Stealing ideas

from metabolomics data exchange repositories

Steffen Neumann

Leibniz Institute of Plant Biochemistry

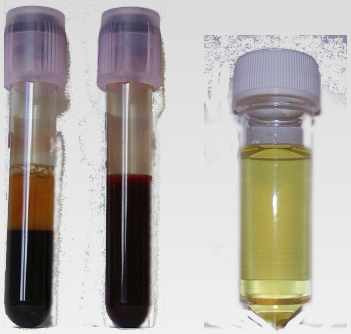
Mass Spectrometry and Bioinformatics group

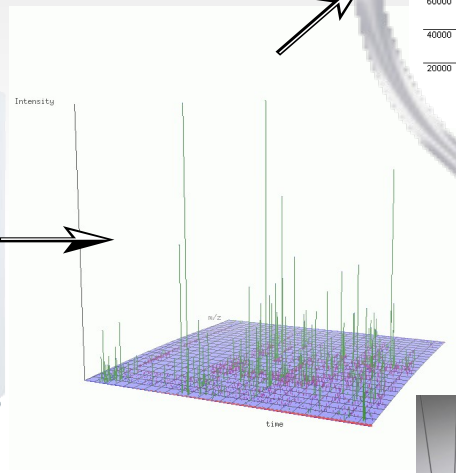
sneumann@ipb-halle.de

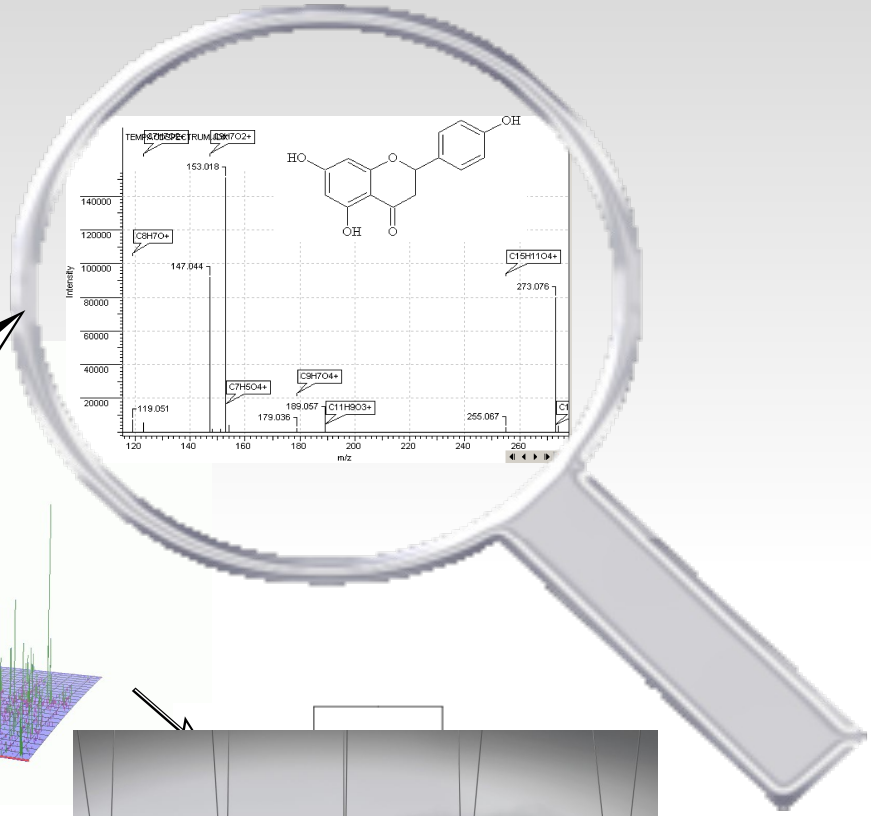
Outline

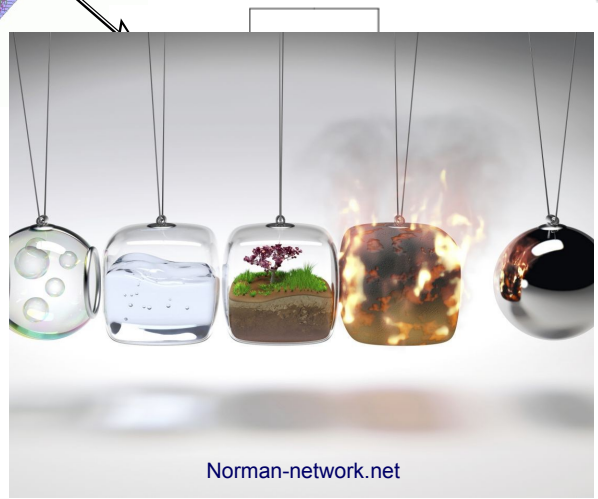
- Why would you want to care about metabolomics?
- Sharing metabolite profiling data
- Data standards in metabolomics
- Sharing reference spectra

Metabolomics @ IPB










WELCOME TO
NORMAN
EST. 1889
wikipedia.org/wiki/Norman,_Oklahoma

Publications: Good Practice

- Provide Information in Supplemental Material
  - But: good looking colourful Excel doesn't help Bioinformaticians
  - Data in complex tables in PDF is even worse
- Better: submit data upstream, any data repository and provide accession numbers!

Not quite there yet:

- DOI's for Data Sets
- DataCite: publication of Data Sets

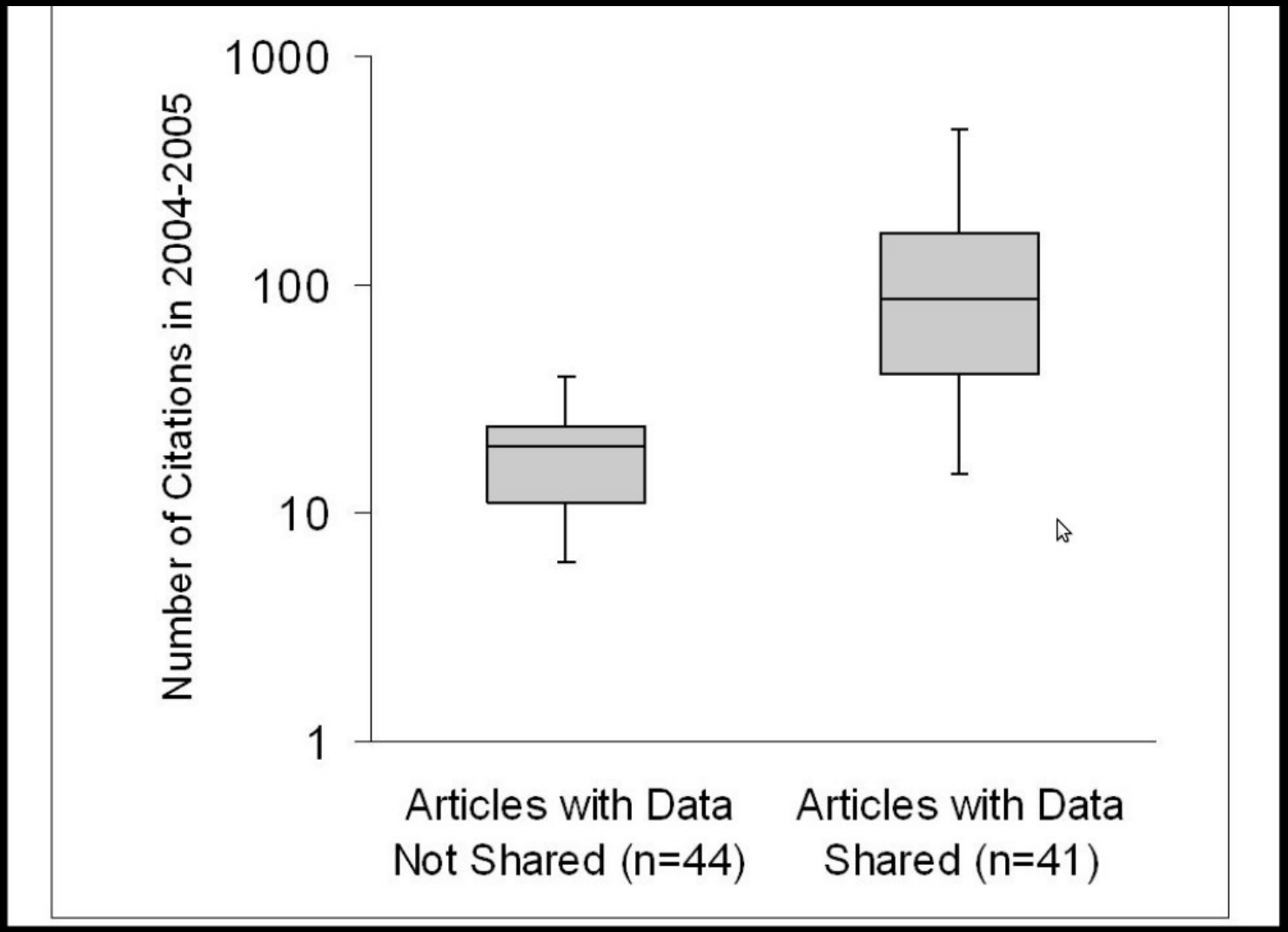
OPEN ACCESS Freely available online

PLoS one

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

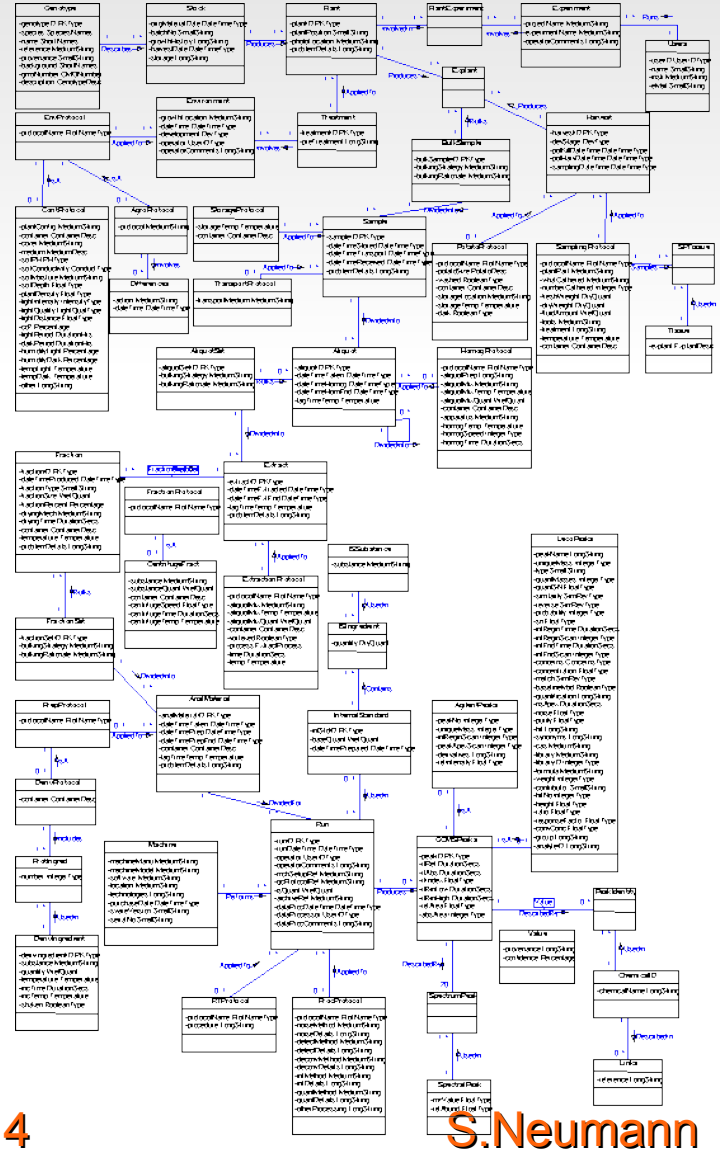
Note: log scale



http://www.slideshare.net/hpiwowar/research-into-open-research-data

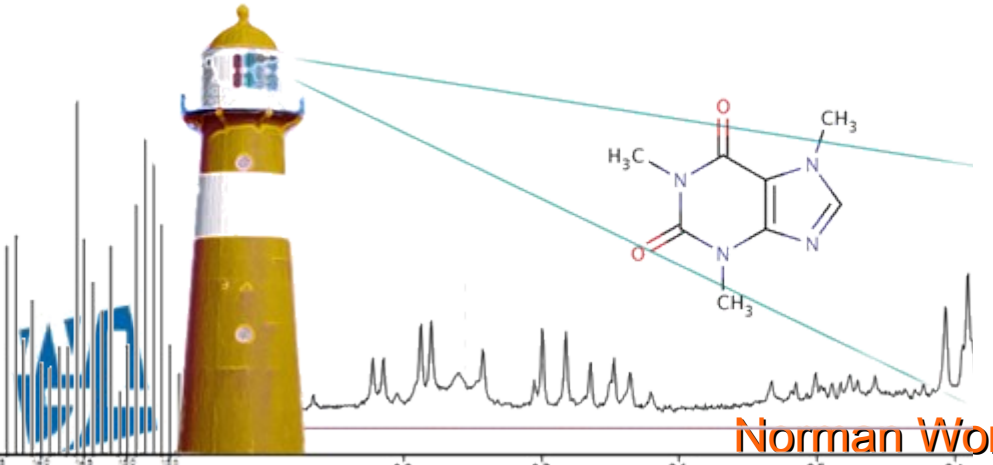
ArMET (2004)

- One of several attempts designing biology/metabolomics databases
- Architecture for METabolomics
  Nigel Hardy, Helen Jenkins
- Very detailed, not so user friendly,
  → didn't take over the world
- Other examples exist …



MetaboLights@EBI

- EBI is a major European life-science resource
- Repository of MS/NMR Metabolomics Data
- Think: Supplemental Experimental Data
- Public data is frozen



Steffen Neumann, Emmanuel Gaquerel,

Study Submission Date: 01-Feb-2012 Study Public Release Date: 02-Feb-2012

characterize changes induced in Nicotiana attenuata leaves 1 h and 5 days after wounding and application of Man
software, we extracted 367 buckets, which were analyzed by principal component analysis and two-factorial ANO
found to be statistically regulated, 128 due to time effects, and 85 due to treatment effects.

Study Design Description | Protocols | Data | Metabolite Identification

Protocol	Description
Sample collection	We used an isogenic line, obtained after 30 generations of inbreeding, of Nicotiana attenuat germinated as described in Krügel et al. (15). All plants were grown in the glasshouse in 1 L light supplied by Philips Sun-T Agro 400- or 600-W sodium lights (Philips, Turnhout, Belgium) Manduca sexta feeding were reproduced by producing with a fabric pattern wheel three row midvein of five fully expanded leaves per plant (5 biological replicates) and directly applyin (OS). Treated leaves from the same plant were harvested, pooled, and flash frozen 1 h and 5 were left unwounded and harvested from other plants at the same time points.
Extraction	One hundred milligrams of ground leaf tissue was weighted and transferred to a Fast Prep tu (BIO 101, Vista, USA). One milliliter of extraction buffer per 100 mg of tissue [50 mM acetate spiked with reserpine (600 ng/mL), atropine (200 ng/mL)] was added, and the samples were rpm, 20 min, 4 °C), the supernatant was collected in a fresh 1.5 mL Eppendorf tube and cen supernatant was transferred to a HPLC vial.
Chromatography	Two microliters of the leaf extract were separated using a HPLC 1100 Series system (Agilent 150 mm Â 2 mm i.d., 3 µm, Phenomenex Gemini NX RP-18 column with a 2 mm Â 4 mm i.d. g (Phenomenex, Germany). The following binary gradient was applied: 0 to 2 min isocratic 95% nitrile [Baker, HPLC grade], and 0.05% formic acid), 5% B (acetonitrile and 0.05% formic acid isocratic for 5 min. The flow rate was 200 µL/min.
Mass spectrometry	Eluted compounds were detected by a MicroToF mass spectrometer (Bruker Daltonik, Breme electrospray ionization source in positive and negative ion modes. Typical instrument setting V; capillary exit, 130 V; dry gas temperature, 200 °C; dry gas flow, 8 L/min. Ions were detec rate of 1 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution containing 0.2% formic acid).

What is EMBL-EBI?

European Bioinformatics Institute

- Part of the
  European Molecular Biology Laboratory
  (Happy Birthday!)
- “Genome Campus” near Cambridge
- EU funded non-profit research institute,
  Europe’s hub for biological data services
  570 staff from 53 nations.
- Capacity ~40PB$_{(08/2014)}$, mirrored data centers
- “Hardware capacity doubles every
  18 months, but biological data
  doubles every 9 months.”

MetaboLights – Logical components

- Data Submission
  - ISAcreator (curation)
  - Online data deposition

- Repository
  - Complete metabolomics experiments/studies
  - Sharing Experiments

- Metabolite Reference Data
  - Metabolite identification
  - Curated

Data Capture Template

Retention time [min]	Mass of ion [m/z] (peak or component)	Intensity of the ion	Intensity of the ion in the blank	Ion type	Other	MS/MS available	Category	Proposed identification (name of the substance or n.i. for not identified)	Molecular formula	Identifier: SMILE
YOUR DATA		YOUR DATA						YOUR DATA		YOU
5.97	134.0711981661	1.81E+09	2.37E+07	M+H	positive	Yes	target	4&5-Methyl-Benzotriazole_167_ p	C7H7N3	c12c(cc(C)cc2)nn[nH]1
4.69	120.0556107233	1.15E+09	1.39E+06	M+H	positive	Yes	target	Benzotriazole_166_ p	C6H5N3	c12c(nn[nH]1)cccc2
3.93	246.1237932546	9.35E+08	0.00E+00	M+H	positive	Yes	target	4-Acetamidoantipyrine (N-Acetyl-4-Am	C13H15N3O2	CC(=O)NC=2C(=O)N(c1cc
3.87	232.1081207512	5.82E+08	0.00E+00	M+H	positive	Yes	target	4-Formylaminoantipyrine (4-FAA)_1038	C12H13N3O2	O=C2C(\NC=O)=C(\C)N(C)
7.93	237.1022905227	3.74E+08	2.38E+06	M+H	positive	Yes	target	Carbamazepine_194_ p	C15H12N2O1	N1(c2c(cccc2)C=Cc2c1ccc
4.09	195.0876061974	3.46E+08	6.01E+06	M+H	positive	Yes	target	Caffeine_303_ p	C8H10N4O2	c12c(n(c(=O)n(c1=O)C)C)r
8.80	192.1382565761	2.50E+08	4.80E+06	M+H	positive	Yes	target	N-N-diethyl-3-methylbenzamide (DEET	C12H17N1O1	c1(C(N(CC)CC)=O)cc(C)ccc
5.10	264.1958593355	1.69E+08	0.00E+00	M+H	positive	Yes	target	Tramadol_2567_ p	C16H25N1O2	COC1=CC=CC(=C1)[C@@
10.40	230.1167174068	1.53E+08	0.00E+00	M+H	positive	Yes	target	Terbutylazine_284_ p	C9H16Cl1N5	c1(nc(nc(n1)Cl)NCC)NC(C)(
4.55	198.1348923768	1.38E+08	0.00E+00	M+H	positive	Yes	target	Atrazin-2-Hydroxy_279_ p	C8H15N5O1	c1(nc(nc(n1)O)NCC)NC(C)C
11.26	436.2341888099	1.25E+08	0.00E+00	M+H	positive	Yes	target	Valsartan_2583_ p	C24H29N5O3	CCCCC(=O)N(CC1=CC=C(
4.79	189.1022584029	1.20E+08	0.00E+00	M+H	positive	Yes	target	Phenazon (Antipyrin- ID 2519)_338_ p	C11H12N2O1	CN1N(C(=O)C=C1C)C2=C(
5.49	256.0151400017	1.10E+08	0.00E+00	M+H	positive	Yes	target	Lamotrigin_2676_ p	C9H7Cl2N5	Cl-c(ccc1)c(Cl)c1-c(nnc2N)
3.96	268.1543604517	1.09E+08	0.00E+00	M+H	positive	Yes	target	Atenololsaeure (Metoprololsaeure)_697	C14H21N1O4	CC(C)NCC(O)COc1ccc(cc1)
6.09	253.0972162649	8.72E+07	0.00E+00	M+H	positive	Yes	target	Carbamazepine-10-11-epoxide_916_ p	C15H12N2O2	NC(=O)N2c1ccccc1C4OC4(
6.94	267.0876513875	8.28E+07	0.00E+00	M+H	positive	No	target	Valsartansaeure_2795_ p	C14H10N4O2	C1=CC=C(C(=C1)C2=CC=
8.55	216.1012139414	7.70E+07	0.00E+00	M+H	positive	Yes	target	Atrazin_288_ p	C8H14Cl1N5	c1(nc(nc(n1)Cl)NCC)NC(C)(
9.57	441.1670258168	7.13E+07	0.00E+00	M+H	positive	Yes	target	Candesartan_2804_ p	C24H20N6O3	c1(ccccc1c1ccc(Cn2c(nc3c
6.10	271.10775803	6.78E+07	0.00E+00	M+H	positive	Yes	target	Carbamazepin-10-11-dihydro-10-11-di	C15H14N2O3	c12[C@@H]([C@@H](O)c3(
3.63	178.054250279	6.06E+07	0.00E+00	M-H	negative	Yes	target	Cyclamat_2813_n	C6H13N1O3S1	C1(CCCCC1)NS(O)(=O)=O
1.07	123.0916004272	5.91E+07	2.59E+06	M+H	positive	No	target	4-Dimethylaminopyridine_2833_ p	C7H10N2	CN(C)C1=CC=NC=C1
5.74	188.0698990433	5.71E+07	6.07E+05	M+H	positive	Yes	target	Atrazine-Desethyl_309_ p	C6H10Cl1N5	c1(nc(nc(n1)Cl)N)NC(C)C
3.56	192.0767182847	4.81E+07	0.00E+00	M+H	positive	Yes	target	Carbendazim_278_ p	C9H9N3O2	c12c(cccc1)N=C(NC(OC)=(
6.38	278.2114644132	4.70E+07	0.00E+00	M+H	positive	Yes	target	Venlafaxin_645_ p	C17H27N1O2	CN(C)CC(C1=CC=C(C=C1)
7.73	202.0854156733	4.70E+07	0.00E+00	M+H	positive	Yes	target	Terbutylazin-desethyl_671_ p	C7H12Cl1N5	Nc1nc(NC(C)(C)C)nc(Cl)n1
4.38	235.1805227377	4.51E+07	0.00E+00	M+H	positive	Yes	target	Lidocain_2572_ p	C14H22N2O1	CCN(CC)CC(=O)NC1=C(C)(
13.21	412.9664972599	4.31E+07	8.25E+05	M-H	negative	Yes	target	Perfluoroctansaeure (PFOA)_2717_n	C8H1F15O2	O=C(O)C(F)(F)C(F)(F)C(F)(F
4.89	152.1434026509	3.67E+07	0.00E+00	M+H	positive	Yes	target	Amantadine_3124_ p	C10H17N1	C1C2CC3CC1CC(C2)(C3)N
4.70	254.0594440809	3.34E+07	0.00E+00	M+H	positive	Yes	target	Sulfamethoxazol_298_ p	C10H11N3O3S	c1(S(Nc2cc(C)on2)(=O)=O
12.04	284.1411922178	3.22E+07	0.00E+00	M+H	positive	Yes	target	Metolachlor_268_ p	C15H22Cl1N1O	CCC1=CC=CC(=C1N(C(C)(
2.94	184.1192596921	2.60E+07	0.00E+00	M+H	positive	Yes	target	Simazin-2-hydroxy_668_ p	C7H13N5O1	c1(nc(nc(n1)Cl)NCC)NCC
2.94	184.1192596921	2.60E+07	0.00E+00	M+H	positive	Yes	target	Terbutylazin-desethyl-2-hydroxy_672_	C7H13N5O1	Nc1nc(NC(C)(C)C)nc(O)n1
13.22	308.1525173486	2.56E+07	0.00E+00	M+H	positive	Yes	target	Tebuconazol_327_ p	C16H22Cl1N3O	CC(C)(C)C(CCC1=CC=C(C=
5.45	307.1113757675	2.00E+07	0.00E+00	M+H	positive	Yes	target	Fluconazol_328_ p	C13H12F2N6O	OC(CN1C=NC=N1)(CN2C=

Organisation_details | ANALYSIS | METHOD LC-MS(MS) | METHOD GC-MS(MS) | Suspect_substances | Sheet7 | Sheet8

EnviLights@EBI ?

- Example: Eawag nontarget screening trial
- Same information as in DCT
- Material & Methods from Paper
- Raw data deposited as mzML files

MTBLS_DEV154: Norman Nontarget screening trial Eawag

Share Study | View all files — Private study — Make it public

Submited: **19-Jun-2014** ,Release date: **17-Jun-2015**

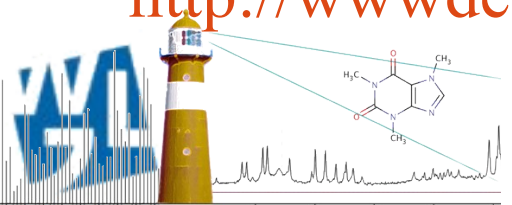
Other identifiers: NormanNontargetTrialEawag

This Collaborative Trial is organised by the NORMAN Association (www.normannetwork.net) as part of its Joint Programme of Activities for the year 2013. The activity is a follow-up action to the NORMAN-JRC workshop in Stresa (2010), where one of the main conclusions was that comparison and harmonisation of non-target screening methods in Europe are needed. The Collaborative Trial should provide in-depth information on the methodologies used by participating laboratories. It will be carried out for the first time in the area of environmental analysis worldwide and therefore aims to result in a common publication in a prestigious refereed journal. The Collaborative Trial will be carried out with a sample from the Danube river using the liquid chromatography-high resolution-mass spectrometry (LC-HR-MS) and gas chromatography-mass spectrometry (GC-MS) methodologies available in participating laboratories. These are the results of the Eawag participation using LC-MS/MS methods.

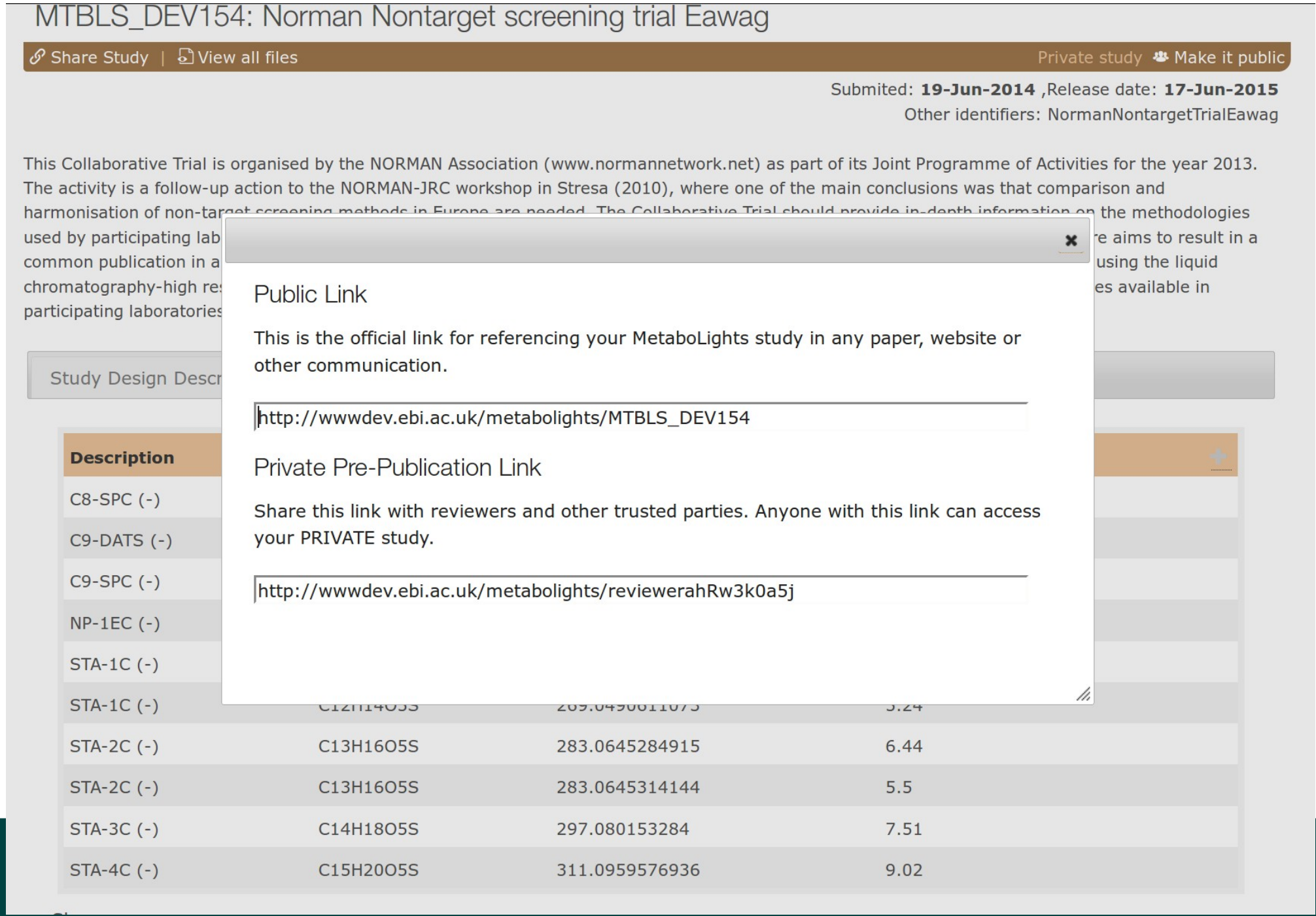
Study Design Description | Protocols | Assay | **Metabolites Identified** | Study Files

Description	Formula	m/z	Retention time
C8-SPC (-)	C14H20O5S	299.0958219827	8.14
C9-DATS (-)	C15H22O3S	281.121647622	13.16
C9-SPC (-)	C15H22O5S	313.1115	9.3
NP-1EC (-)	C17H26O3	277.1807672249	16.9
STA-1C (-)	C12H14O5S	269.0490519962	5.89
STA-1C (-)	C12H14O5S	269.0490611073	5.24
STA-2C (-)	C13H16O5S	283.0645284915	6.44
STA-2C (-)	C13H16O5S	283.0645314144	5.5
STA-3C (-)	C14H18O5S	297.080153284	7.51
STA-4C (-)	C15H20O5S	311.0959576936	9.02

- http://wwwdev.ebi.ac.uk/metabolights/reviewerahRw3k0a5j

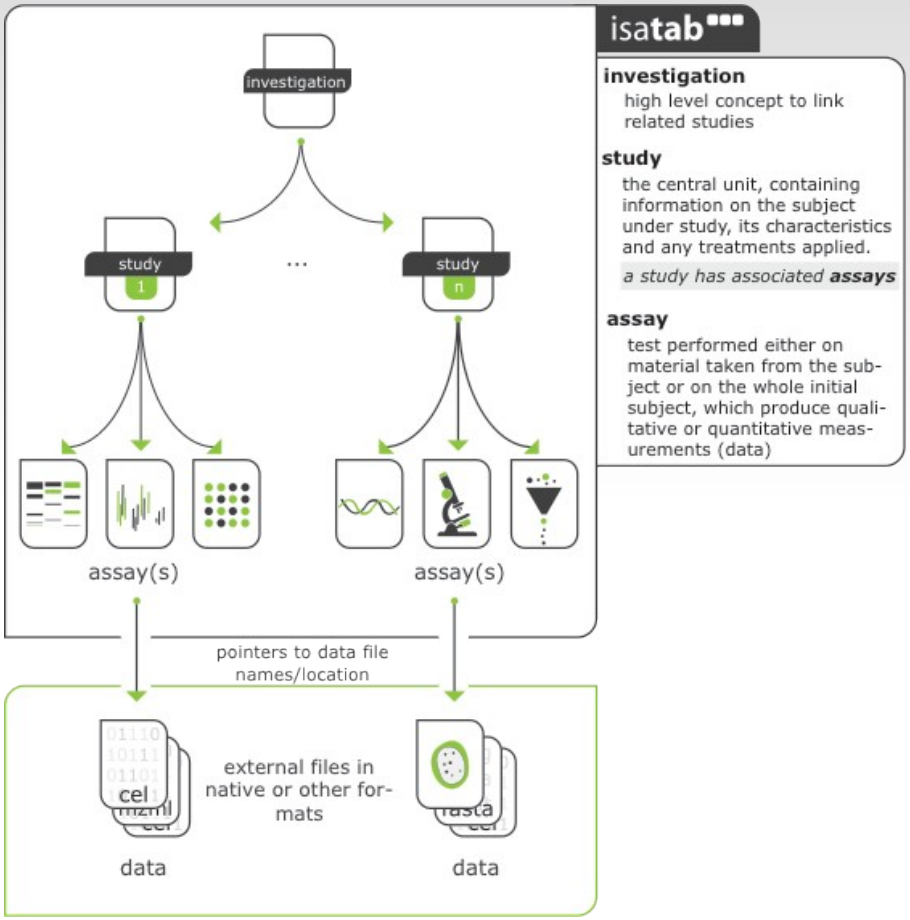


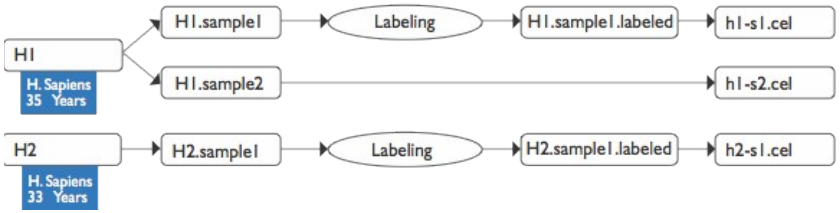
MetaboLights – Private Data



ISAtab in a nutshell

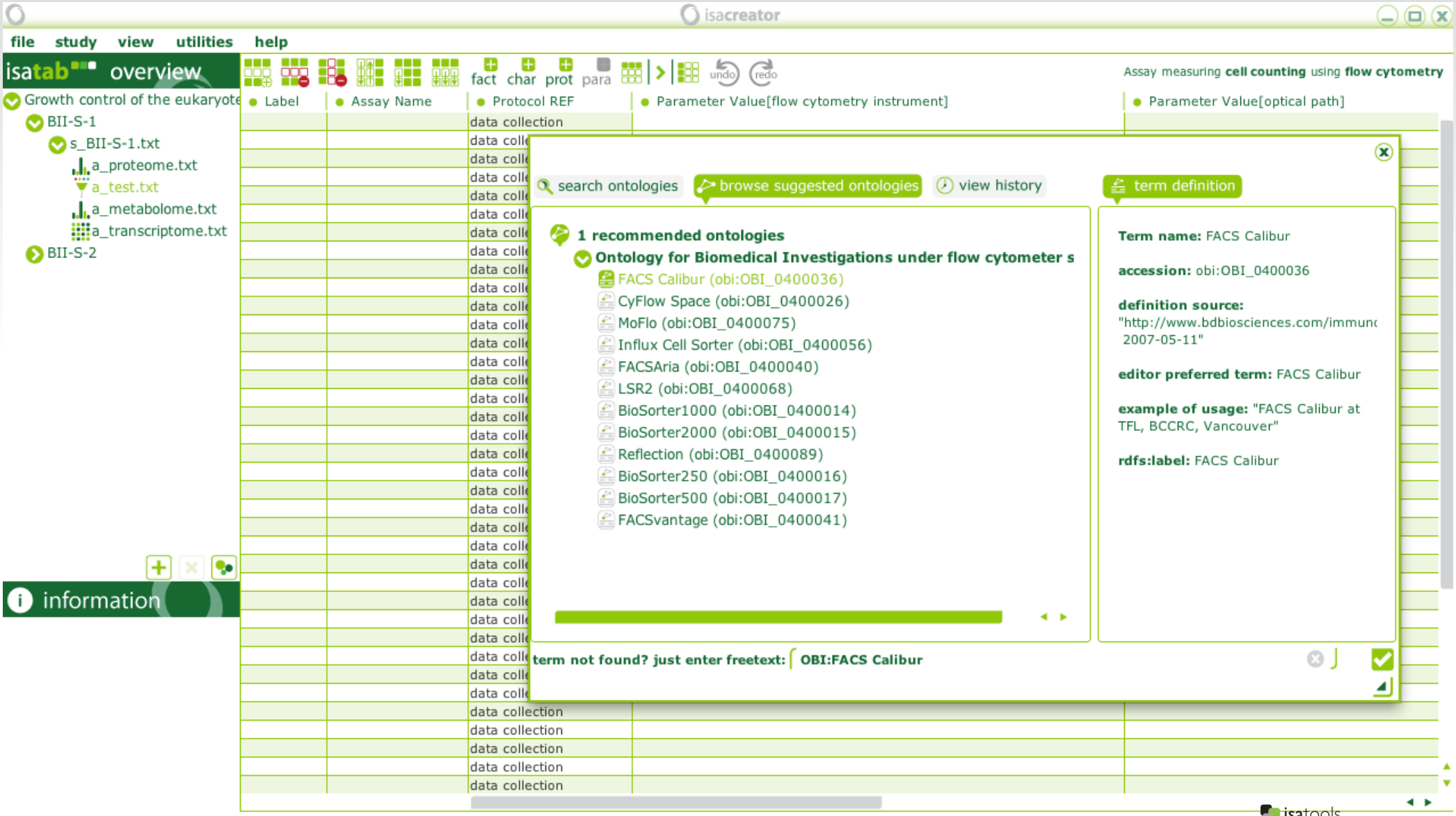
- Tab separated format
- Investigation, Study, Assay
- Checklist enforcement
- Ontology support
- Validation





Source Name	Organism	Age	Unit	Sample Name	Protocol REF	Labeled Extract Name	Data File
H1	H. Sapiens	35	Years	H1.sample1	Labeling	H1.sample1.labeled	h1-s1.cel
H1	H. Sapiens	35	Years	H1.sample2			h1-s2.cel
H2	H. Sapiens	33	Years	H2.sample1	Labeling	H2.sample1.labeled	h2-s1.cel

ISAcreator: the editor

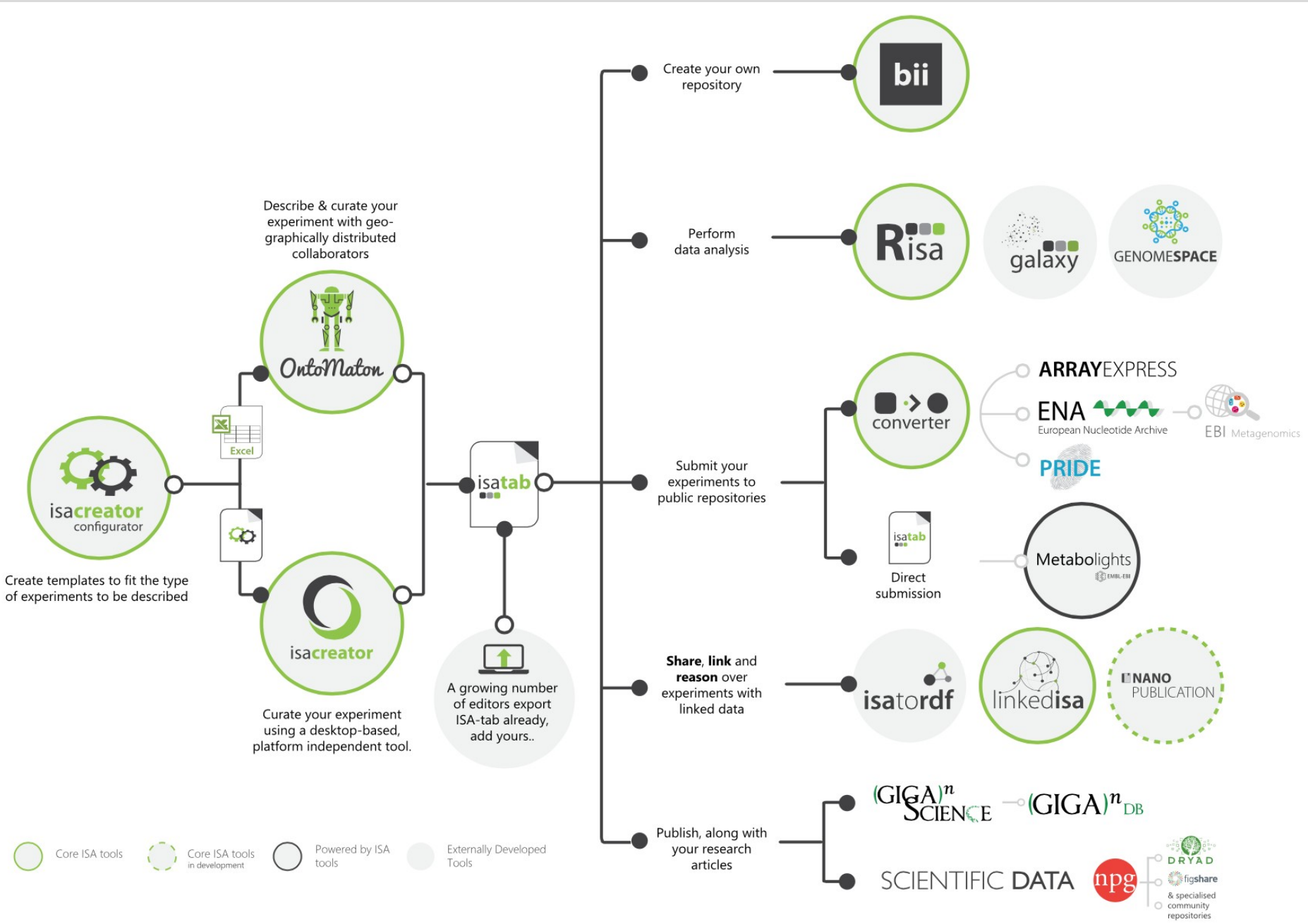


https://github.com/ISA-tools/ISAcreator

ISAcreator - Metabolite Annotation plugin

database_identifier	chemical_formula	smiles	inchi	metabolite_identification	chemical_shift	multiplicity
CHEBI:1148	C4H8O3	CCC(C(=O)O)O	InChI=1S/C4H8O3/c...	a-Hydroxy-n-butyrate	[0.85 .. 0.90]	t 0.90
CHEBI:60630	C5H9O3-	CCCC(C(=O)[O-])O	InChI=1S/C5H10O3/...	a-Hydroxy-n-valerate	[0.90 .. 0.96]	t 0.92
CHEBI:24898	C6H13NO2	C(CCC(=O)O)CCN	InChI=1S/C6H13NO2...	isoleucine	[0.90 .. 0.96]	t 0.94
CHEBI:25017	C6H13NO2	CC(C)CC(C(=O)O)N	InChI=1S/C6H13NO2...	leucine	[0.96 .. 1.01]	t 0.96
CHEBI:57762	C5H11NO2	CC(C)C(C(=O)O)N	InChI=1S/C5H11NO2...	valine	[0.96 .. 1.01]	t 0.97
CHEBI:64669	C5H10O3	CC(C)(C(=O)OC)O	InChI=1S/C5H10O3/...	a-Hydroxyisovalerate	[0.96 .. 1.01]	
CHEBI:24898	C6H13NO2	C(CCC(=O)O)CCN	InChI=1S/C6H13NO2...	isoleucine	[1.01 .. 1.03]	d 1.01
CHEBI:35932	C6H10O3	CCCC(=O)C(=O)OC	InChI=1S/C6H10O3/...	a-keto-b-methyl-N valerate	[1.03 .. 1.09]	d 1.075
CHEBI:16530	C5H8O3	C1CC(=O)OC1CO	InChI=1S/C5H8O3/c...	2-oxoisovalerate	[1.09 .. 1.13]	d 1.11
CHEBI:16236	C2H6O	CCO	InChI=1S/C2H6O/c1...	ethanol	[1.09 .. 1.13]	t 1.11
					[1.13 .. 1.16]	
					[1.16 .. 1.18]	
CHEBI:20067	C4H8O3	CC(CC(=O)O)O	InChI=1S/C4H8O3/c...	b-hydroxy-n-butyrate	[1.18 .. 1.22]	d 1.20
				unknown_t_1.205	[1.18 .. 1.22]	t 1.205
				unknown_d_1.235	[1.22 .. 1.26]	d 1.235
CHEBI:30860	C4H6O4	C(CC(=O)O)C(=O)O	InChI=1S/C4H6O4/c...	methylmalonate	[1.22 .. 1.26]	d 1.255
				unknown_s_1.27	[1.26 .. 1.32]	s 1.27
CHEBI:28358	C3H6O3	CC(C(=O)O)O	InChI=1S/C3H6O3/c...	lactate	[1.32 .. 1.35]	d 1.335
				unknown_s_1.35	[1.35 .. 1.38]	s 1.35
CHEBI:50129	C4H8O3	C(CC(=O)[O-])CO	InChI=1S/C4H8O3/c...	a-hydroxyisobutyrate	[1.35 .. 1.38]	s 1.36
					[1.38 .. 1.43]	
					[1.43 .. 1.47]	
CHEBI:16449	C3H7NO2	CCOC(=O)N	InChI=1S/C3H7NO2/...	alanine	[1.47 .. 1.50]	d 1.485
					[1.50 .. 1.55]	
CHEBI:64390	C8H15NO3	CC(C)C(C(=O)OC)NC(...	InChI=1S/C8H15NO3...	n-caproylglycine	[1.55 .. 1.60]	
					[1.60 .. 1.63]	
CID638154	C8H15NO3	CC(C)C(C(=O)OC)NC(...	InChI=1S/C8H15NO3...	methyl N-acetylvalinate	[1.63 .. 1.65]	
					[1.65 .. 1.70]	
CHEBI:25017	C6H13NO2	CC(C)CC(C(=O)O)N	InChI=1S/C6H13NO2...	leucine	[1.70 .. 1.76]	m 1.71
CHEBI:25094	C6H14N2O2	CC(CCCN)(C(=O)O)N	InChI=1S/C6H14N2O...	lysine	[1.70 .. 1.76]	m 1.73
					[1.76 .. 1.81]	
CHEBI:18257	C5H12N2O2	C(CC(C(=O)O)N)CN	InChI=1S/C5H12N2O...	ornithine	[1.81 .. 1.87]	m 1.81
CHEBI:66922	C6H12N3O3-	C(CC(C(=O)[O-])N)CN...	InChI=1S/C6H13N3O...	citrullinate	[1.81 .. 1.87]	m 1.88

ISA ecosystem

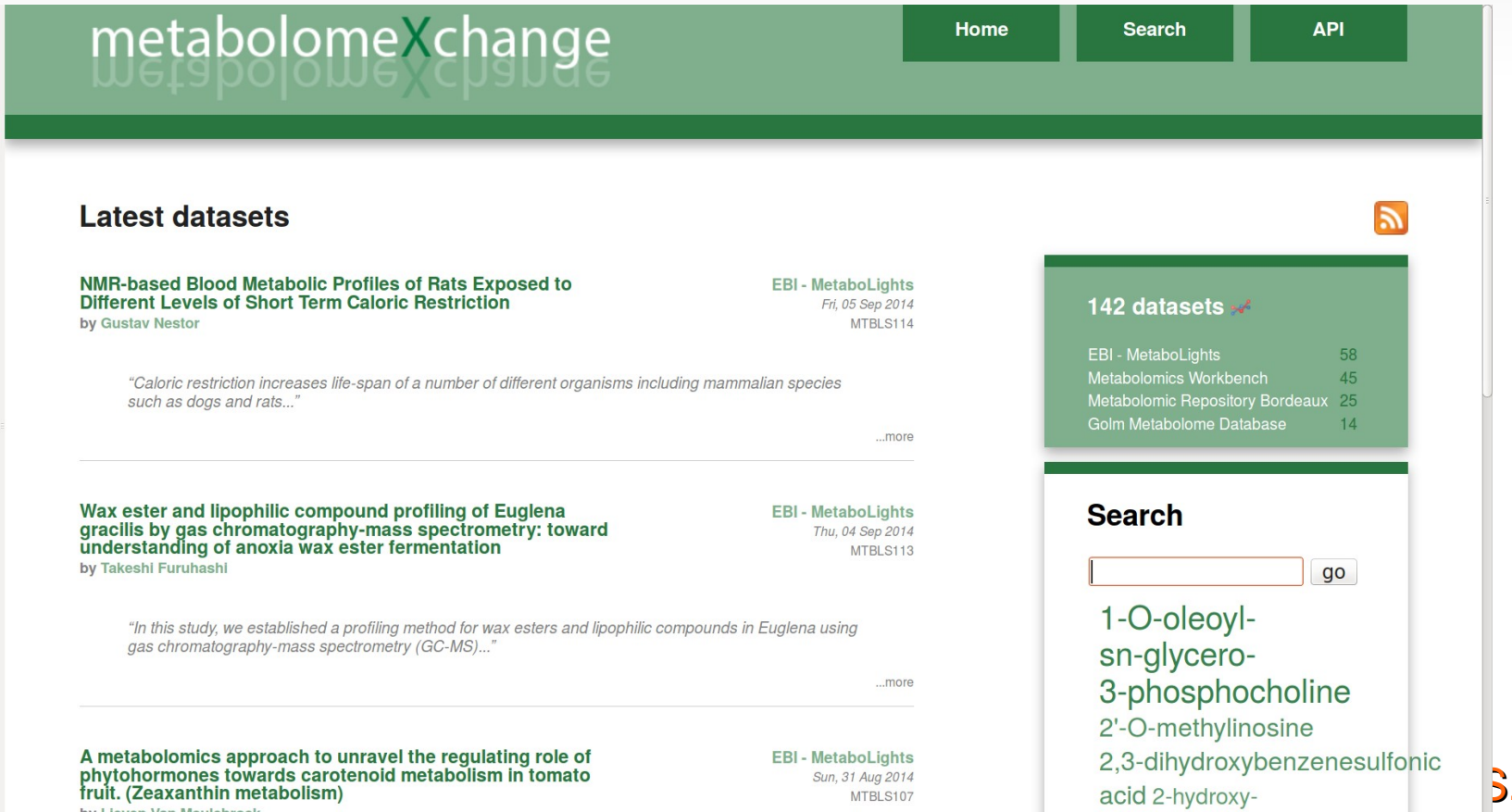


MetabolomeXchange.org

- Network of repositories (EU/US/...)
- Metadata exchange, storage & search

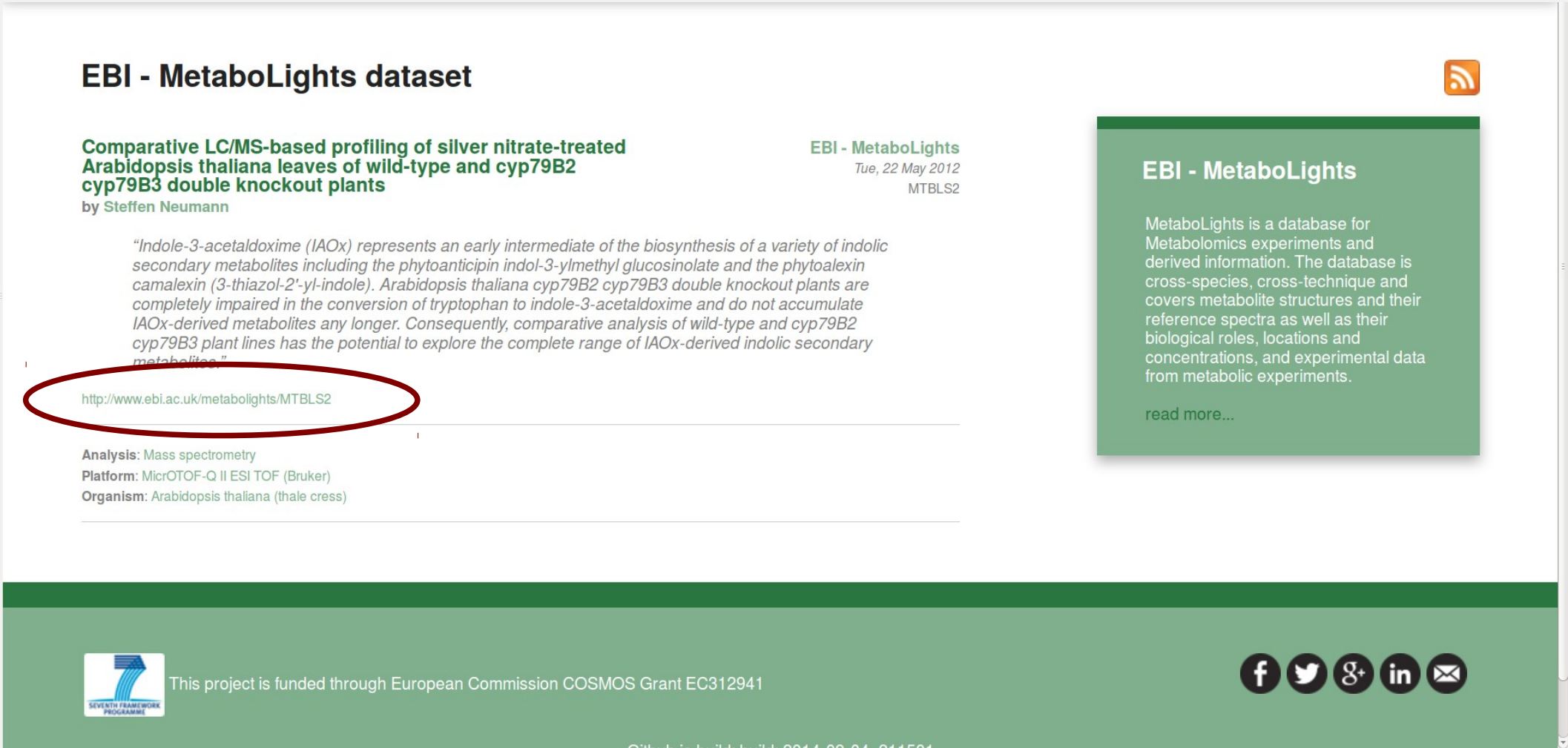
142 datasets (08.09.2014)

EBI - MetaboLights	58
Metabolomics Workbench	45
Metabolomic Repository Bordeaux	25
Golm Metabolome Database	14



MetabolomeXchange.org

- No actual data: Link back to repo



EBI - MetaboLights dataset

Comparative LC/MS-based profiling of silver nitrate-treated Arabidopsis thaliana leaves of wild-type and cyp79B2 cyp79B3 double knockout plants
by Steffen Neumann

EBI - MetaboLights
Tue, 22 May 2012
MTBLS2

"Indole-3-acetaldoxime (IAOx) represents an early intermediate of the biosynthesis of a variety of indolic secondary metabolites including the phytoanticipin indol-3-ylmethyl glucosinolate and the phytoalexin camalexin (3-thiazol-2'-yl-indole). Arabidopsis thaliana cyp79B2 cyp79B3 double knockout plants are completely impaired in the conversion of tryptophan to indole-3-acetaldoxime and do not accumulate IAOx-derived metabolites any longer. Consequently, comparative analysis of wild-type and cyp79B2 cyp79B3 plant lines has the potential to explore the complete range of IAOx-derived indolic secondary metabolites."

http://www.ebi.ac.uk/metabolights/MTBLS2

Analysis: Mass spectrometry
Platform: MicrOTOF-Q II ESI TOF (Bruker)
Organism: Arabidopsis thaliana (thale cress)

EBI - 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.

read more...

This project is funded through European Commission COSMOS Grant EC312941

MetabolomeXchange.org

- API to integrate into software and websites
- Push and Pull between data consumers and data providers

GET :: /api/dataset/search/@in/@for

Find all datasets where @for is found in @in. The @in value is case insensitive.

All datasets submitted by Reza (max 2)

cURL: *curl -i http://metabolomexchange.org/api/dataset/search/submitter/Reza/?limit=2*


```
[
 {
  "accession": "MTBLS3",
  "url": "http://www.ebi.ac.uk/metabolights/MTBLS3",
  "title": "A metabolomic strategy defines the regulation of lipid content and global metabolis
  "description": "Caenorhabditis elegans provides a genetically tractable model organism to
metabolism and how regulation is perturbed to produce the complex phenotype of obesity. C
including the delta-9 desaturases expressed by fat-5, fat-6 and fat-7. They regulate the biosy
synthesis of lipids including phospholipids, triglycerides and cholesteryl esters. Gas chroma
define the metabolome of all the possible knock-outs for the delta 9 desaturases, including fo
similar enzymatic roles, excellent discrimination was achievable for all single and viable dou
and fat-7, both expressing steroyl-CoA desaturases. The metabolomic changes extend to ac
9 desaturases have on regulating global metabolism and highlighting how comprehensive m
used dyes for fat staining. The propagation of metabolic changes across the network of meta
desaturases places C.elegans into a catabolic state compared with wildtype controls.",
  "date": "1329177600",
  "submitter": "Reza Salek",
  "meta": {
   "analysis": "mass spectrometry",
   "platform": "DSQ II Single Quadrupole GC/MS (Thermo Scientific)",
   "organism": "Caenorhabditis elegans",
   "metabolites": [
    "tyrosine",
    "ornithine",
    "L-histidine zwitterion",
```

Granularity of metadata

- There can be too little and too much metadata
  - Too little: meaningless data dump
  - Too much: over-engineered database, problematic User Acceptance
- Future: Text-Mining the MetaData ?
  - Automatic RDF extraction as suggestion during submission
- Different granularity for in-house “LIMS” and published data-sets ?

Ontologies

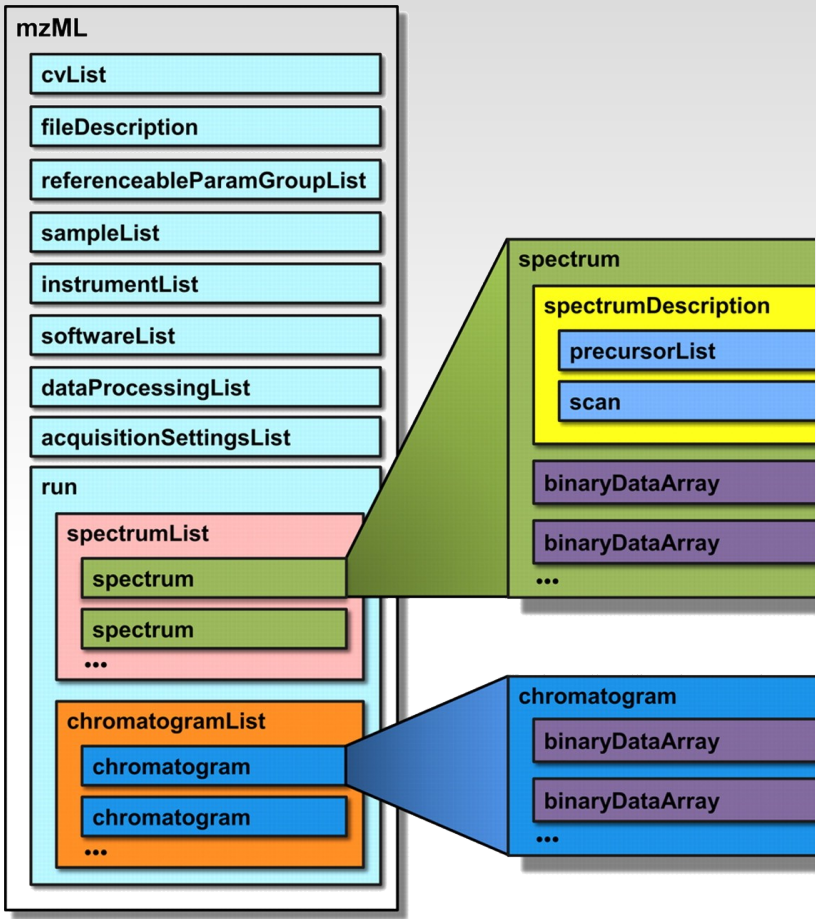
- Controlled vocabulary/ies
- Relationships: e.g.
  `is_a, contained_in, tributary_of`
- Different formats (OWL, OBO)
- Online queries: www.ebi.ac.uk/ols
  or http://bioportal.bioontology.org/
- Centrally or community maintained

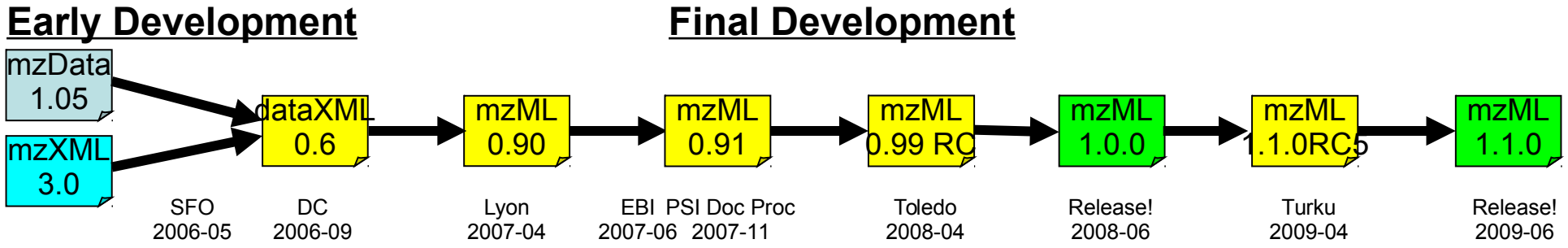
Change gears: low-level data standards

- Standardised Metadata allows to compare and integrate knowledge between Experiments
- Proprietary data formats:
  *.d (Bruker) or *.RAW (Waters) or …
- Standardised (raw) data formats allow to mix-and-match software and workflow pipelines
- Open Formats: well documented, I/O and parsing available for Java, R, Matlab, …

Example: mzML

- mzData and mzXML flawed: too little/much structure
- Joint open development
- ABI, Bruker, Waters, Thermo, …
- Implementations: jmzML, mzR, OpenMS, pymzML, proteowizard, …



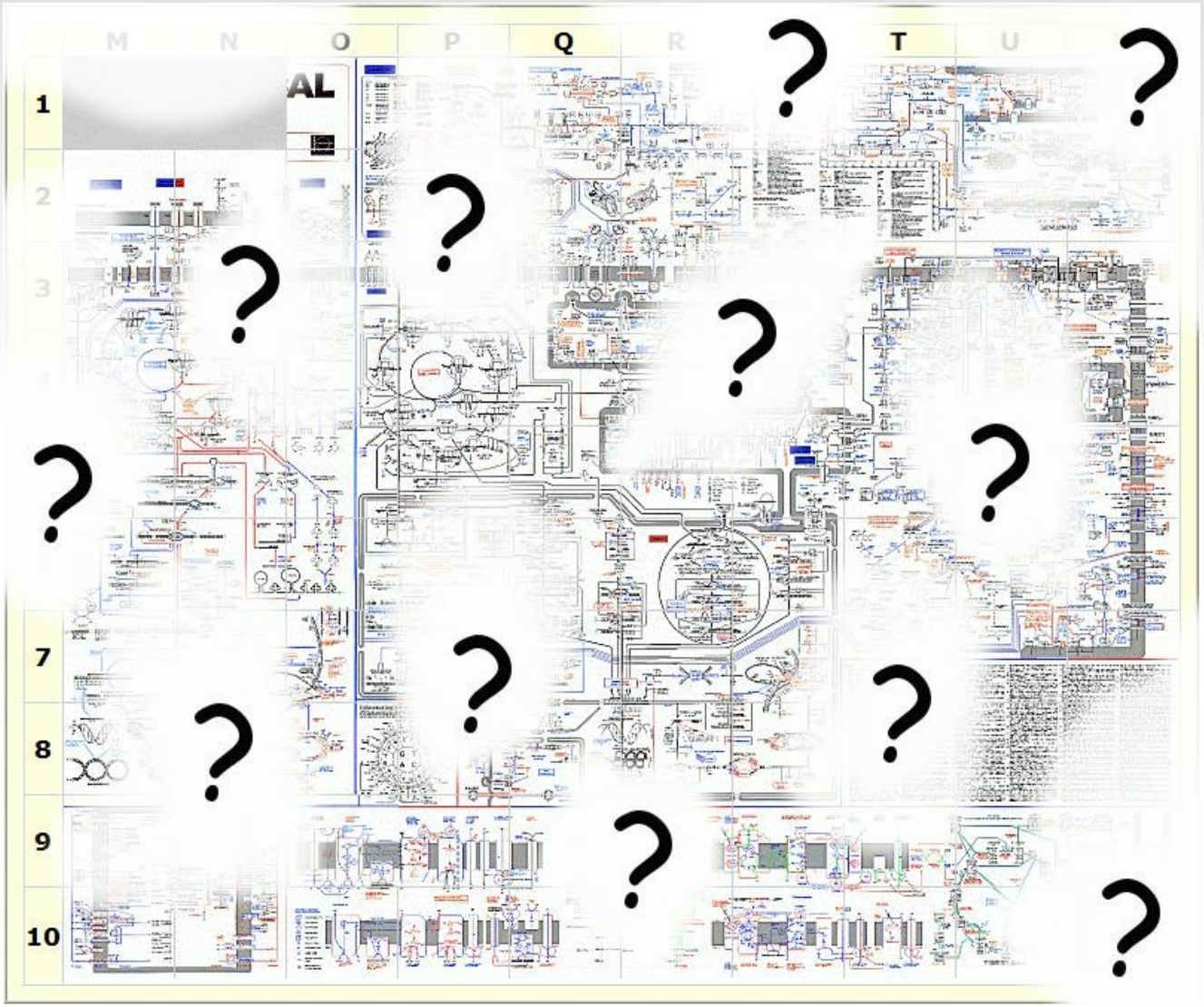


Acknowledgements

Many:

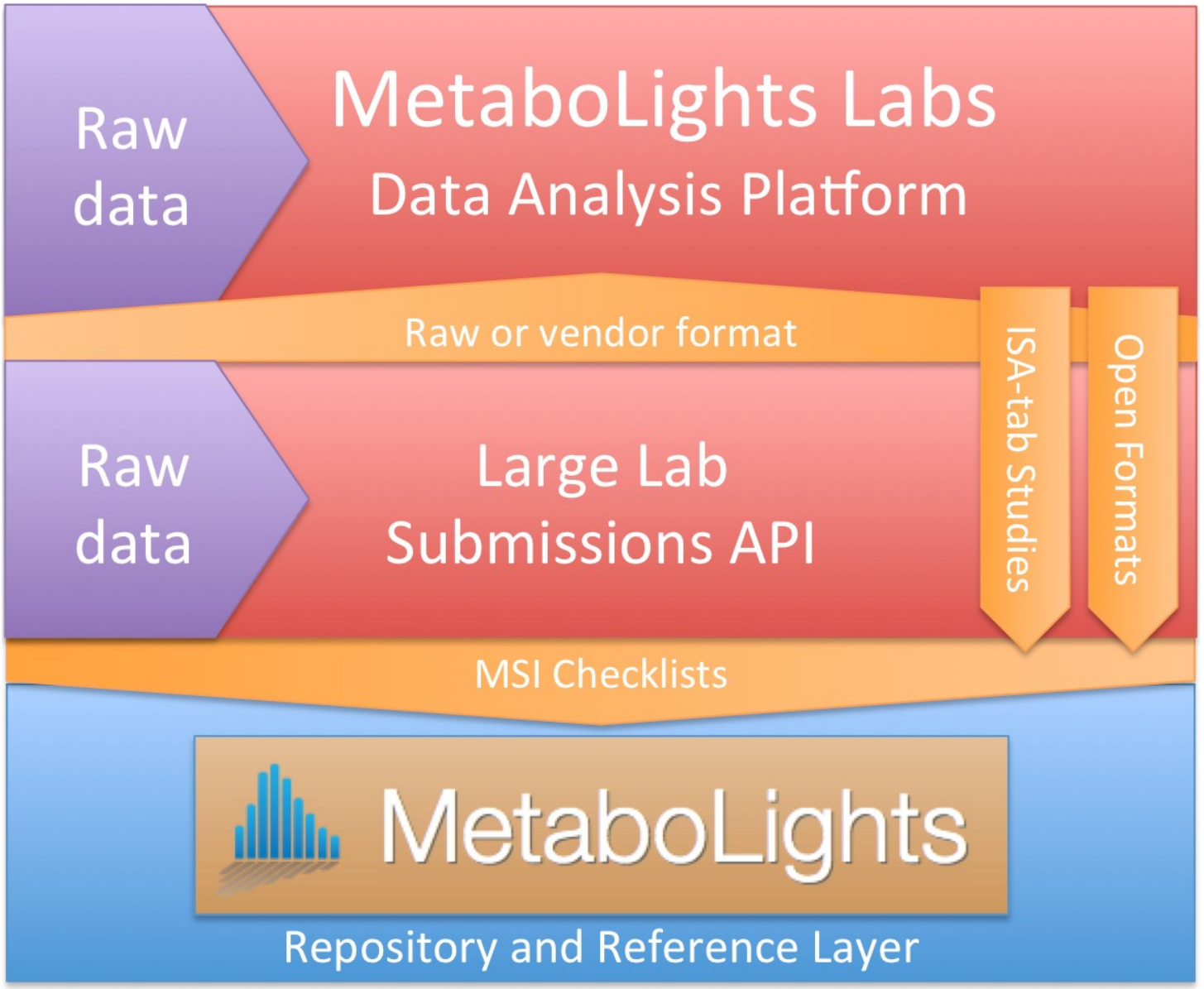
- Eawag (Emma, Michele, Heinz, Juliane, …)
- EBI Metabolights team Hinxton
- ISA team Oxford
- COSMOS project members
- IPB Halle team

Questions ?



Only supplemental slides beyond this point

MetaboLights – What are we working on



MetaboLights – Analysis

EMBL-EBI

Services | Research | Training | About us

MetaboLights DEV

Search

Examples: alanine, human, urine, MTBLS1

Home | Browse Studies | Browse Compounds | Browse Species | Analysis | Download | Help | Give us feedback | About | Submit Study | Login

MetaboLights > MetaboAnalyst

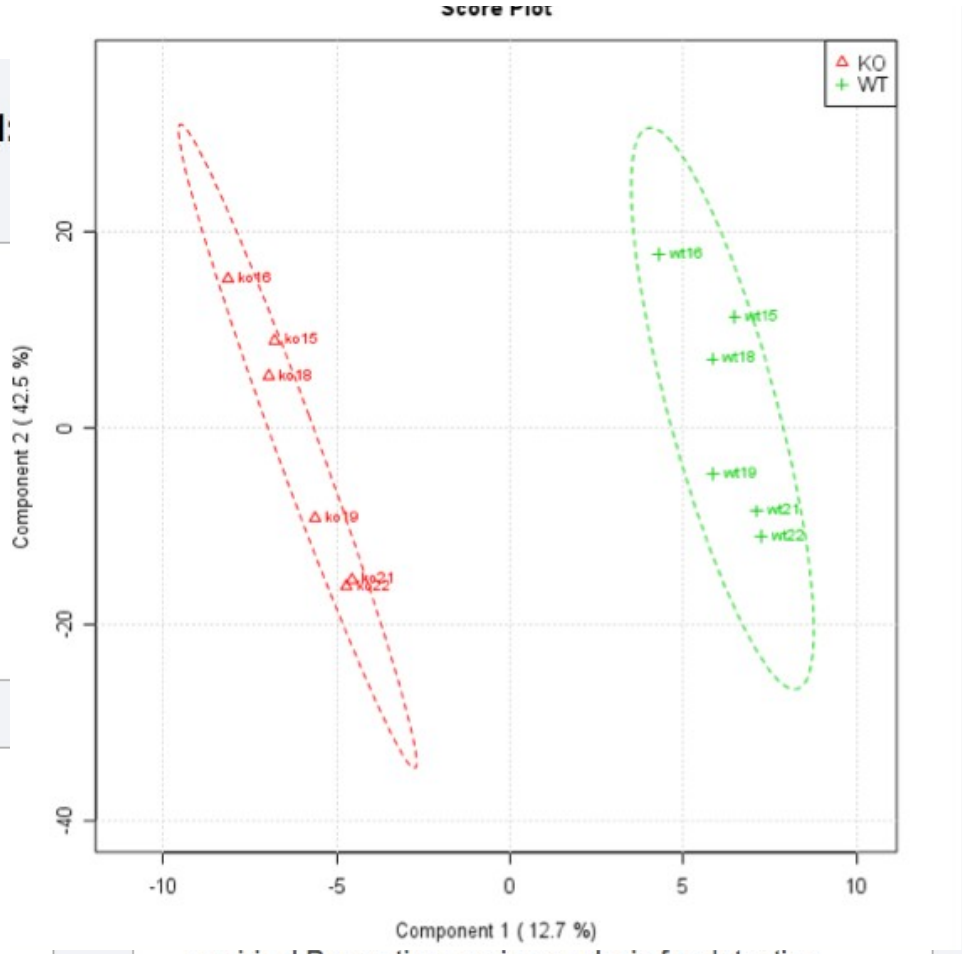
Please choose a functional module to proceed:

Statistical Analysis

This module offers various commonly used statistical and machine learning methods from t-tests, ANOVA to PCA and PLS-DA. It also provides clustering and visualization such as dendrogram, heatmap, K-means, as well as classification based on random forests and SVM.

Pathway Analysis

This module supports pathway analysis (including pathway enrichment analysis and pathway topology



EMBL-EBI

MetaboLights – What more are we working on

- Tools integration, R focus (MetaboAnalyst, rISA, MetFrag, BATMAN)
- Online creation and edit of studies
- Extensive international data exchange
- Considering “Community Portals” on top of MetaboLights
  - Branded for each community
  - Bespoke curation and submission templates
  - Data exchange
  - Linking community resources

License: rights+obligations

- CreativeCommons: Family of licenses
  - ByAttribution (BY): Obligation to cite the origin
  - ShareAlike (SA): Derived must be CC-SA as well
    The "viral" license, it spreads!
  - NoDerivatives: Redistribution only unchanged
  - NC (NonCommercial): problematic definition
  - And (almost) combinations thereof
- And many other: `opendatacommons.org/licenses`
- No license is the worst choice, because the "default rights/obligations" differ across legislations

MetaboLights – The team

Kenneth Haug

Pablo Conesa

Kalai Jayaseelan

Software development

Reza Salek

Mark Williams

Namrata Kale

Curation

Christoph Steinbeck

Jules Griffin (UC/MRC)

PIs

Previous: Paula de Matos, Mark Rijnbeek, Tejasvi Mahendraker

Susanna **Sansone**
Team Leader
University of Oxford
LinkedIn Twitter

Philippe **Rocca-Serra**
Technical Co-ordinator
University of Oxford
GitHub LinkedIn Twitter

Eamonn **Maguire**
Lead Software Engineer
University of Oxford
Figshare GitHub LinkedIn Twitter

Alejandra **Gonzalez-Beltran**
Senior Software Engineer
University of Oxford
GitHub LinkedIn Twitter

funding

2012-2015, latest funding stream for metabolomics standards from the EU COSMOS.

2012-2015, travel funds for a UK-China partnership on harmonization of data curation from the BBSRC, a collaboration with TGAC and BGI/GigaScience

2012-2015, more funding for ISA project from the BBSRC MGportal project, a collaboration with the EMBL-EBI.

2011-2014, the ISA project has received another new funding stream from BBSRC and NERC.

2010-2013, the ISA project has received new funding stream from BBSRC and NERC.

2007-2010, the ISA project has been mainly supported by funds from the EU IP CarcinoGenomics, with contribution from EU NoE NuGO, NERC-NEBC, BBSRC and EMBL-EBI.

isatools

UNIVERSITY OF OXFORD

OXFORD e-Research CENTRE