Introduction to MassBank

Takaaki Nishioka
MassBank Project, Graduate School of Information Science,
Nara Institute of Science and Technology

MassBank

MassBank

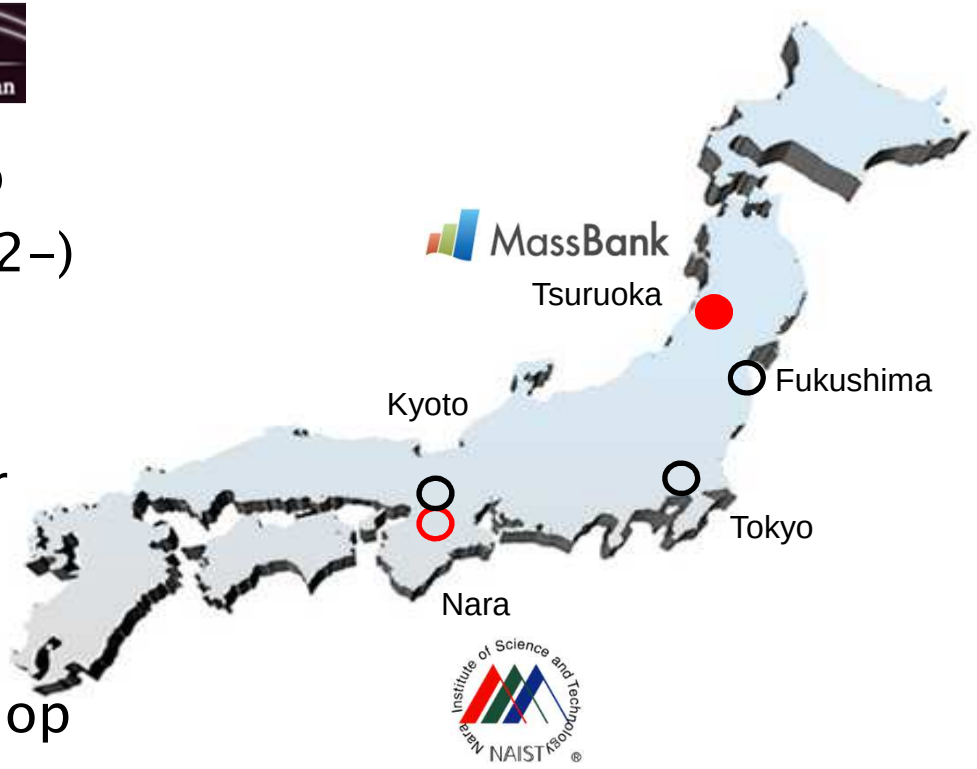
- is a public repository (2006–),
- has no instrument to produce data by itself,
- is the official DB of the Mass Spectrometry Society of Japan (2008–).

Mass Spectrometry
The Mass Spectrometry Society of Japan

NORMAN Association joined to
MassBank consortium (2012–)

We visit

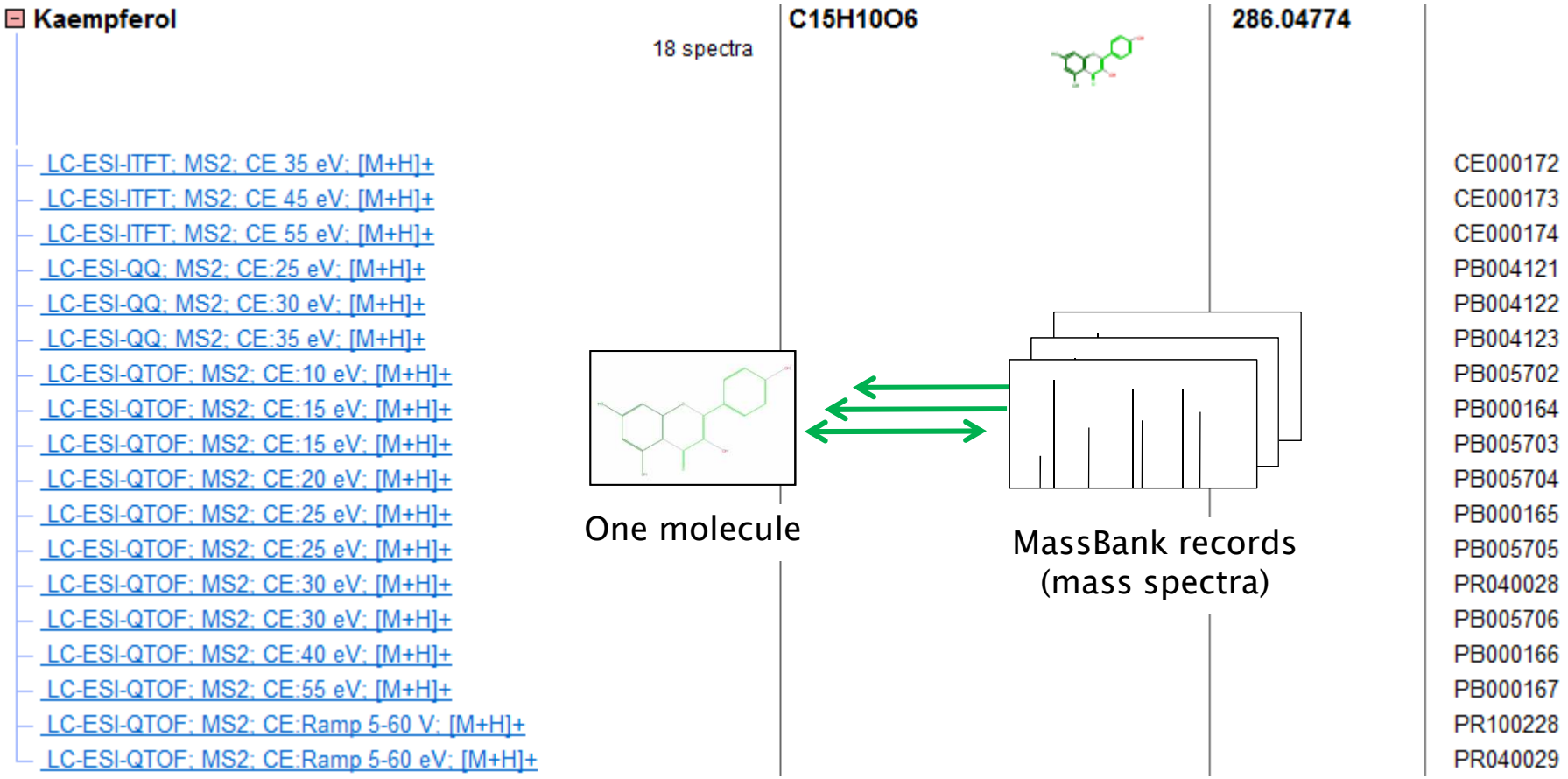
- UFZ to setup a mirror server
  (MassBank.eu)
- Amsterdam to attend
  NORMAN MassBank workshop



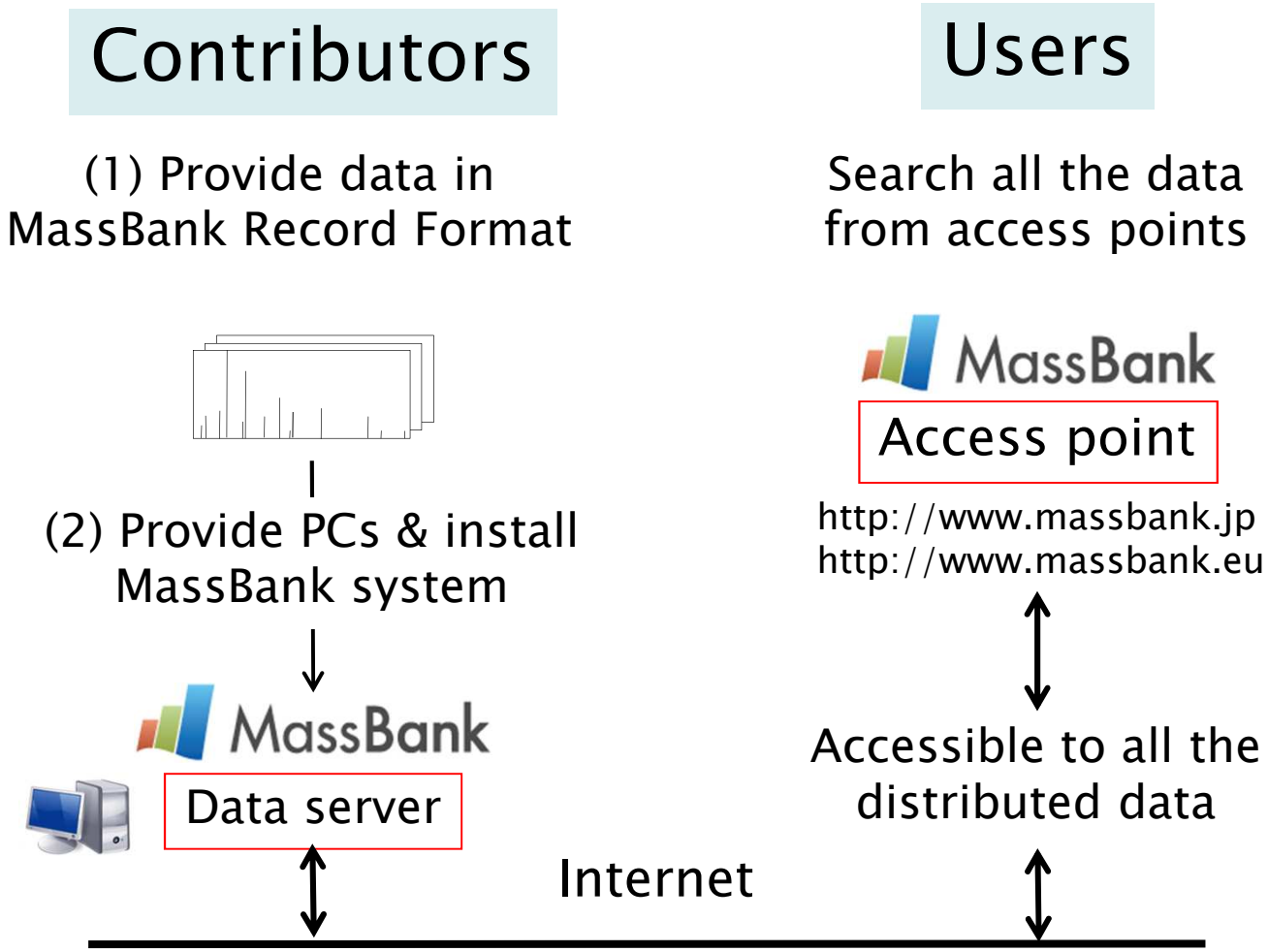
Concept (1)

- Public repository
- No standard analytical method
  - Reference data analyzed by as many as different analytical methods should be available.

Reference ESI-MS^2 Data of Different Methods



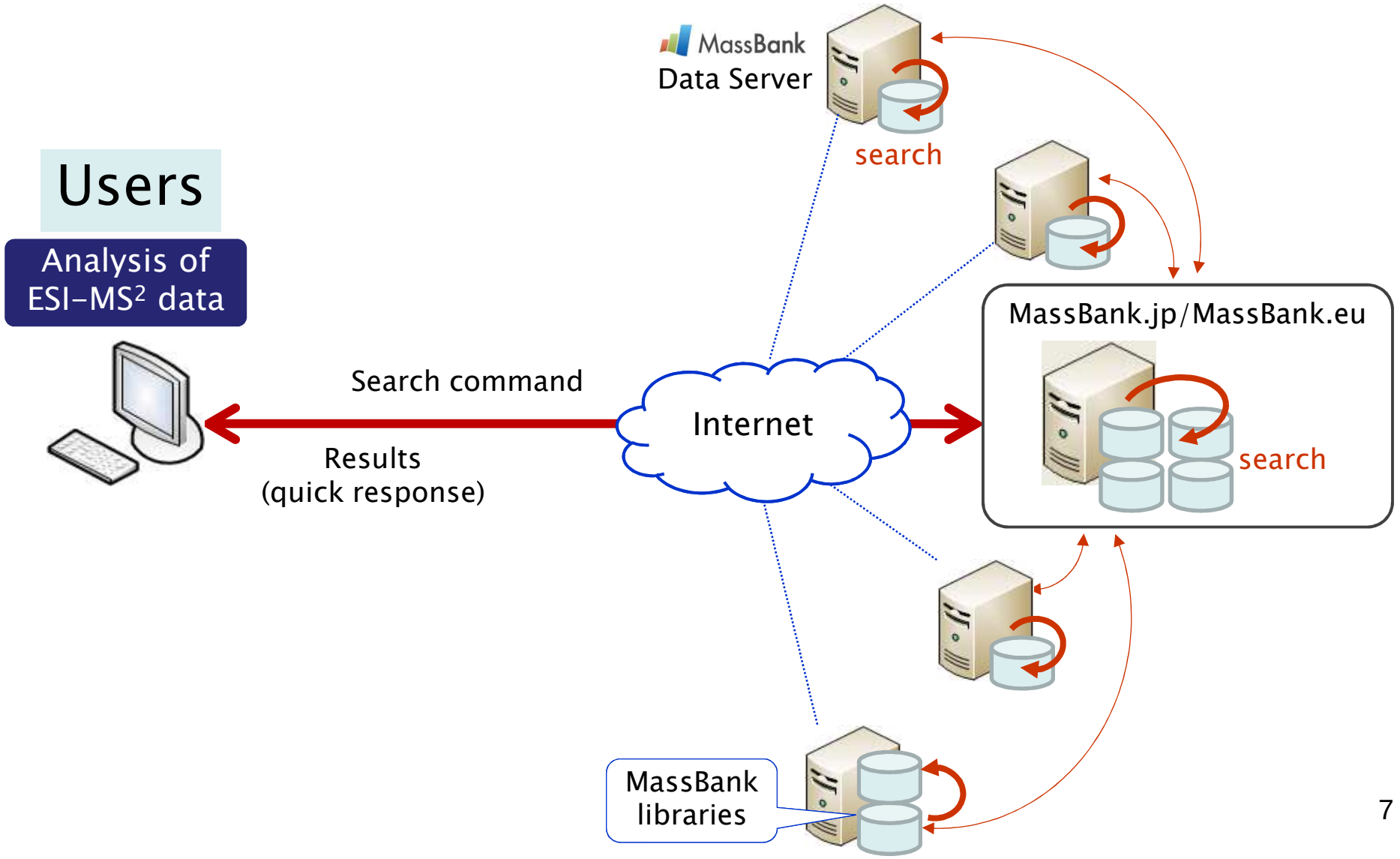
Concept (2): Distributed Database



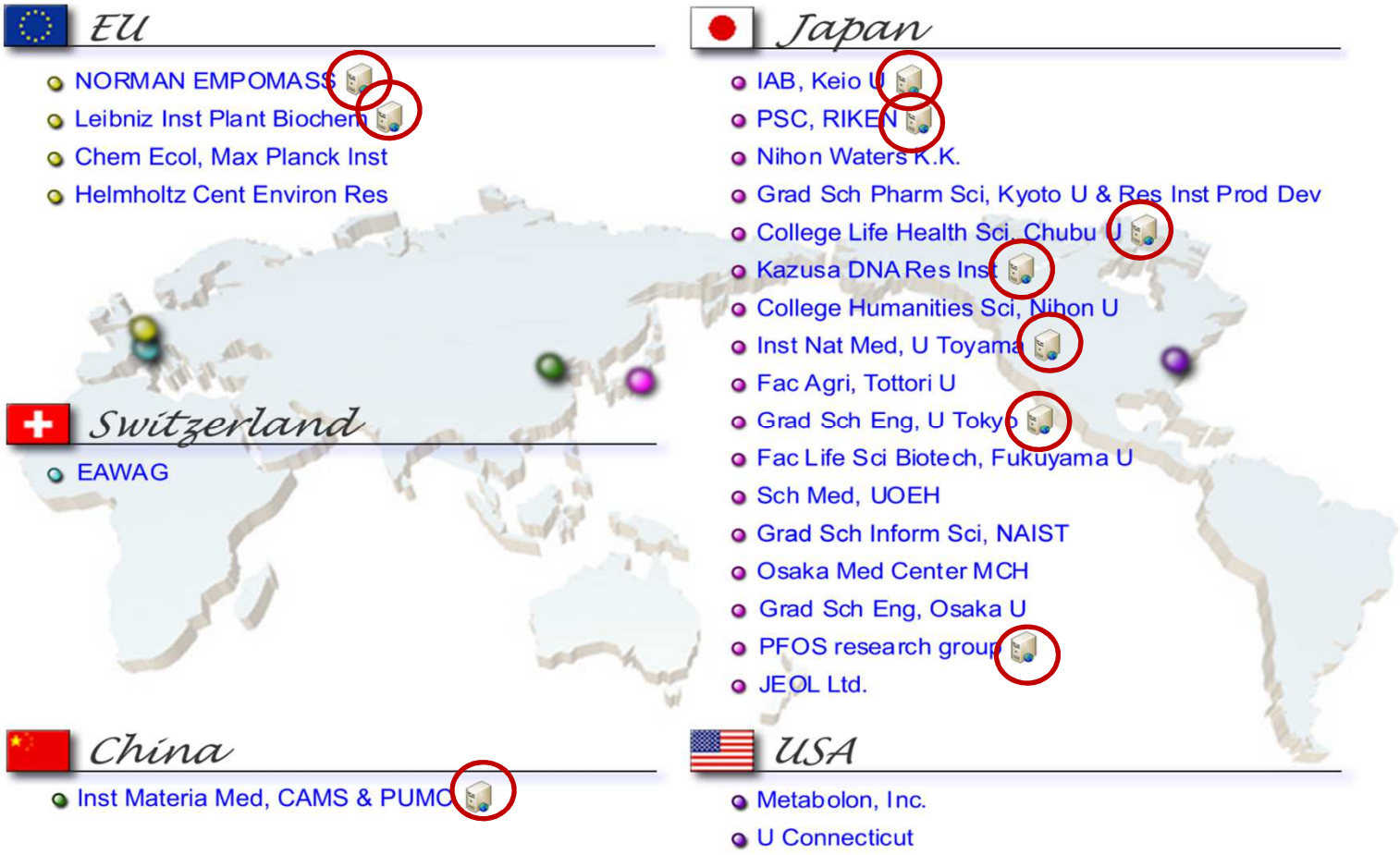
Features of Distributed Database

- Low-cost management
  - Contributors take part of managing the database in proportion to the amount of data.
- Quality of data
  - by contributor's responsibility
- Search in parallel
  - No super computers such as centralized database servers are necessary.

Parallel-Search Database



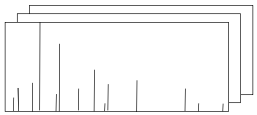
Current Statistics of MS Data on MassBank



A total of 25 research groups (17 Japan, 4 EU, 2 USA, 1 Switzerland and China groups) contribute 31,153 data from 10 data servers (August 2012). Monthly access is 9,980 unique IPs (October 2012). API accesses.

Tools for contributors and users

Tools for Contributors



Raw data

Mass++

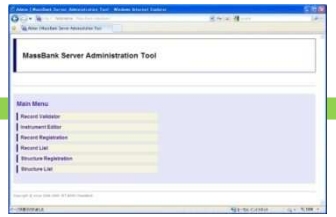
Mass++ (developed by Dr. Y. Oda at Eisai Co. Ltd)

- imports raw data (mz/XML/ML, netCDF)
- detects peaks from profile data
- outputs analytical methods (AC$, MS$) and peak data (PK$) in MassBank Record Format.

Contributors

Record Editor

MassBank record

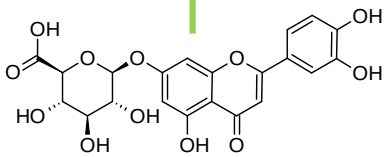


Administration Tool

MassBank

Data server

Chemical structure information (CH$)

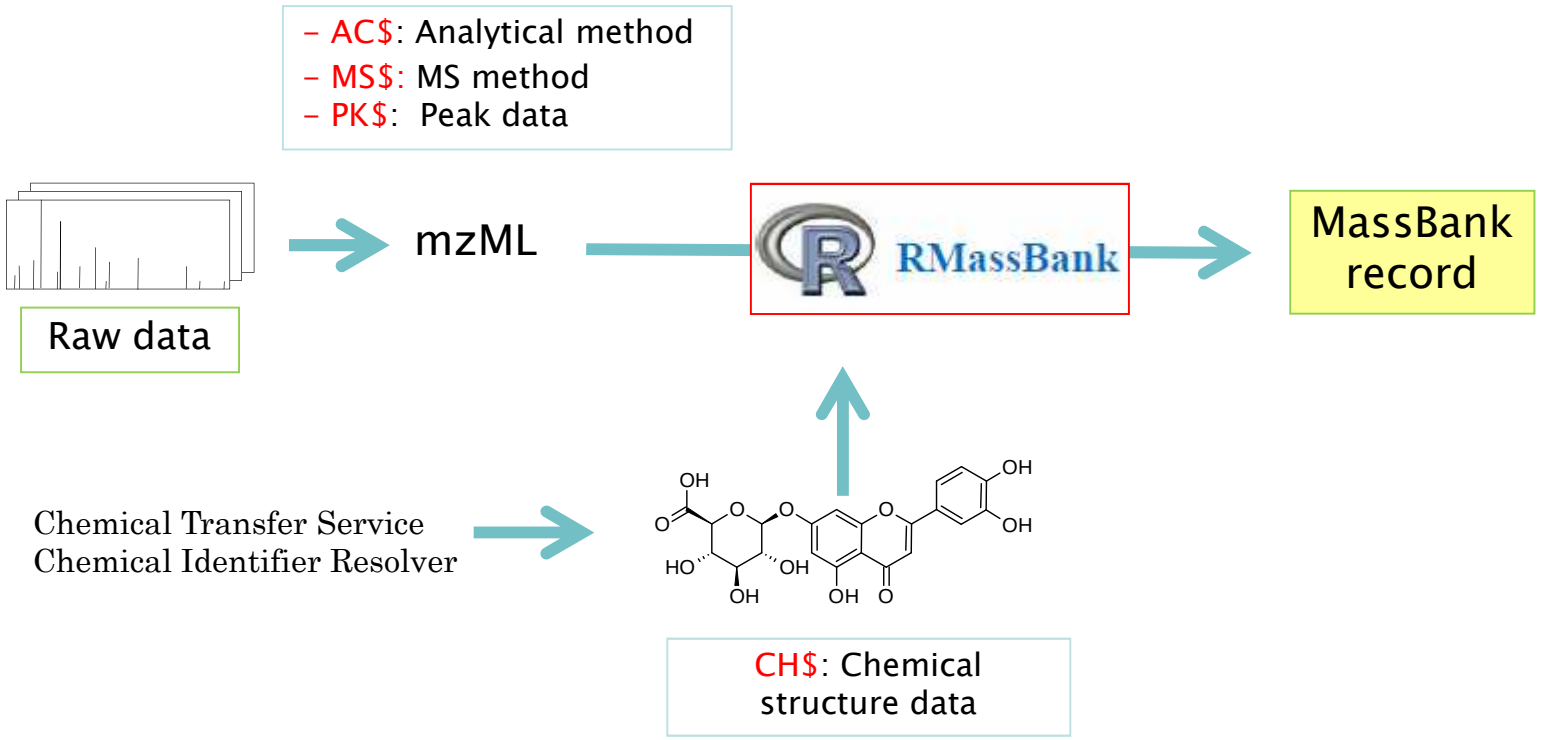


Molfile

Internet

・Contributor's manuals & workshop

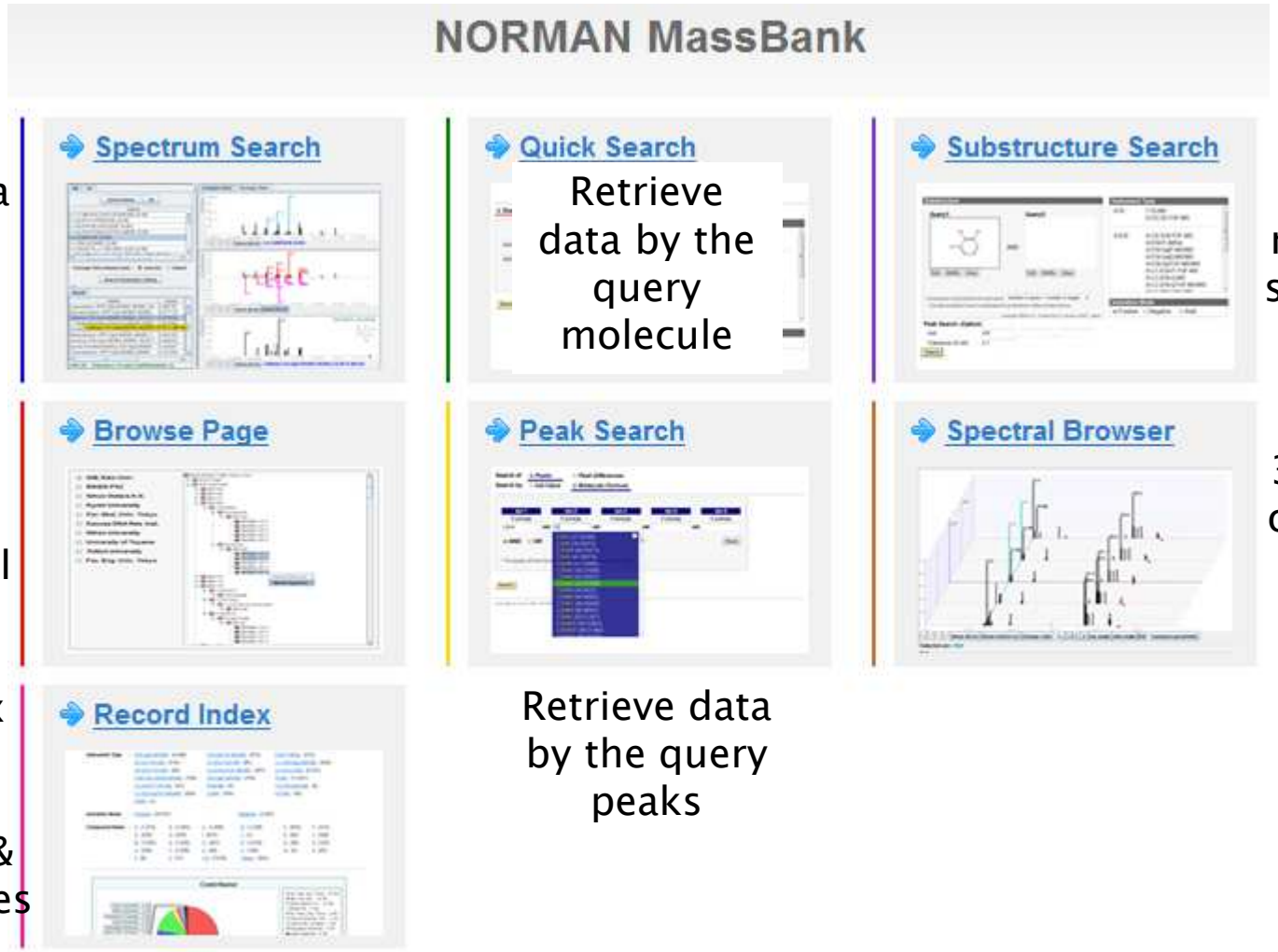
MassBank Record Preparation Tool: "RMassBank"



RMassBank

http://bioconductor.org/packages/devel/bioc/html/RMassBank.html

RMassBank was developed by Dr. Emma Schymanski (EAWAG).



Search data similar to the query data

Retrieve data by the query molecule

Retrieve molecules by substructures

Browsing data by a hierarchical display

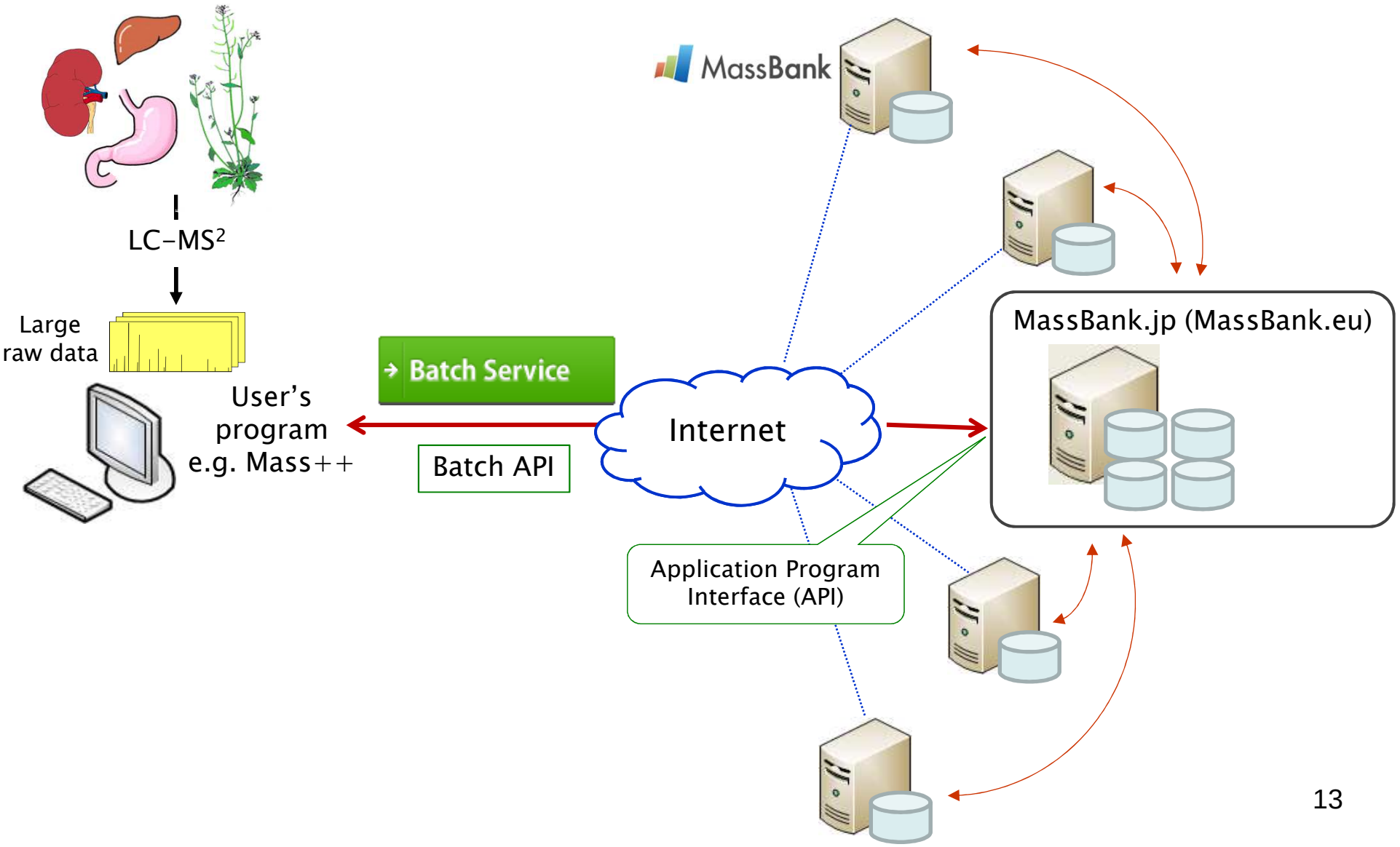
3D display of multiple spectra

Record index classified by contributors, instruments & chemical names

Retrieve data by the query peaks

Application program interface (API)

Batch Spectral Search

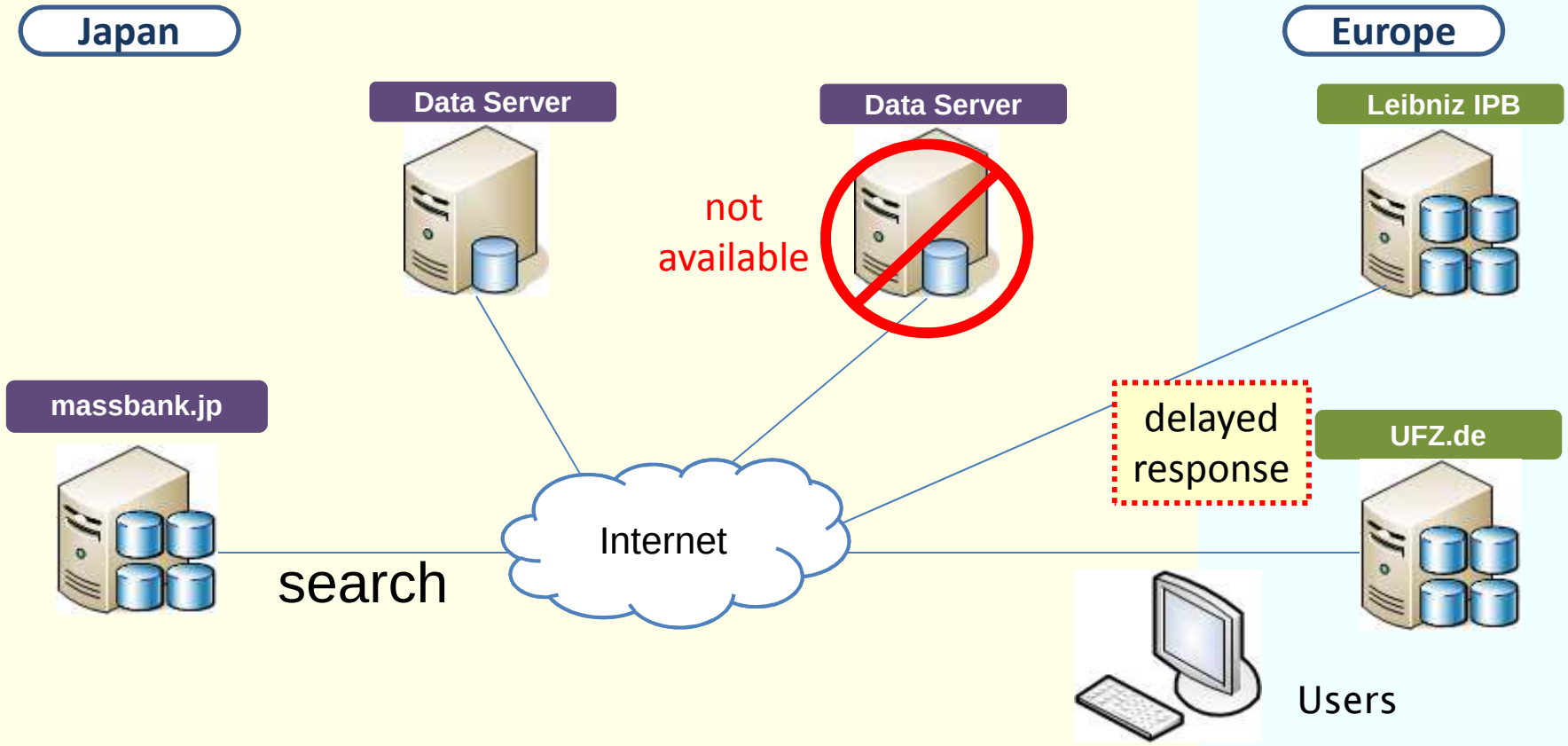


Setup of Mirror Server in Europe

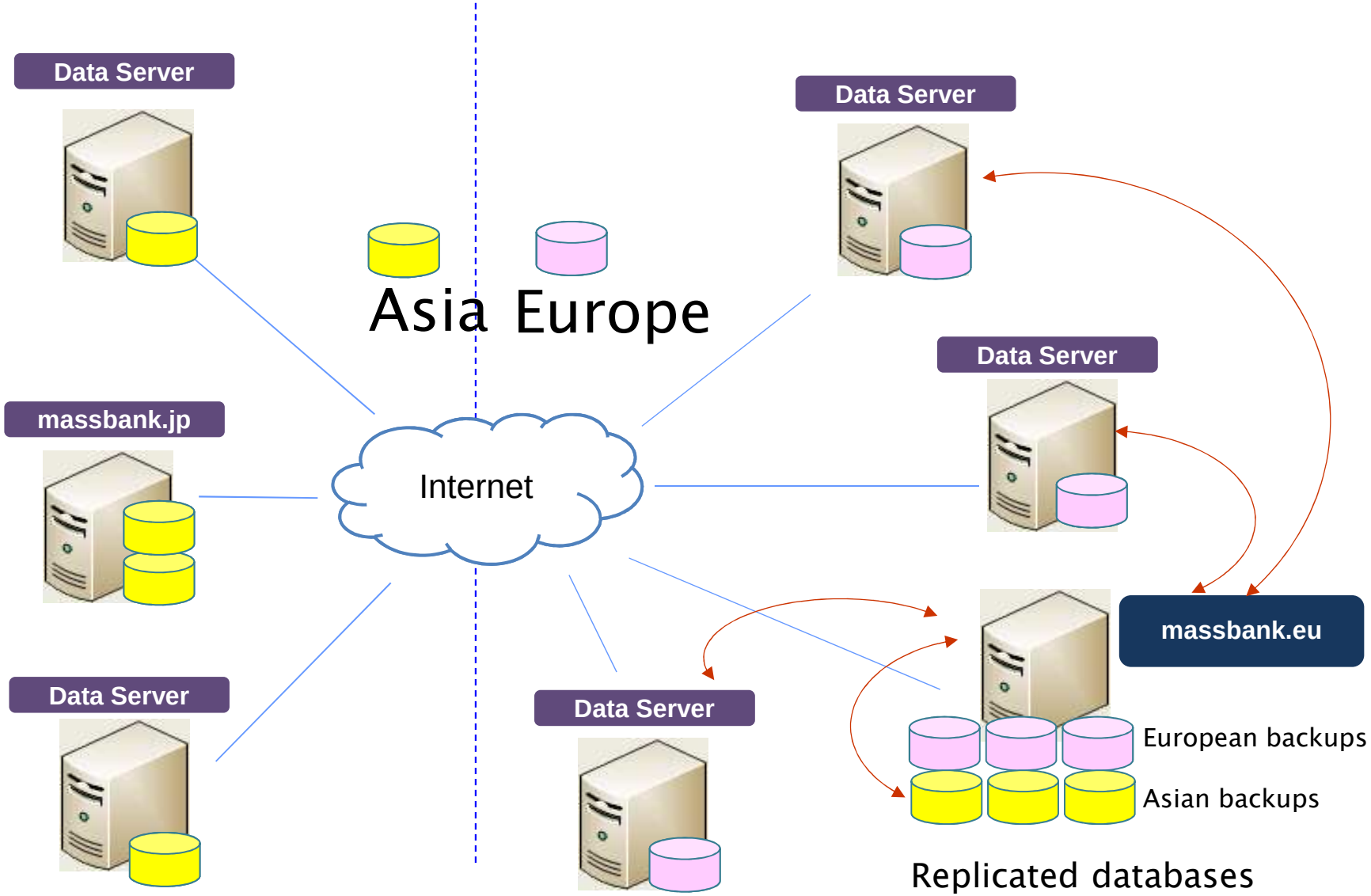
Distributed Database has Two Problems

1. Transcontinental accesses to MassBank.jp from European users are slow in the responses.

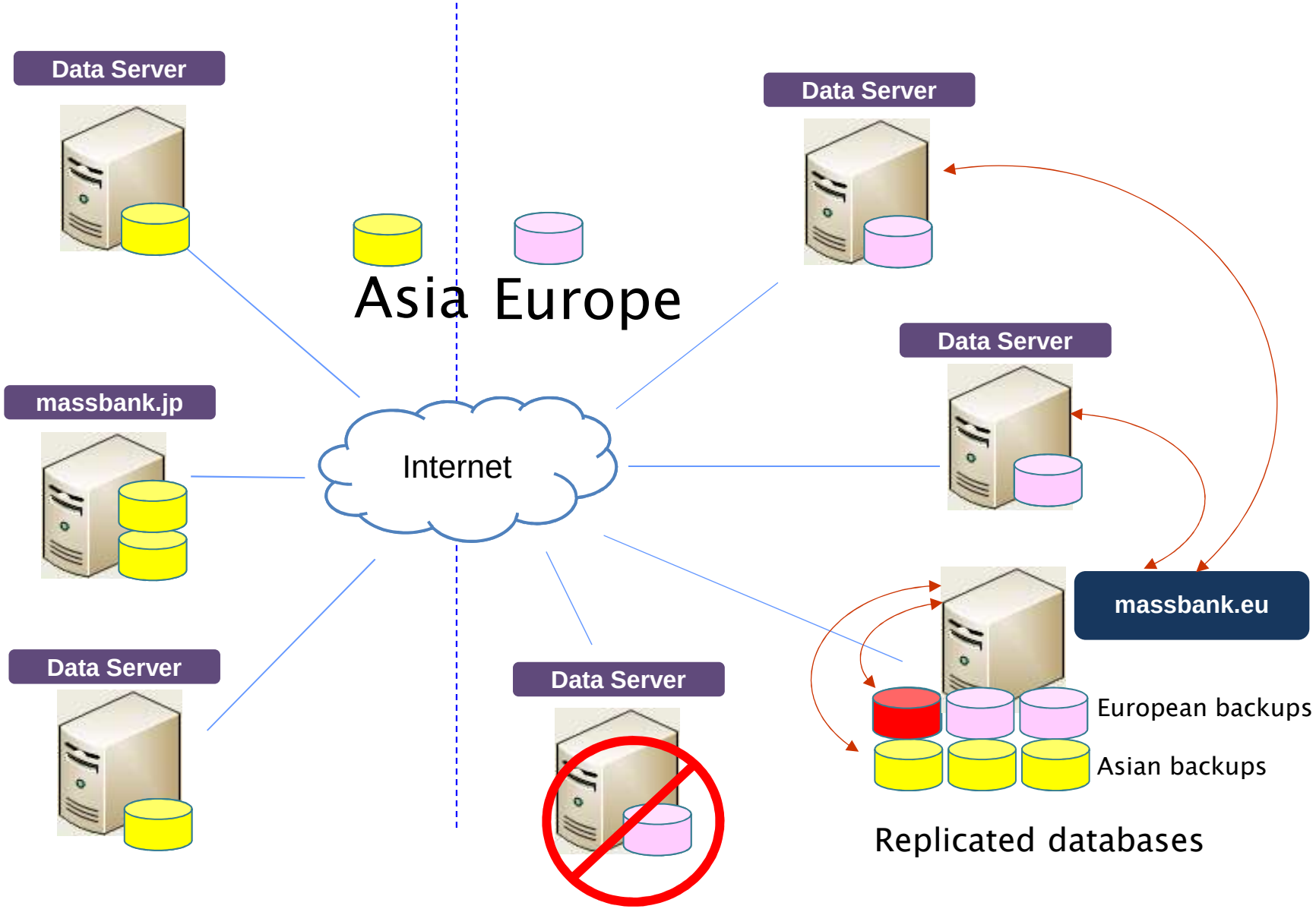
2. When one of the data servers is unavailable, it affects the search results significantly.



1. Mirror Server in UFZ



2. Backup Databases



From "*m/z*" to "chemical formula"

"Chemically Exact" ESI-MS^2 data

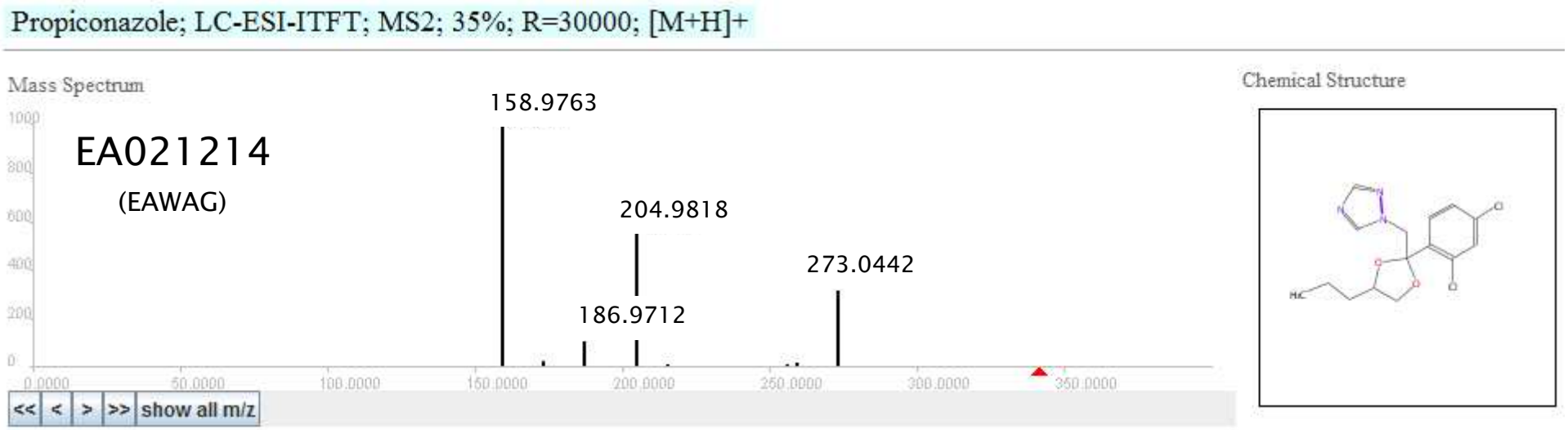
Accurate Data Increase Rapidly

- Accurate reference MS data rapidly increase in MassBank.

3,251	ESI-Q/TOF-MS^2 data	Metabolites	
4,504	ESI-IT/FT-MS^2 data	General chemicals	
253	ESI-IT/TOF-MS^2 data	Saponins	
45	ESI-TOF/TOF-MS^2 data	Metabolites	(August 2012)

MassBank provide MS data that are chemically annotated with **molecular formula**.
"RMASSBANK" has this functionality.

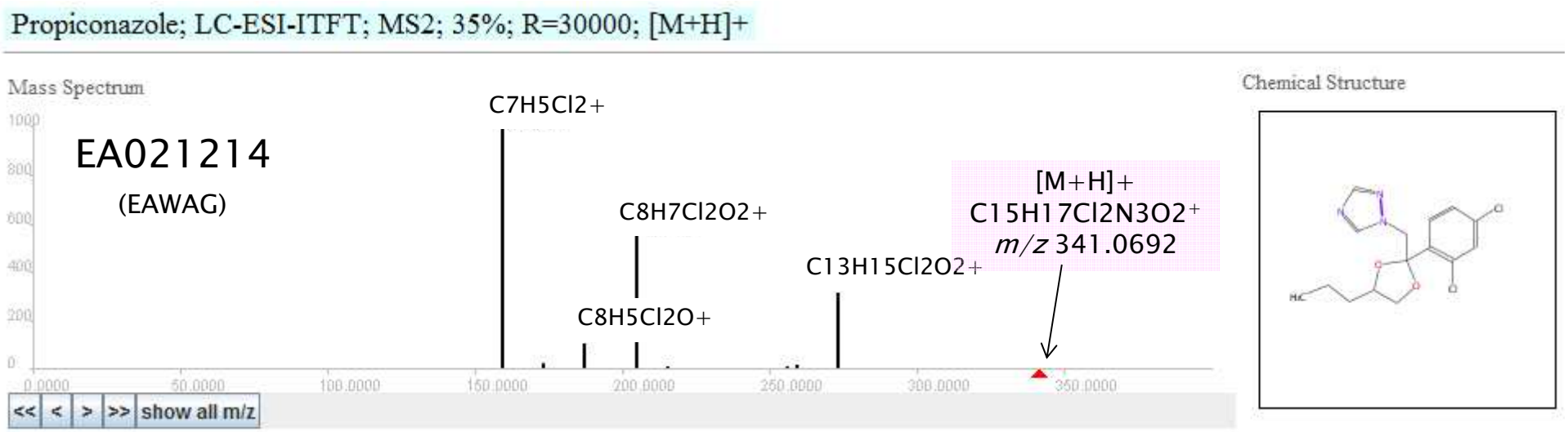
Accurate ESI-MS^2 Data




```
PK$PEAK: m/z int. rel.int.
  158.9763 1082088.9 999
  172.9558 3258.9 3
  172.9918 27399.4 25
  186.9712 113116.7 104
  204.9818 594435.4 548
  215.0387 11294 10
  256.0034 3781.4 3
  259.0284 17049.6 15
  273.0442 341509.9 315
```

Conventional presentation of peaks (*m/z*)

Chemically Exact ESI-MS^2 Data




```
PK$ANNOTATION: m/z num {formula mass error(ppm)}
  158.9763 1 C7H5Cl2+ 158.9763 0.37
  172.9558 1 C7H3Cl2O+ 172.9555 1.23
  172.9918 1 C8H7Cl2+ 172.9919 -0.65
  186.9712 1 C8H5Cl2O+ 186.9712 -0.2
  204.9818 1 C8H7Cl2O2+ 204.9818 0.24
  215.0387 1 C11H13Cl2+ 215.0389 -0.71
  256.0034 1 C10H8Cl2N3O+ 256.0039 -1.85
  259.0284 1 C12H13Cl2O2+ 259.0287 -1.13
  273.0442 1 C13H15Cl2O2+ 273.0444 -0.77
```

Chemical annotation of peaks (by formula)

- quality check
- merged reference data
- peak-substructure relationships

Acknowledgement

“Special thanks to all the contributors!”

MassBank team

Yoshito Nihei (NAIST)
Tasuku Ikeda
Yuya Ojima
Rie Matsuzawa
Shigehiko Kanaya
Masanori Arita (U Tokyo)
Hisayuki Horai (Ibaraki CT)

Mass++ team

Satoshi Tanaka (FIRST program)
Shigeki Kajihara
Ken Aoshima (Eisai Co Ltd)
Yoshiya Oda

MassBank Consortium

Steffen Neumann (IPB, Halle)
Emma Schymanski (EAWAG)
Tobias Schulze (UFZ)

Grants

- National Bioscience Database Center, JST (2011-2013)

NBDC
National Bioscience Database Center

- BIRD, JST (2006-2010)
- Yamagata Grant to IAB, Keio (2006-2010)