

Spectra: A Flexible Infrastructure for Mass Spectrometry Data

Laurent Gatto¹, Johannes Rainer² and Sebastian Gibb³

¹ de Duve Institute, UCLouvain, Brussels, Belgium
² Institute for Biomedicine, Eurac Research, Italy
³ Department of Anaesthesiology and Intensive Care, University Medicine Greifswald, Germany

Mass Spectrometry (MS) Data

- ▶ Paired measurements:
 - ▶ intensity (counts) vs m/z (mass-to-charge ratio).
 - ▶ intensity vs retention time.
- ▶ Additional various metatdata.

Problems

- ▶ Vendor-specific, closed file formats.
- ▶ Various open-formats, e.g. mzML, mzXML, mz5, CDF, MGF, CSV.
- ▶ Spectra libraries/Annotation information in databases, e.g. SQL.
- ▶ Multiple R -packages require different file formats and/or data structures.
- ▶ Larger experiments don't fit into RAM.

Solution: Spectra[1]

- ▶ Common, flexible data structure and interface.
- ▶ Supports various file formats by different backends.

Spectra class

- ▶ Handles one or multiple MS spectra.
- ▶ Implements typical processing functionality.
- ▶ Isolation from type of data storage (handled by backend classes)

MsBackend classes

- ▶ Isolation from spectra handling and processing.
- ▶ Specific class for each type of data storage.

class	source.files	storage	writeable
MsBackendDataFrame[1]	manual	in-memory	yes
MsBackendHdf5Peaks[1]	hdf5	on-disk	yes
MsBackendMzR[1]	mzML, mzXML and CDF	on-disk	no
MsBackendHmdbXml[2]	XML	on-disk	yes
MsBackendRawFileReader[3]	Thermo .raw	on-disk	no

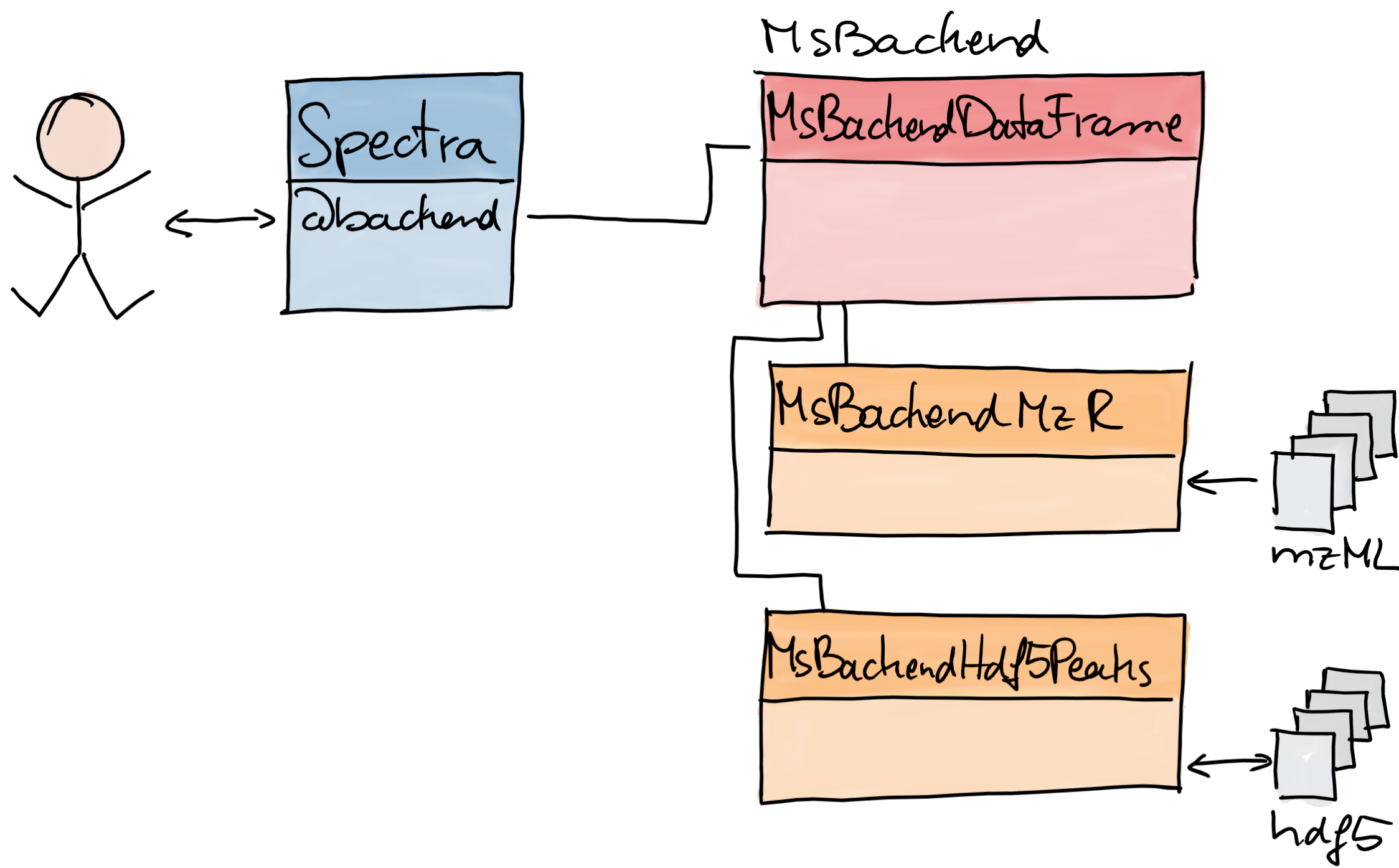


Figure 1: The user interacts with a Spectra object. It takes care of the data storage by different backend classes.

Example

```
# Load packages
library("Spectra")
library("magrittr")

# Import data
#sps <- Spectra("data/20191107_Mix2_CE20.mzML", backend = "mzML")
sps <- Spectra()

# Select all MS2 spectra for a [M+Na]+ ion of Fructose
mzr <- 203.0526
mzr <- 156.07675 # Histidine
fruct <- sps %>%
  filterMsLevel(2) %>%
  filterPrecursorMz(
    mz = mzr + ppm(c(-mzr, mzr), 20)
  )

# Centroid and clean spectra.
fruct <- fruct %>%
  pickPeaks() %>%
  removePeaks(threshold = 500) %>%
  clean(all = TRUE)

# Change backend: load data in memory.
#fruct <- setBackend(fruct, MsBackendDataFrame())
```

Contact

- ▶ <https://github.com/rformassspectrometry/Spectra>
- ▶ <https://rformassspectrometry.org>
- ▶ <https://twitter.com/search?q=%23rformassspec>

References

Laurent Gatto, Johannes Rainer, and Sebastian Gibb.
Spectra: Spectra Infrastructure for Mass Spectrometry Data.
R package version 0.3.0.

Laurent Gatto, Johannes Rainer, and Sebastian Gibb.
MsBackendHmdb: Mass spectrometry data backends for The Human Metabolome Database, 2019.
R package version 0.1.0.

Christian Panse and Tobias Kockmann.
MsBackendRawFileReader: Bridging Spectra and ThermoFinnigan raw files, 2019.
R package version 0.0.1.

To use the Spectra package install the **R for Mass Spectrometry** packages and load the core packages:

```
BiocManager::install("RforMassSpectrometry/RforMassSpectrometry")
library("RforMassSpectrometry")
```