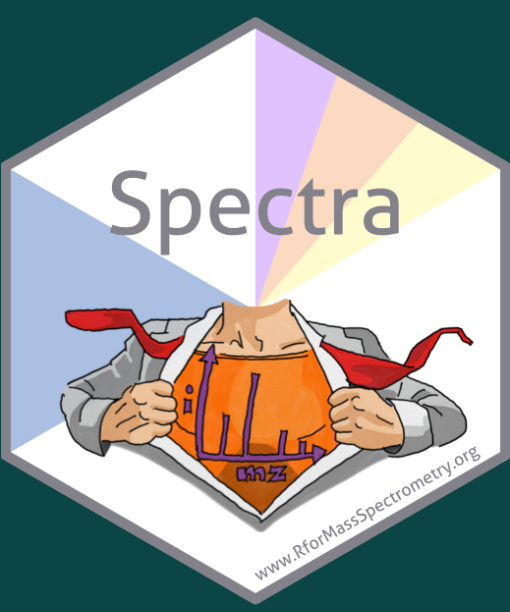


Flexible Infrastructure for Mass Spectrometry Data



Laurent Gatto¹, , Johannes Rainer², , Sebastian Gibb³, 

¹ Computational Biology and Bioinformatics Group, de Duve Institute, UCLouvain
² Institute for Biomedicine, Eurac Research

³ Department of Anaesthesiology and Intensive Care, University Medicine Greifswald



Introduction

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.

Solution

Spectra offers a common, flexible data structure and interface.

Data Structure

class	storage	writeable
MsBackendDataFrame ¹	in-memory	yes
MsBackendHdf5Peaks ¹	on-disk	yes
MsBackendMzR ¹	on-disk	no
MsBackendHmdbXml ²	on-disk	yes
MsBackendRawFileReader ³	on-disk	no

Interface

¹

³

Contact

[rformassspectrometry/Spectra](https://github.com/rformassspectrometry/Spectra)

References

1. Gatto, L., Rainer, J. & Gibb, S. *Spectra: Spectra infrastructure for mass spectrometry data*.
2. Gatto, L., Rainer, J. & Gibb, S. *MsBackendHmdb: Mass spectrometry data backends for the human metabolome database*. (2019).
3. Panse, C. & Kockmann, T. *MsBackendRawFileReader: Bridging spectra and thermofinnigan raw files*. (2019).

