Scale Computing for Medical Metabolomics

http://phenomenal-h2020.eu/home/about/objectives/

Joint Research Activities

Work Package 9: Tools, Workflows, Audit and Data Management

WP9 (led by IPB) aims to develop and maintain the primary scientific- and technological tools and corresponding interfaces. We will support the data standards defined by WP8 and facilitate the interoperability of tools both within this consortium and those externally developed by the community. We will establish distributed tools for phenomics, metabolomics and bioinformatics processing pipelines and workflows, including longitudinal primary research data management (continuous availability to avoid data lock-in) and data audit mechanisms, as well as quality assurance schemes. Thus, this work package will produce several tailored VMIs which will be the basis for the service activities in WP5.

Objectives:

- Specify and integrate software pipelines and tools utilised in the PhenoMeNal e-Infrastructure into VMIs, adhering to data standards developed in WP8 and supporting the interoperability and federation middleware developed in WP5. Most tools will be already available (see table 1.1) and we will develop new applications to complete 'missing links' in pipelines. Although two explicit releases for VMIs are listed as deliverables below, we will use public repositories and continuous integration to always provide development snapshots of the infrastructure VMIs.
- Develop methods to scale-up software pipelines for high-throughput analysis, supporting distributed execution on e.g. local clusters, private clouds, federated clouds, or GRIDs.
- Add quality control and quality assurance to pipelines to ensure high quality and reliable data, keep an audit trail of intermediate steps and results.
- Develop methods to present and summarize the results of the pipelines in biomedical and disease contexts.

PhenoMeNal
Large-Scale Computing for Medical Metabolomics

http://phenomenal-h2020.eu/home/about/objectives/

Joint Research Activities

Work Package 9: Tools, Workflows, Audit and Data Management

Deliverables:

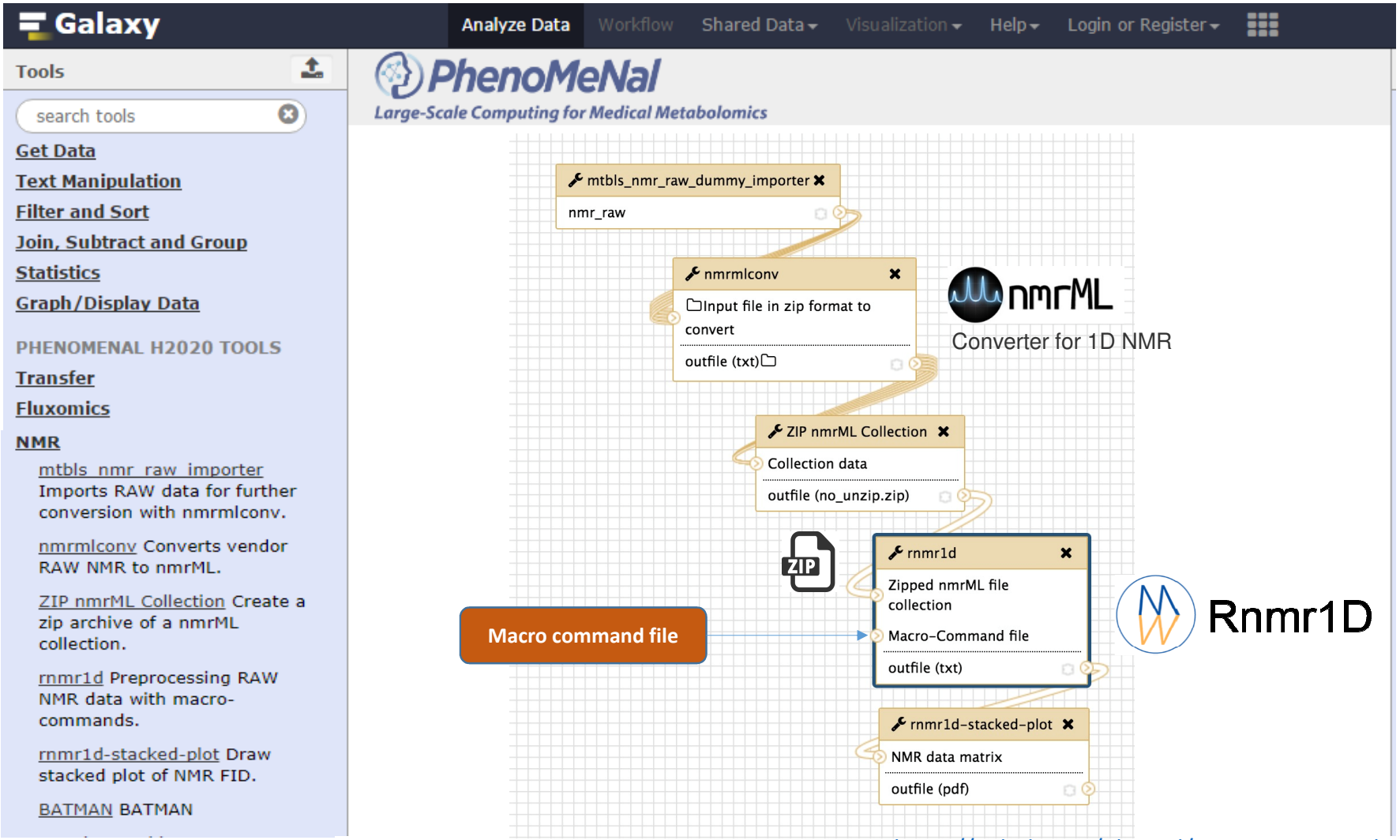
- Report on existing software tools, workflows and analytical pipelines initially supported in the PhenoMeNal grid – D9.1
- PhenoMeNal-Preprocess Virtual Machine Image to enable data producers to locally process raw data into standard formats supported in PhenoMeNal – D9.2.1
- PhenoMeNal-Data Virtual Machine image to enable sharing and dissemination of standardised and processed omics data to participating online repositories, like MetaboLights – D9.2.2
- Services Virtual Machine Image to facilitate the PhenoMeNal toolsets and pipelines, both locally and in the grid – D9.2.3
- Compute Virtual Machine Image to enable standardised compute capabilities for all the grid supplying partners – D9.2.4

Work Package Title	WP9 Tools, Workflows, Audit and Data Management
Deliverable Title	D9.2.1 *PhenoMeNal-Preprocess* Virtual Machine Image to enable data producers to locally process raw data into standard formats supported in PhenoMeNal
Delivery Date	M12
Work Package leader	IPB
Contributing Partners	EMBL-EBI, IPB, CRS4, UU, SIB, CEA, ICL, UB
Authors	Kristian Peters, Daniel Schober, Steffen Neumann, Pablo Moreno, Reza Salek, Luca Pireddu, Rico Rueedi, Roger Mallol, Ibrahim Karaman, Tim Ebbels, Pedro de Atauri, Etienne Thévenot, Pierrick Roger
Abstract: This deliverable report on the development of container images of tools to convert raw data into standard formats and their integration into the PhenoMeNal build and deployment infrastructure.	

D9.2.1 - 3.5.Integration of tools into Workflows

http://public.phenomenal-h2020.eu/

D9.2.1 - 3.5.NMR Workflow

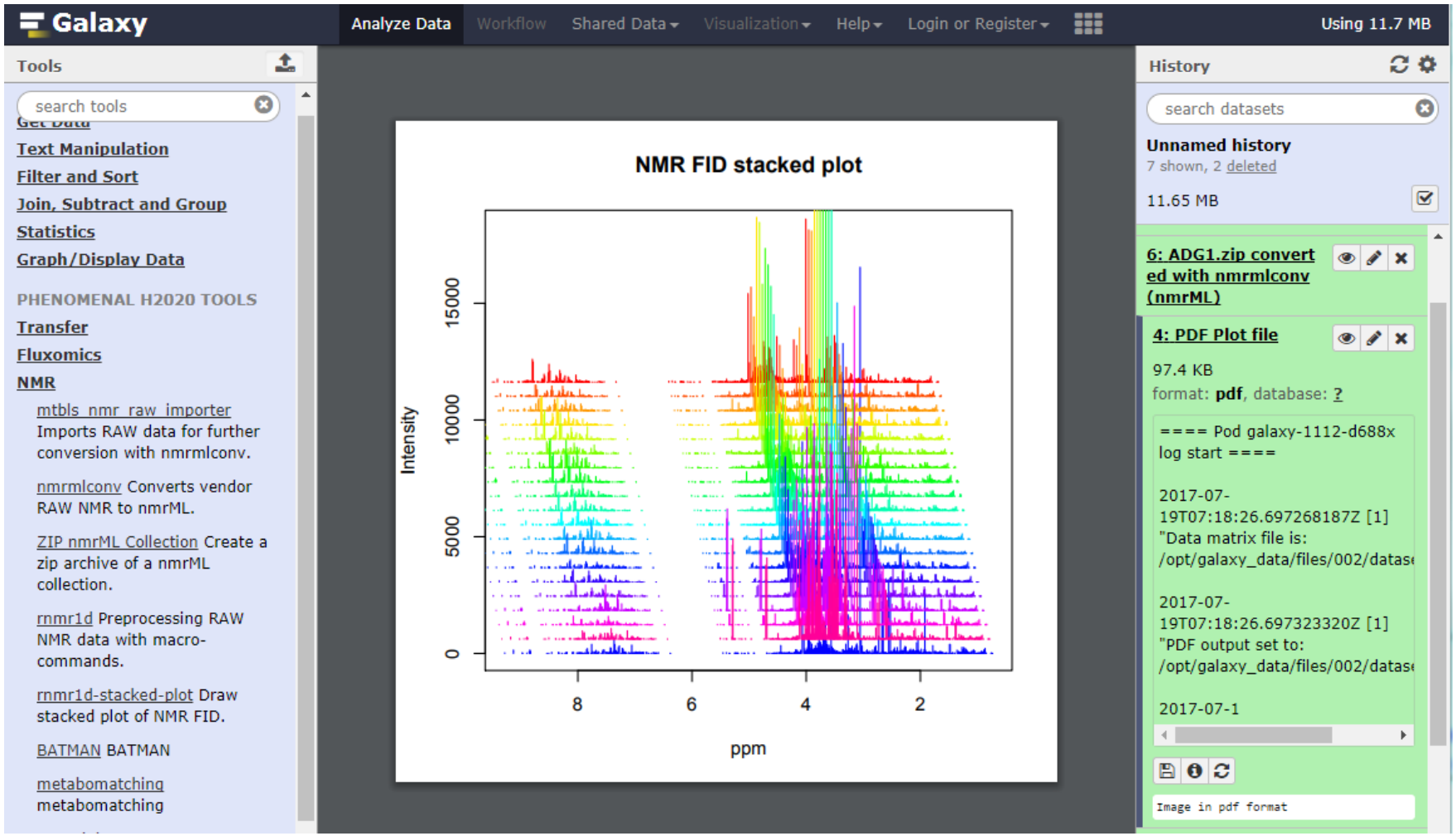


https://github.com/phnmnl/container-rnmr1d

Thanks to
Kristian Peters

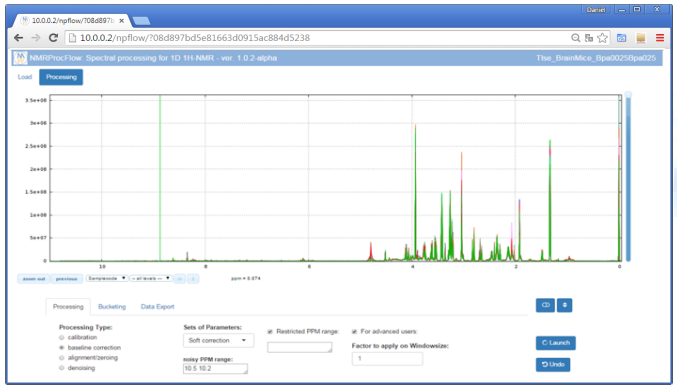
Leibniz Institute of Plant Biochemistry

D9.2.1 - 3.5.NMR Workflow



Thanks to
Kristian Peters

NMRProcFlow



An efficient GUI tool
for spectra processing
from 1D 1H-NMR metabolomics data

http://nmrprocflow.org/

Daniel Jacob

PhenoMeNal
Large-Scale Computing for Medical Metabolomics

Galaxy

http://phenomenal-h2020.eu/home/

combined with the computing power of
a Cloud Research Environment

Thank you for your attention

Thanks to
Kristian Peters

Leibniz Institute of Plant Biochemistry