

1H BT transformation scheme used with permission from Huntscha et al. 2014, DOI: 10.1021/es405694z, Copyright (2014) American Chemical Society

www.umweltprobenbank.de

Perspectives of repositories for environmental mass spectral and metadata for non-target and retrospective analysis

Tobias Schulze, Martin Krauss, Werner Brack
 UFZ Helmholtz Centre for Environmental Research, Leipzig, Germany
 Oslo, 02 Mar 2016



Introduction

- Scientists, companies and agencies hold a vast amount of high quality mass spectral information related to environmental samples
- Larger parts of data usually neither used nor shared with others
- No share may be related to privacy issues or unawareness of Open Data benefits
- Missing data repositories for raw mass spectral files, derived peak lists and associated meta data

Implications for digital 'specimen' repositories

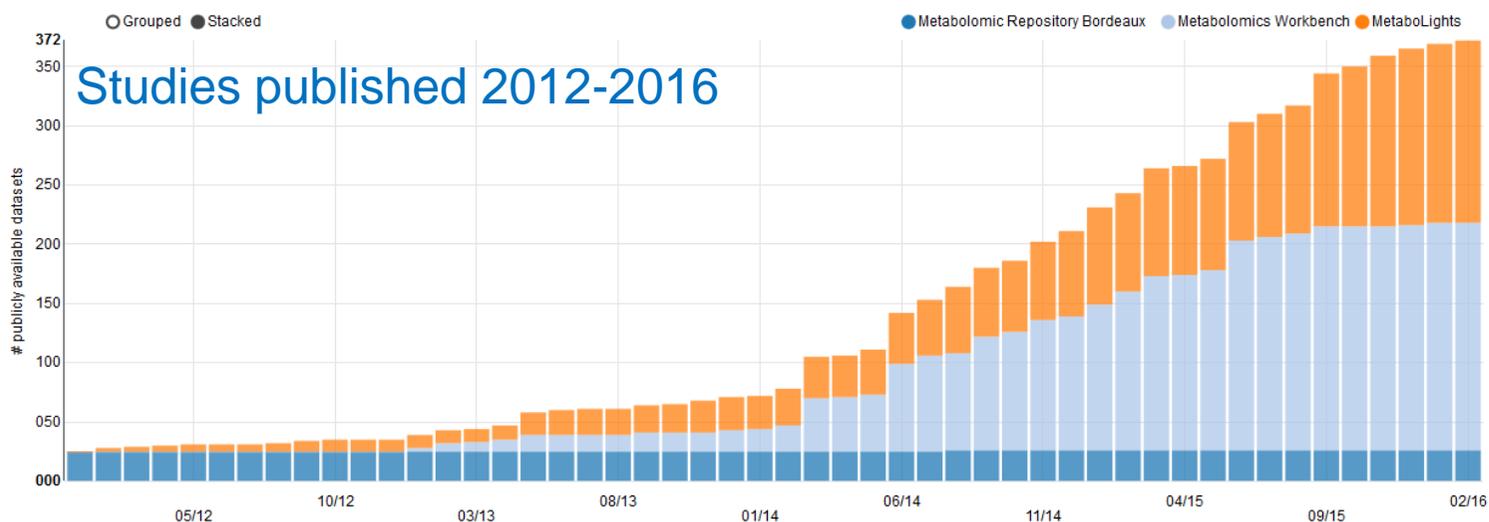
- **Community building** (good examples, social media, workshops)
- **Data quality** (reliability, instruments and software settings)
- **Standardisation** (methods, vocabularies / ontologies)
- **Curation** (data quality, enrichment of meta data)
- **Data quantity** (handling of large data files)
- **User-interfaces** (easy discovery, access and re-use of data)
- **Machine-interfaces** (retrieval of data for automated workflows)
- **Privacy** (anonymisation, secured data services for internal use)
- **Long-term availability and referencing** (e.g. DOI deployment)

DOI: digital object identifier



Learning from metabolomics community

High activity in release, publication, announcement and exchange of publicly available metabolomics datasets that may be useful for other researchers



Standards in metabolomics



cosmos

<http://cosmos-fp7.eu/>



<http://metabolomexchange.org>

Example: MetaboLights

The screenshot shows the MetaboLights website. At the top left is the MetaboLights logo, which consists of a blue bar chart icon followed by the text 'MetaboLights'. To the right of the logo is a search bar with a 'Search' button. Below the search bar, there are examples of search terms: 'alanine, human, urine, MTBLS1'. A navigation menu is located below the search bar, containing links for 'Home', 'Browse Studies', 'Browse Compounds', 'Browse Species', 'Analysis', 'Download', 'Help', 'Give us feedback', and 'About'. On the right side of the navigation menu, there are links for 'Submit Study' and 'Login'. The main content area is divided into three columns. The first column is titled 'MetaboLights' and contains a paragraph describing the database as a repository for metabolomics experiments and derived information, covering cross-species, cross-technique data. The second column is titled 'Download' and contains two sections: 'Pre-packaged ISAcreator download' and 'Experiments'. The third column is titled 'Tweets' and displays three tweets from the MetaboLights account (@MetaboLights) regarding new methods and studies.

MetaboLights

MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments. [About MetaboLights](#).

Download

Pre-packaged ISAcreator download. To make it easy for new users, please download and just unzip our pre-packaged ISAcreator with plugin and configurations.

Experiments. All public MetaboLights experiments can be downloaded from our public [ftp archive](#). Please find zip archives under the "studies" folder. Each public study can be found in the corresponding MTBLS-id folder. Complete experiments can be opened with [ISAcreator](#) or you can extract the archives using your normal unzip program.

Tweets

MetaboLights @MetaboLights 4 Jan
MTBLS270: A new method for quantitative determination of bitter components of Belgian endi... [ebi.ac.uk/metabolights/M...](#)

MetaboLights @MetaboLights 31 Dec
MTBLS168: Metabolism triggers system-wide epistasis on transcriptome, proteome and metabol... [ebi.ac.uk/metabolights/M...](#)
Expand

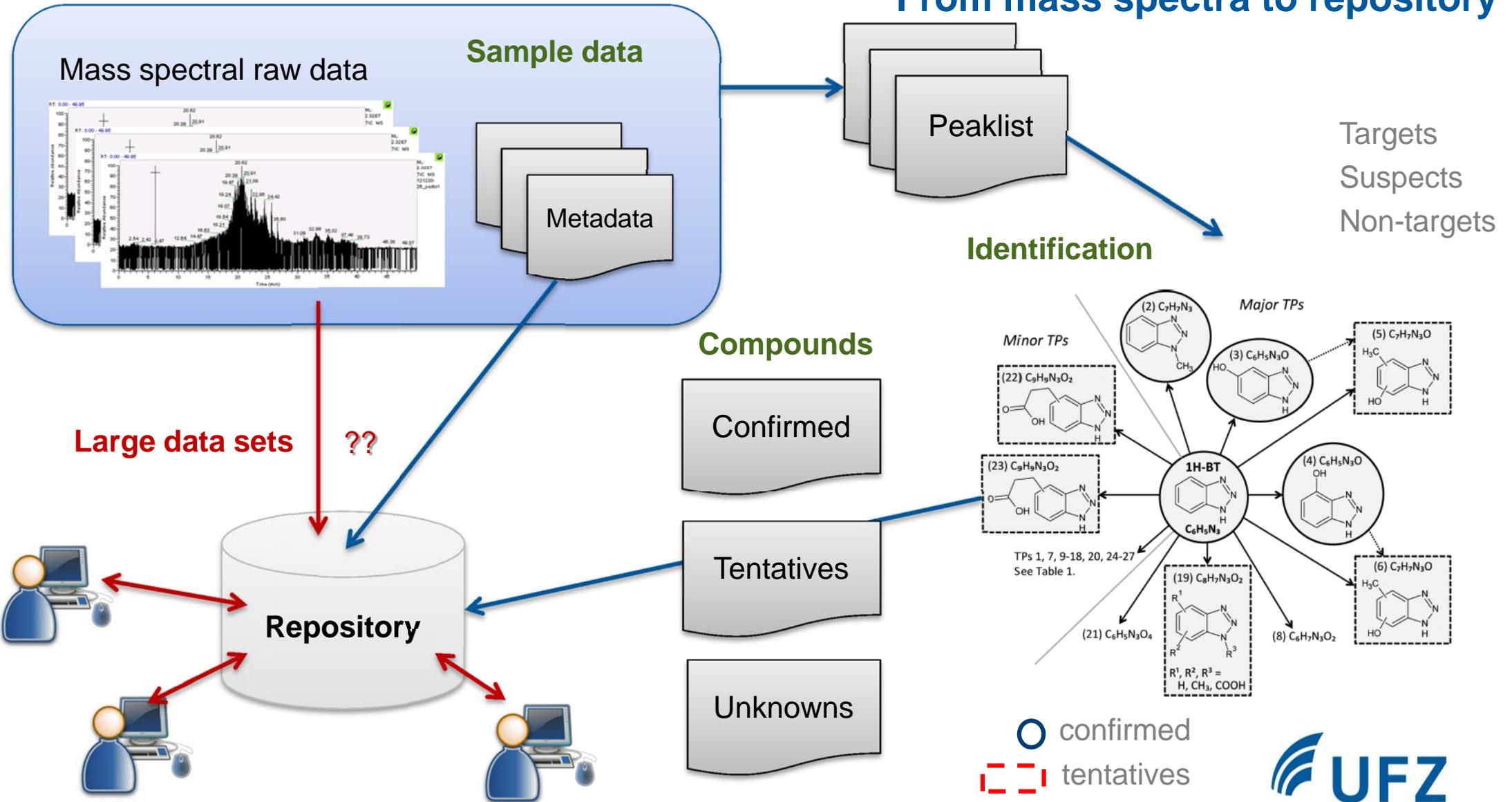
MetaboLights @MetaboLights 17 Nov
MTBLS173: Comprehensive Metabolic Profiling of

Tweet to @MetaboLights

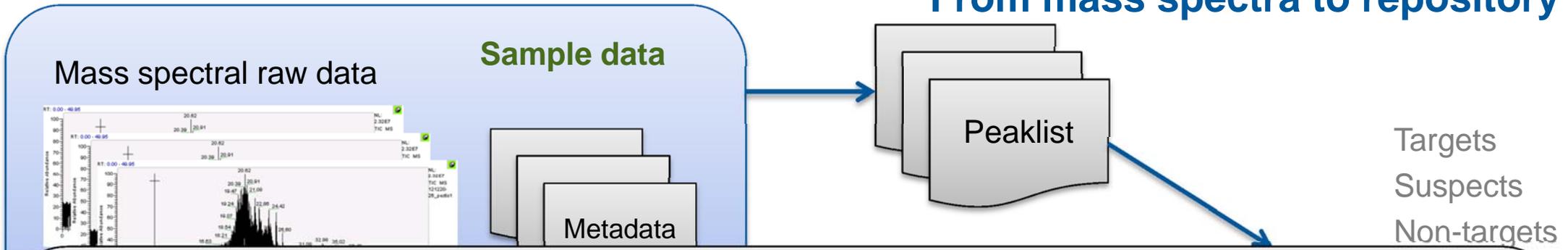
- Database and repository for metabolomics experiments and derived information
- Presents meta data on studies, experiments, identified metabolites and MS and NMR data

<http://www.ebi.ac.uk/metabolights/> NMR: Nuclear Magnetic Resonance Spectroscopy

From mass spectra to repository

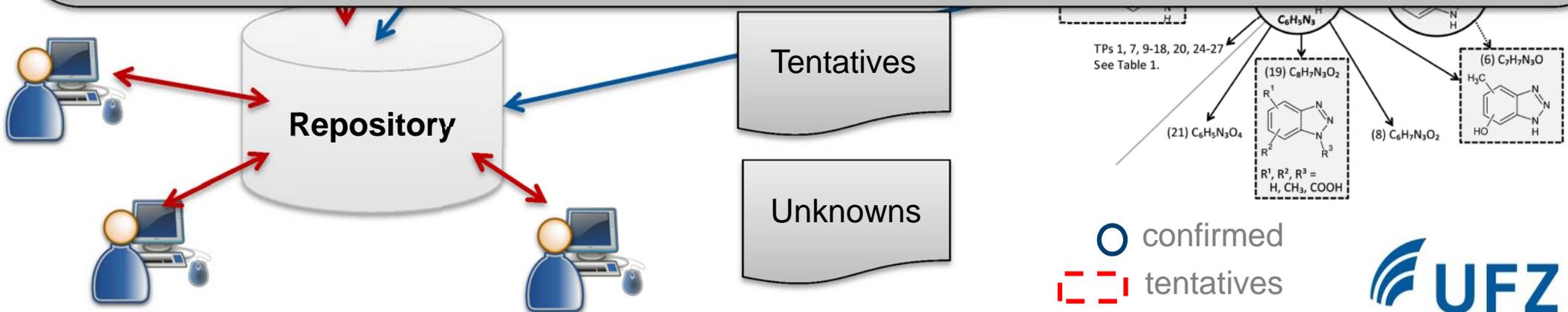


From mass spectra to repository

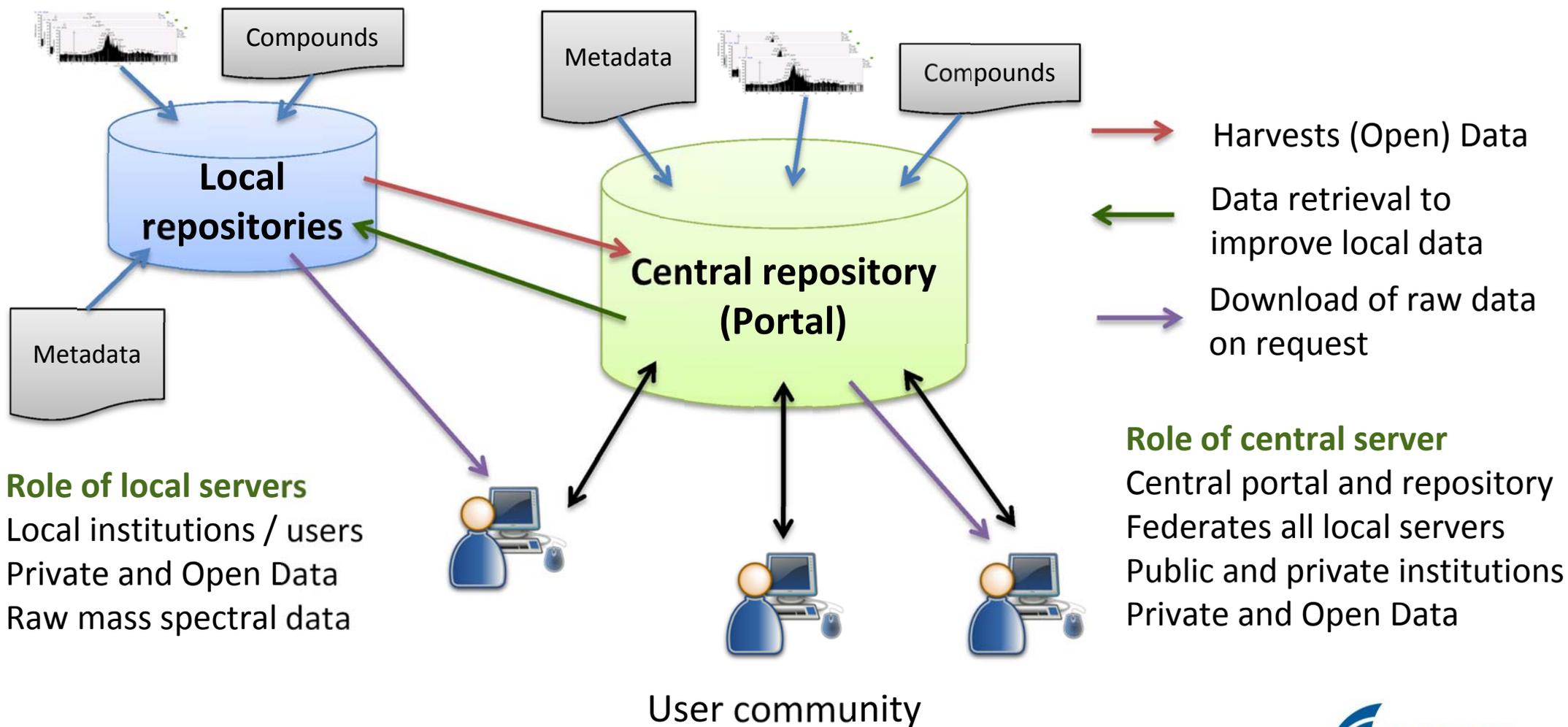


How to handle large datasets?

- 1) Large MS raw data (e.g. 400-600 MB for one run at Q Exactive Orbitrap in full scan / AIF mode)
- 2) No one has the capacities to store large data of many researchers on in kind basis, even upload is a challenge



Solution: Building a distributed system – NormanNTXchange



Demonstration of NormanNTXchange mockup

Data portal based on ckan

The screenshot displays a web interface for NormanNTXchange, a data portal based on CKAN. The top navigation bar includes the Norman logo, menu items for Datasets, Organizations, Groups, and About, and a search field. The main content area is divided into several sections:

- Search data:** A search bar with the placeholder text "E.g. environment" and a search icon. Below it are "Popular tags" for "target", "suspect", and "non-target".
- NormanNTXchange statistics:** A section showing counts for "2 datasets", "1 organization", "0 groups", and "2 related items".
- NORMAN Non-Target Group:** A section featuring a dataset titled "Joint Danube Survey 2013 - LVSPE - Non-Target Screening" with the note "This dataset has no description". Below it is another dataset titled "NORMAN Collaborative Trial Targets and Suspects" with a description: "As part of a series of workshops in September 2014, NORMAN members expressed the need to exchange various lists of...".
- Left sidebar:** A "Datasets" section with a home icon. It lists "Organizations" (NORMAN Non-Target G... (2)), "Groups" (with a note "There are no Groups that match this search"), and "Tags" (target (1), suspect (1), non-target (1), collaborative trial (1), NORMAN (1), LVSPE (1), LTQ Orbitrap XL (1), JDS3 (1)). A red arrow points from the "NORMAN Collaborative Trial Targets and Suspects" dataset in the main content to the "NORMAN (1)" tag in the sidebar.
- Right sidebar:** An "Add Dataset" button, a search bar for datasets, and a section titled "2 datasets found" with an "Order by: Relevance" dropdown. It lists the same two datasets as the main content, with file format icons for CSV and mzML.

<https://ckan.org>

Demonstration of NormanNTXchange mockup

Joint Danube Survey 2013 - LVSPE - Non-Target Screening

Data and Resources

Examples of exported peak lists derived with ...

- JDS3_LVSPE_JDS27_Hercegszanto_FTMS_full_MS_ESIpos...
Joint Dabube Survey 3 - Large volume solid phase extraction sample JDS27 full...
- JDS3_LVSPE_JDS41_Melika_Morava_FTMS_full_MS_ESIpos...
Joint Dabube Survey 3 - Large volume solid phase extraction sample JDS41 full...
- JDS3_LVSPE_JDS59_Downstream_Arges_FTMS_full_MS_ESI...
Joint Dabube Survey 3 - Large volume solid phase extraction sample JDS59 full...
- JDS3_LVSPE_JDS67_Sulina_FTMS_full_MS_ESIpos_non-ta...
Joint Dabube Survey 3 - Large volume solid phase extraction sample JDS67 full...

JDS3 LTQ Orbitrap XL LVSPE non-target

Examples of exported peak lists

derived with ...

Example of a peak list stored in the database

URL: <http://ckan.io/dataset/735df029-7765-41ff-8f9f-bbec4ae8dc8b/resource/b50e653c-aac1-4605-babe-e404e49821a7/download/jds3ex...>

Data Explorer DETECTED Peaks DETECTED Peaks in Sulina

Embed

Add Filter

Grid Graph Map 12656 records 1 - 100

Search data ... Go Filters

_id	Sample	Locatio...	Latitude	Longitu...	Peak st...	Peak m/...	Peak RT	Peak RT...	Peak RT...	Peak he...
1	140521-...	Sulina	45.1945	28.95933	DETECT...	315.171...	25.8683...	25.7276...	27.1642...	7672347...
2	140521-...	Sulina	45.1945	28.95933	DETECT...	114.090...	11.64345	10.8438...	13.2718	1650489...
3	140521-...	Sulina	45.1945	28.95933	DETECT...	279.159...	2.351645	1.42360...	3.51280...	883025...
4	140521-...	Sulina	45.1945	28.95933	UNKNOWN	N/A	N/A	N/A	N/A	N/A
5	140521-...	Sulina	45.1945	28.95933	UNKNOWN	N/A	N/A	N/A	N/A	N/A
6	140521-...	Sulina	45.1945	28.95933	DETECT...	353.266...	29.8947...	29.4744...	30.6222...	1467252...
7	140521-...	Sulina	45.1945	28.95933	DETECT...	329.005...	2.43652...	1.45144...	3.51280...	530962...
8	140521-...	Sulina	45.1945	28.95933	UNKNOWN	N/A	N/A	N/A	N/A	N/A
9	140521-...	Sulina	45.1945	28.95933	UNKNOWN	N/A	N/A	N/A	N/A	N/A

Download page for raw mass spectral file

UFZ DATAINVESTIGATIONPORTAL Keywords ... DE | EN | Imp

back to search results

140521-14_JDS3_LVSPE_JDS27_Hercegszanto_FTMS_full_MS_ESIpos_non-target.zip

Project WANA sample data
Description Joint Dabube Survey 3 - Large volume solid phase extraction sample JDS27 full scan mass spectra ESI positive LTQ Orbitrap XL
Size 52.0 MB
Dataquality quality assured data



Metadata Catalogues

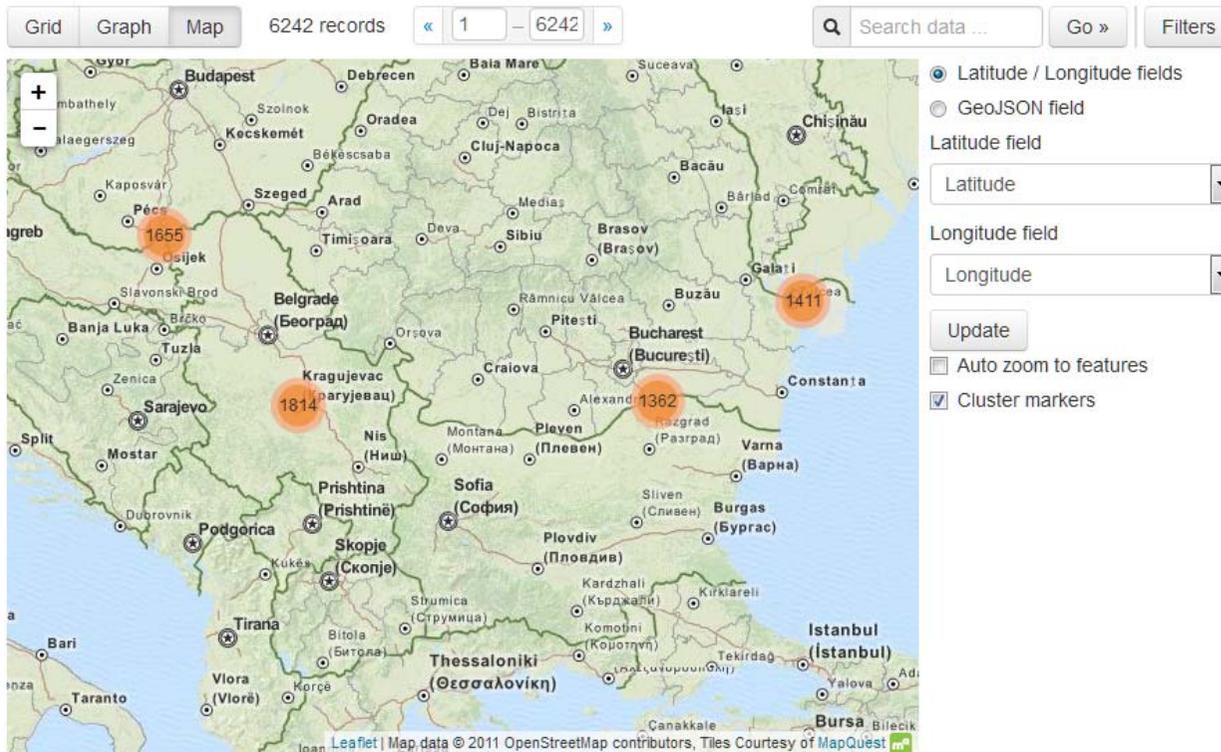
Request Data for Download

Contact <https://www.ufz.de/record/dmp/archive/2853/en/>

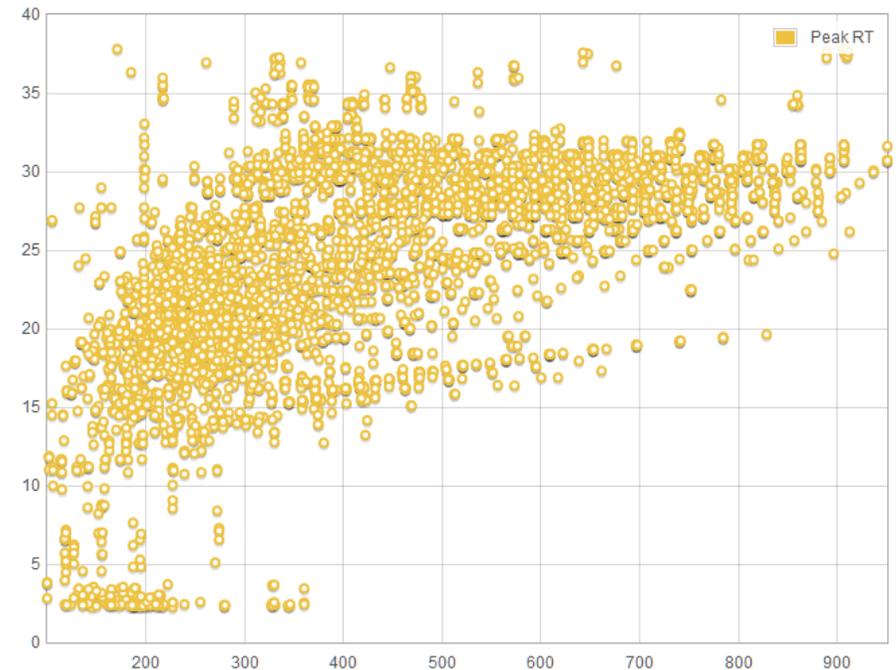


Demonstration of NormanNTXchange mockup

Map view of the detected non-target peaks



Graphical view of the not-target peaks (m/z vs. retention time)



Conclusions

- Archiving of environmental MS data is achievable
- A distributed and federated system shares costs
- Privacy issues are addressed and guaranteed
- Common standards need to be developed and committed
- Curation ensures data quality and value adds data
- Open Data enhances awareness and builds a community



"Opened up a Pandora's box" by F.S. Church (<http://en.wikipedia.org>)

Acknowledgements

- Erik Müller (UFZ) and Steffen Neumann (IPB) for help with the ckan mockup
- Martin Krauss (UFZ) for providing example data sets
- Contributors to the NORMAN NT collaborative trial
- Werner Brack (UFZ) for supporting the activities
- NORMAN for funding and supporting
- And you for your attention!

