

# Mass spectrometry and Free Software in Debian

Filippo RUSCONI, Ph.D.  
filippo.rusconi@u-psud.fr

Laboratoire de Chimie Physique  
CNRS UMR 8000  
Université Paris-Sud 11  
F-91405 Orsay

FOSDEM — Brussels — February the 2<sup>nd</sup>-3<sup>rd</sup> 2013

## Mass spectrometry:

- ▶ Uses in the biochemical sciences;
- ▶ Cultural differences with other disciplines;
- ▶ Why free software is ever more considered essential;
- ▶ Software available in Debian and other packaging work.

# A mass spectrometer

Source, analyser and detector (ion counter)

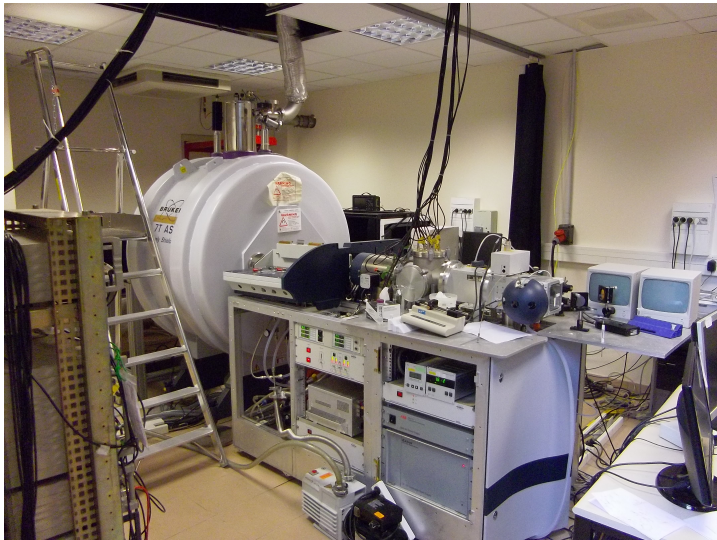
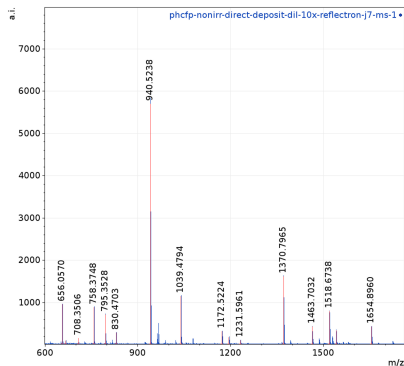


Photo: Vincent Steinmetz (U-Psud, Orsay)

# A mass spectrum

Detected ion masses *versus* the count of the ions



- ▶ Proteins, DNA | RNA, Sugars → function and regulation (disease understanding);
- ▶ The applied sciences (big pharma) → drug development needs checking at the cell level the effect of a given drug. Mass spec shows the changes occurring upon treatment.

# Why is Free Software so important here?

In the context of mass spectrometry for biology?

Hardware manufacturers are fiercely struggling to gain exclusive control both on the mass data and on the users themselves...

*(“vendor lock-in”)*

- ▶ Software is used as a sales pitch (particularly, LIMS<sup>†</sup>);
- ▶ Proprietary formats lock terabytes of (sensitive) data;
- ▶ Inter-project interoperability is hindered by the file format wars;
- ▶ MS facility users cannot themselves analyze their data;
- ▶ ⇒ Always ask for an “Export to mzML”<sup>‡</sup> feature.

<sup>†</sup> Laboratory Information Management System.

<sup>‡</sup> Or, at least, to a simple  $(x,y)$  format.

# Cultural differences with other disciplines

Popularization of biopolymer mass spectrometry is recent

## Genomics/Bioinformatics Old-style biochemistry

- ▶ Mid 1970's: sequencing of proteins/nucleic acids  
⇒ simple-text raw data  
⇒ rather easy data processing;
- ▶ PDP-11, VAX/VMS, DEC-ALPHA, UNIX...
- ▶ Easy text-based formatted-data display.

## Mass spectrometry New-age hyper-technicity

- ▶ Mid 2000's: popularity of mass spectrometry  
⇒ (x,y) text raw data  
⇒ more complex data processing;
- ▶ MS-Windows NT4, Millennium, 2000...
- ▶ Complex graphics-based formatted-data display.

These chrono-technological differences impacted the unconscious computer-based behaviour of scientists

Since a few years, I see a behavioural shift...  
that might foreshadow a mindset transition...

### **Pride in stating:**

- ▶ *“Developed under GNU/Linux with FREE SOFTWARE;”*
- ▶ *“We can use tens of computers; spending nothing in licenses;”*

### **But not yet something like:**

- ▶ *“We packaged our soft and set up a public repos;”*
  - ▶ *“How about having a powerful environment for both mass spec-centric software development and data analysis?”*
- ⇒ This is the proper time window to push *freedom* forward

# Debian as a full-featured mass spectrometry setup?

## Three kinds of users

- ▶ The massist — acquires the data and handles them:
  - ▶ Databases;
  - ▶ Nicely crafted graphics-based site-specific workflows;
  - ▶ Some reporting.
- ▶ The user — a biologist (or else) mainly willing to click:
  - ▶ Nicely crafted application-specific number crunching;
  - ▶ Reporting tools;
- ▶ The developer: the guy willing to help the former.
  - ▶ Interfacing with databases (*i.e.* web interfaces);
  - ▶ Number-crunching coding C/C++/Python/Java;
  - ▶ Build systems (CMake, GNU autotools, bjam. . . );
  - ▶ Documentation-generation tools (sgml, LaTeX, TexInfo);



# Debian source (binary) packages for mass spectrometry

- ▶ Before 2002: nothing;
- ▶ Since my switching to Debian:
  - ▶ `polyxmass` (2) and `massxpert` (3)
  - ▶ `lutefisk` (1)
  - ▶ `mmass` (2)
- ▶ Since a recent push to package new stuff:
  - ▶ libraries as tools to craft specific workflows:
    - ▶ `openms` (5)
    - ▶ `libpwiz` (3)
    - ▶ `tandem-mass` (1)
  - ▶ libraries/executables to handle data (format conversion, quantitation...):
    - ▶ `python-mzml` (2)
    - ▶ `r-cran-readbrukerflexdata` (1)
    - ▶ `r-cran-maldivant` (1)

# One example of a large software project

OpenMS (<http://open-ms.sourceforge.net/>)

- ▶ C++; LGPL 2.1; CMake-based;
- ▶ 2 libraries (fundamental and GUI);
- ▶ 114 binaries;
- ▶ Large documentation;
- ▶ **contrib** directory stuff with external libraries;
- ▶ “Hackish” DESTDIR-based design decisions to be revisited;
- ▶ Useful interactions with various project members;
- ▶ Build-Depends: debhelper ( $\geq 7.0.50$ ), dpkg-dev ( $\geq 1.16.1$ ), quilt ( $\geq 0.60-2$ ), cmake ( $\geq 2.6.3$ ), libxerces-c-dev ( $\geq 3.1.1$ ), libgsl0-dev ( $\geq 1.15+dfsg$ ), libboost1.49-dev, libboost-iostreams1.49-dev, libboost-date-time1.49-dev, libboost-math1.49-dev, seqan-dev ( $\geq 1.3.1$ ), libsvm-dev ( $\geq 3.12$ ), libglpk-dev ( $\geq 4.45$ ), zlib1g-dev ( $\geq 1.2.7$ ), libbz2-dev ( $\geq 1.0.6$ ), cppcheck ( $\geq 1.54$ ), libqt4-dev ( $\geq 4.8.2$ ), libqt4-opengl-dev ( $\geq 4.8.2$ ), libqtwebkit-dev ( $\geq 2.2.1$ ), coinor-libcoinutils-dev ( $\geq 2.6.4$ ), imagemagick, doxygen ( $\geq 1.8.1.2$ ), texlive-extra-utils, texlive-latex-extra, latex-xcolor, texlive-font-utils, ghostscript, texlive-fonts-recommended

## Using Debian for such a project — PROS...

- ▶ Huge amount of pre-packaged software (libraries, particularly);
- ▶ Basis of many derivatives;
- ▶ Debian Pure Blends infrastructure (*Debichem*);
- ▶ Wonderful and welcoming infrastructure for collaborative packaging ([alioth.debian.org](http://alioth.debian.org));
- ▶ Robustness of the *stable* (even *testing*) distributions;

## Using Debian for such a project — CONS...

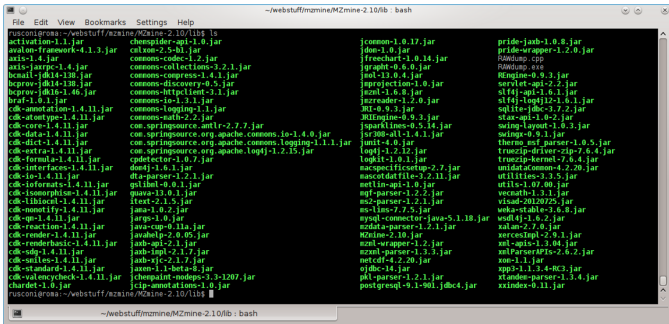
- ▶ Packaging is highly involved and may hinder the involvement of colleagues who do not want to invest too much time. Mentoring plays a *huge* role here;
- ▶ Intimidating distribution for the average biochemist (specifically at install time). This should become ever less true with the installer progress since some time;

# Challenges...

Rocky packaging ahead...

Java-based sophisticated mass spectrum viewer at

<http://mzmine.sourceforge.net>



```
~/webstuff/mzmine/MZmine-2.10.10b : bash
File Edit View Bookmarks Settings Help
rusccon@rora:~/webstuff/mzmine/MZmine-2.10.10b$ ls
activation-1.1.jar      commons-collections-3.2.1.jar
avalon-framework-4.1.3.jar  commons-codem-1.2.jar
axis-1.4.jar           commons-collectio-3.2.1.jar
axis-jaxrpc-1.4.jar    commons-copress-1.4.1.jar
bcmail-jdk14-138.jar   commons-discovery-0.5.jar
bcprov-jdk14-138.jar   commons-httpclient-3.1.jar
bcprov-jdk15-1.46.jar  commons-io-1.3.1.jar
brotli-0.1.jar         commons-logging-1.1.jar
cdk-annotation-1.4.11.jar  commons-math-2.2.jar
cdk-atontype-1.4.11.jar  com.springsource.antlr-2.7.7.jar
cdk-core-1.4.11.jar    com.springsource.org.apache.commons.io-1.4.0.jar
cdk-datas-1.4.11.jar   com.springsource.org.apache.commons.logging-1.1.1.jar
cdk-dict-1.4.11.jar    com.springsource.org.apache.log4j-1.2.15.jar
cdk-extra-1.4.11.jar   cpdetector-1.0.7.jar
cdk-formula-1.4.11.jar  dnetlib-1.0.1.jar
cdk-interfaces-1.4.11.jar  dita-parser-1.2.1.jar
cdk-io-1.4.11.jar      gslib-0.0.1.jar
cdk-ioformats-1.4.11.jar  guava-13.0.1.jar
cdk-isomorphism-1.4.11.jar  hext-2.1.5.jar
cdk-libhsca-1.4.11.jar  jama-1.0.2.jar
cdk-notify-1.4.11.jar  jargs-1.0.jar
cdk-qn-1.4.11.jar      java-cup-0.11a.jar
cdk-reaction-1.4.11.jar  javahelp-2.0.05.jar
cdk-render-1.4.11.jar  jaxb-api-2.1.jar
cdk-renderbasic-1.4.11.jar  jaxb-api-2.1.7.jar
cdk-sdg-1.4.11.jar     jaxb-impl-2.1.7.jar
cdk-smiles-1.4.11.jar  jaxb-rt-2.1.7.jar
cdk-standards-1.4.11.jar  jaxws-1.3-beta-8.jar
cdk-valuation-1.4.11.jar  jchempaint-nodesp-3.3-1207.jar
chardet-1.0.jar        jcip-annotations-1.0.jar
ruscon@rora:~/webstuff/mzmine/MZmine-2.10.10b$
jcommon-1.0.17.jar      pride-jaxb-1.0.8.jar
jdon-1.0.jar            pride-wrapper-1.2.0.jar
jfreechart-1.0.14.jar  RAndMup.cpp
jgraph-0.6.0.jar       RAndMup.exe
jmol-13.0.4.jar        REngine-0.9.3.jar
jprojection-1.0.jar    servlet-api-2.2.jar
jzml-1.6.8.jar         slf4j-api-1.6.1.jar
jzrender-1.2.0.jar     slf4j-log4j12-1.6.1.jar
JRI-0.9.3.jar          sqlite-jdbc-3.7.2.jar
JRIEngine-0.9.3.jar    stax-api-1.0-2.jar
jsparklines-0.5.14.jar  swing-layout-1.0.3.jar
jsr308-all-1.0.1.jar  swing-0.9.1.jar
junit-4.0.jar          thermo esf parser-1.0.5.jar
log4j-1.2.12.jar       truezip-driver-zip-7.6.4.jar
logkit-1.0.1.jar       truezip-kernel-7.6.4.jar
mascopelicesetup-2.7.jar  undatadommon-4.2.20.jar
mascotdatfile-3.2.11.jar  utilites-3.3.5.jar
netlin-api-1.0.jar     utils-1.07.00.jar
ngt-parser-1.2.2.jar   vecnath-1.1.1.jar
niz-parser-1.2.1.jar   vlnsd-20120725.jar
nc-lins-7.7.5.jar      wka-stable-3.6.8.jar
mysql-connector-java-5.1.18.jar  vsd4j-1.6.2.jar
nzdat-parser-1.2.1.jar  xalan-2.7.0.jar
nZIndex-2.10.jar       xercesImpl-2.8.1.jar
nxml-wrapper-1.2.jar  xml-apix-1.3.04.jar
nxml-parser-1.3.3.jar  xmlParserAPIs-2.6.2.jar
nxml-1.1.jar           xom-1.1.jar
nxml-1.3.4-RC3.jar    xmlsec-1.3.4-RC3.jar
pk1-parser-1.2.1.jar  xtndee-parser-1.3.4.jar
postgresql-9.1-901.jdbc4.jar  xxindex-0.11.jar
```

- ▶ Non-free but highly useful software;
- ▶ Databases of natural data (undistributable);