Project deliverable report

DELIVERABLE NUMBER AND TITLE	D2.2.2.
TITLE	ONTOLOGIES FOR METABOLOMICS
AUTHOR(S)	LARS OVE DRAGSTED, MARIONA PINART, FEDERICO MARINI, JILDAU BOUWMAN
WORK PACKAGE	WP 2
TASK	TASK 2.2.
WP LEADER	LARS OVE DRAGSTED
BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE	UNIV. COPENHAGEN, MDC BERLIN, UNIV. SAPIENZA ROME, TNO
STATUS – VERSION	FINAL - VERSION 1.0
DELIVERY DATE (MONTH)	M18
SUBMISSION DATE	M30
DISSEMINATION LEVEL – SECURITY*	PU
DELIVERABLE TYPE**	R

* Security: PU – *Public*; PP – *Restricted to other programme participants (including JPI Services)*: RE – *Restricted to a group specified by the consortium (including JPI Services)*; CO – *Confidential, only for members of the consortium (including JPI Services)*

** Type: R – *Report*; P – *Prototype*; D – *Demonstrator*; - O - *Other*

CONTENTS

Status of Ontologies for metabolomics 3

MS metabolomics 4

Existing terms 4

New terms needed and work plan for continuation 6

NMR metabolomics 7

Existing terms 7

New terms needed and further work 8

Further work to support data sharing 8

Literature 8

Appendix 1: detailed NMR ontology 9

STATUS OF ONTOLOGIES FOR METABOLOMICS

Ontologies exist at the major website Ontobee (http://www.ontobee.org/) for metabolomics. There are at least two definitions for metabolomics. The first one (ID http://edamontology.org/topic_3172) is under EDAM bioinformatics topics/omics and defines Metabolomics as "the systematic study of metabolites, the chemical processes they are involved in, and the chemical fingerprints of specific cellular processes in a whole cell, tissue, organ or organism", whereas the second one is under Medical subject headings/Natural science disciplines/ Chemistry/Biochemistry/ (ID http://purl.bioontology.org/ontology/MESH/D055432) defining Metabolomics as "the study of metabolite patterns in biological samples and correlation with xenobiotic challenge and disease states", noting that the term was coined by Nicholson in 1999 to mean "the systematic identification and quantitation of all the metabolic products of a cell, tissue, organ, or organism under varying conditions" and that "the METABOLOME of a cell or organism is a dynamic collection of metabolites which represent its net response to current conditions". The latter entry also includes the terms, Metabonomics, Metabolomic, and Metabonomic that are now considered synonymous by most groups.

Metabolomic profiling is also part of the ontology tree for "Experimental factor" (http://www.ebi.ac.uk/efo/EFO_0000752). Metabolic profiling, metabolomics profiling or metabolomic profiling are synonyms and may be defined as "an assay in which a defined subset or all measurable metabolites are analysed". When a subset of metabolites are defined in advance the method is labelled as "targeted", while methods where all measurable metabolites are recorded are labelled "untargeted".

Common to all of these websites are that the terms for metabolomics or profiling are parents without links to any children, which in effect means that there is no detailed ontology for the practise of metabolomics profiling or the related data handling and analysis procedures. The best practice of metabolic profiling has been described by Sumner et al. (2007). This paper provides minimal requirements for reporting and details also the need for detailed information about the steps preceeding metabolic profiling, i.e. the study outline and sample collection procedures (not covered here). The terms for subdivision of the steps needed to conduct metabolic profiling include (new entries in brackets).

Metabolomics

- Sample analysis
  - Sample preparation
    - Chromatography
      - Mass spectrometry
    - NMR
    - FT-IR spectroscopy
    - (other spectrometry)
  - Instrumental performance and method validation
- Data analysis
  - Data preprocessing

(Statistical data analysis)

Metabolite identification

Several subdivisions of these steps should help to further create an ontology for the full process. This work should be further developed based on the output of the COordination of Standards in MetabOlomicS (COSMOS; Salek et al, 2015) collaboration and especially using the terms used in the MetaboLights database of metabolomics data.

Ontologies exist for the two major technologies: Nuclear Magnetic Resonance spectrometry (NMR, http://www.ontobee.org/ontology/NMR) and mass spectrometry (MS, http://www.ontobee.org/ontology/MS). Both of them are downloadable as Excel and text files containing links for each term included.

These tend to be extremely detailed and should be linked with metabolomics ontology. However, the outline of profiling may be less detailed to cover the steps and terms most often used in metabolomics.

MS METABOLOMICS

Ontologies for MS exist in several recent versions. From Bioportal the latest ontology is ver. 4.0.1, uploaded 30 Aug 2016 (https://bioportal.bioontology.org/ontologies/MS). From Ontology Lookup Service the ms.owl ontology file ver. 4.0.3 from 31 Oct 2016 can be downloaded (http://purl.obolibrary.org/obo/ms/4.0.3/ms.owl). From Ontobee the ms ontology ver. 3.88 from 13. May 2016 can be obtained (http://www.ontobee.org/ontology/MS). All of these are versions of the same work under continuous development by the HUPO proteomics standards initiative (PSI) and some terms are therefore directed towards the application of MS for proteomics. However, a range of terms are generic and should therefore be shared between these existing ontologies and an MS-metabolomics ontology. As opposed to the situation for the term metabolomics, ontology for MS and proteomics is detailed with several layers of parent/children relationships.

EXISTING TERMS

All terms related to the MS ontology are listed with explanations and parent terms (but without links) in the overall MS ontology file for Excel. The list contains most generic terms used in MS analyses as well as a large number of vendor-specific terms and product names. This forms an easy entry for looking up terms.

The recent proteomics MS ontology work has added a range of additional terms to cover specifically the application of MS for proteomics.

&: The MS ontology has the following parent terms, most of which have additional children, i.e. derived terms lower in the hierarchy of terms, marked below with '&'. The derived (child) terms may be seen by clicking the links.

*****: Those marked with * could be used also for metabolomics as they are; for the remainder, there are two possibilities:

#: There is a need to add a definition that covers the use of the term better for metabolomics (marked #),

¤: to ignore the term (marked ¤):

- ambiguous residues &¤
- binary data array &*
- binary data compression type
- binary data type &*
- chemical compound &#
- chemical compound attribute &#
- chemical compound formula &*
- chromatogram type &*
- Cleavage agent regular expression &¤
- data file checksum type &*
- data file content &#
- data processing parameter &*
- data transformation &*
- detector acquisition mode &*
- detector attribute &*
- detector type &*
- dissociation method &*
- external reference identifier &#
- file format &*
- inlet type &*
- instrument model &#
- ion attribute ¤
- ion chemical type ¤
- ion optics &*
- ion reaction ¤
- ion role ¤
- ionization type &*
- laser attribute &*
- laser type &*
- m/z Separation Method ¤
- MALDI matrix application &*
- mass analyzer attribute &*
- mass analyzer type &*
- mass spectrometry ¤
- mass table options &*
- matrix application type &*
- measurement method &*
- modification parameters &*
- native spectrum identifier format &*
- object attribute &*
- peak &*
- peptide attribute ¤ (a similar term is needed for metabolite modifications)
- peptide modification details ¤ (a similar term is needed for metabolite modifications)
- polarity ¤
- precursor activation attribute &*

- prefix &*
- protein attribute &¤
- purgatory ¤
- quantification data processing &*
- quantification information &*
- resolution type ¤
- retention time window attribute &#
- scan direction &*
- Scan Function ¤
- scan law &*
- scan polarity &*
- scanning method ¤
- search input details &*
- sequential m/z separation method ¤
- software &*
- source attribute &*
- spectra combination &*
- spectrum identification result details &#
- spectrum representation &*
- spectrum type &*
- standard &*
- target inclusion exclusion priority &*
- transition validation method &*
- unit &*
- unit ¤

As seen from this list only the five parent terms marked with **#** (chemical compound attribute; external reference identifier; instrument model ; retention time window attribute; spectrum identification result details) may need to be changed or added with definitions of children specific to metabolomics. Most of the terms, marked with ¤ have been defined as obsolete at the Ontobee site, however some of them need terms for metabolomics to cover the analytical procedures.

NEW TERMS NEEDED AND WORK PLAN FOR CONTINUATION

Based on the analysis above of the existing terms for proteomics a set of terms or adjustments must be included in order to form a coherent MS ontology useful for metabolomics. This work must be done iteratively with the following steps:

1. Defining suggestive amendments to the five sets of parent and daughter (child) terms marked above with **#** and for additional terms needed.

2. Looking up these suggestive terms or possible alternatives in the overall MS ontology, defined by the Excel sheet at http://www.ontobee.org/ontology/MS.

    a. If an existing term exists that may cover the needs in metabolomics, we should then consider whether it is placed correctly in the hierarchy

b. If there is no current term, we should look it up more generally in Ontobee to ensure it is not present already in another ontology, e.g. for chemistry

3. To prepare definitions for each new, unique terms needed for metabolomics MS.

The degree of detailing actually needed for the metabolomics ontology may be similar to the current practice of detail provided in acceptable scientific reporting. Further work to set the detailing level needed should be done before each entry above is finalized for a first version of a metabolomics ontology.

NMR METABOLOMICS

From Bioportal the latest ontology version is ver. 0.1, uploaded on 20 Aug 2015 (https://bioportal.bioontology.org/ontologies/NMR). The nmrCV.owl ontology (version v1.0.rc1, http://nmrml.org/cv/) momentarily contains ≈ 600 classes, most of them under nmr namespace but others imported (via MIREOT in Ontofox) from the units, Chebi, ChemO, PSI-MS, OBI and BFO top level ontologies. The NMR.owl includes the descriptors relevant to Nuclear Magnetic Resonance (NMR) experimental conditions in the context of metabolomics investigations (See appendix 1). The terms also fulfil the "Proposed Reporting Requirements for the Description of NMR-based Metabolomics Experiments" (Rubtsov et al Metabolomics 2007, 3(3): 223–229 (See doc attached)). Currently the NMR.owl is a pure taxonomy of 247 classes in owl-DL format and only annotation properties have been assigned to the classes. The terms are 'binned' under some 'web imported' OBI and BFO classes (bfo.owl being secondary imported via OBI).

From Ontology Lookup Service nuclear magnetic resonance (http://www.bioassayontology.org/bao#BAO_0000160) states that “Nuclear magnetic resonance (NMR) is used to analyse protein-small molecule interactions. This binding interaction is commonly monitored by the chemical shift of the compound. With this technique, both the binding interaction and bound compounds can be detected, which allows pools of 5-100 compounds to be tested for binding”. From Ontobee the nmr ontology ver. 0.1 from 13. October 2006 can be obtained (http://www.ontobee.org/ontology/NMR). The NMR ontology is developed by the ontology working group (http://msi-ontology.sourceforge.net/) of the msi-metabolomics society (msi-workgroups.sf.net).

EXISTING TERMS

All terms related to the NMR ontology are listed with explanations and parent terms in the overall NMR ontology file for Excel. This list contains most generic terms used in NMR analyses. However, vendors like Bruker, Varian and JEOL typically provide also the software to process the vendor specific NMR data.

The recent proteomics NMR ontology work has added a range of additional terms to cover specifically the application of NMR for proteomics. The NMR ontology has the following parent terms, most of which have additional children, i.e. derived terms lower in the hierarchy of terms, marked below with ’&’. The derived (child) terms may be seen by clicking the links. Those marked with # a definition that covers the use of the term better for metabolomics should be added, and those marked with ¤ are obsolete terms and should be therefore ignored:

- entity &#
- deleted_classes &¤

NEW TERMS NEEDED AND FURTHER WORK

Based on the analysis above of the existing terms for proteomics a set of terms or adjustments must be included in order to form a coherent NMR ontology useful for metabolomics. This work must be done iteratively with the following steps:

1. Defining suggestive amendments to the parent and daughter terms marked above with #

2. Looking up these suggestive terms or possible alternatives in the overall NMR ontology, defined by the Excel sheet.

   a. If an existing term exists that may cover the needs in metabolomics we should consider whether it is placed correctly in the hierarchy

   b. If there is no current term we should look it up more generally in Ontobee to assure it is not present already in another ontology, e.g. for chemistry

3. To prepare definitions for each new, unique term needed for metabolomics NMR.

FURTHER WORK TO SUPPORT DATA SHARING

The work on ontologies in ENPADASI aimed to structure the creation of ontology for metabolomics in order to support the sharing and analysis of metabolomics data. A dedicated system for sharing of experimental metabolomics data already exists () and the dbNP is also able to contain and share metabolomics data from nutrition research but the ontology to support analysis across uploaded data sets in these systems as part of the DASH-IN infrastructure needs further development. This task is large and should be part of a coordinated effort between ontology and metabolomics experts in future projects.

LITERATURE

Sumner, L., Amberg, A., Barrett, D., Beale, M. H., et al. (2007). Proposed minimum reporting standards for chemical analysis. Metabolomics, 3, 211–221.

Salek, R.M., Neumann, S., Schober, D. et al. (2015) COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. Metabolomics 11: 1587. doi:10.1007/s11306-015-0810-y

APPENDIX 1: DETAILED NMR ONTOLOGY

owl:Thing

- OBI_10
  - data_pre-processing_software
  - spectrum_post-processing_software
  - NMR_software
    - Bruker_NMR_software
      - TopSpin
      - NMR_Suite
      - SampleTrack
      - Paravision
      - AURELIA
      - AUREMOL
      - AMIX_VIEWER_&_AMIX-TOOLS
    - JEOL_NMR_software
      - Delta
    - Varian_NMR_software
      - VNMR
    - Other NMR software
  - NMR software version
- OBI_11
  - NMR_parameter_set
    - NMR_two_dimensional_J-resolved_processing_parameter_set
    - NMR_spectral_projection_parameter_set

NMR_spectrum_post-processing_parameter_set

NMR_quality_check_parameter_set

pre-processing_parameter_set

NMR_data_pre-processing_parameter_set_by_dimension

NMR_data_pre-processing_parameter_set

NMR_instrument_parameter_set

NMR_instrument_acquisition_parameter_set

one_dimensional_NMR_acquisition_parameter_set

two_dimensional_NMR_acquisition_parameter_set

three_dimensional_NMR_acquisition_parameter_set

OBI_20

tilt

quality_check_peak_feature

number_of_scans

number_of_steady_state_scans

atom_environment_encoding

relaxation_delay

x-axis_unit

y-axis_unit

additional_axis_unit

x_start_value

x_end_value

window_function_parameter

line_broadening

line_sharpening

shifted_bell

shifted_sine_bell

gaussian multiplication

sample_introduction_parameter

magic_angle_spinning_rotor

flow_probe

water_suppression_parameter

OBI_3

number_of_data_points

concentration

hadamard_frequency

projection_axis

calibration_reference_shift

ninety_degree_pulse_duration

dwell_time

irradiation_frequency

field_strength

magnetic_field_strength

spinning_rate

decoupled_nucleus

instrument_configuration

NMR_instrument_configuration

one_dimensional_NMR_configuration

two_dimensional_NMR_configuration

three_dimensional_NMR_configuration

reference

pulse_sequence_literature_reference

NMR_sample_reference

file_reference

acquisition_parameter_set_file_reference

free_induction_decay_file_reference

shaped_pulse_file_reference

pulse_sequence_file_reference

processing_parameter_set_file_reference

literature_reference

vendor

NMR_vendor

Bruker

Varian

JEOL

MR_Resources

tecmag

Doty_Scientific

Wilmad

JS_Research

Acorn_NMR_Inc

sample_temperature

sample_temperature_in_autosampler

sample_temperature_in_magnet

OBI_43

complex_datatype

cardinal_part_of_NMR_data_set

one_dimensional_free_induction_decay

two_dimensional_free_induction_decay

NMR_spectrum

projected_spectrum

bucketed_spectrum

peak-picked_spectrum

one_dimensional_spectrum

two_dimensional_spectrum

three_dimensional_spectrum

NMR_data_set

pulse_sequence

one_dimensional_carr_purcell_meiboom_gill

two_dimensional_total_correlation_spectroscopy

one_dimensional

two_dimensional_J-resolved

two_dimensional_hadamard_total_correlation_spectroscopy

one_dimensional_diffusion_edited

OBI_47

optional_part_of_NMR_instrument

autosampler

Bruker_autosampler

NMR_Case

Bruker_BEST_NMR

SampleJet

B-ACS

JEOL_autosampler

ASC24

ASC30

cardinal_part_of_NMR_instrument

acquisition_computer

NMR_probe

Bruker_NMR_probe

CryoProbe

high_resolution_magic_angle_spin_probe

solid_magic_angle_spinning_probe

1mm_MicroProbe

flow_high_resolution_probe

High_resolution_probe

HR_Probes_with_ATM

Micro_Imaging_probe

JEOL_NMR_probe

CapNMR_Probe

liquid_NMR_probe

solid_NMR_probe

tecmag_EAGLE_probe

NMR_imaging_probe

NMR_cryo_probe

console

- tecmag_console
  - DISCOVERY
  - APOLLO
- tube
- NMR_magnet
  - Bruker_NMR_magnet
    - UltraShield
    - Ultrastabilized
    - US_2
    - UltraShield_Plus
- NMR_sample_holder
  - Bruker_MATCH

NMR_instrument

- Bruker_NMR_instrument
  - AVANCE_II_spectrometer
  - Metabolic_Profiler
  - Hyphenation
  - Capillary_LC-NMR
  - AMX
  - AC
- JEOL_NMR_instrument
  - ECX
  - ECA
- Varian_NMR_instrument
  - MERCURY

INOVA

UNITY

VXR

GEMINI

tecmag_NMR_instrument

CAT

NMR_tube_washing_system

Bruker_AutoClean

SampleRail

OBI_50

NMR_plattform

Object

chemical_compound

RealizableEntity

method

algorithm

post-processing_algorithm

analysis

NMR_analysis

data_processing_method

data_post-processing

NMR_spectral_post-processing

data_pre-processing

NMR_data_pre-processing

data_transformation

fourier_transformation

non-fourier_transformation

normalisation_strategy

spectral_denoising

baseline_correction

zero_order_phase_correction

first_order_phase_correction

symmetrise

spectrum_post-processing

peak_picking

bucketing

encoding_method

quantum_momentum_field_gradient

'time_proportional_phase incrementation'

states

'states-time_proportional_phase incrementation'

Hadamard_encoding

radon_encoding

Frydman_encoding

echo_anti-echo

projection_method

G_matrix_fourier_transform_projection

maximum_intensity_projection

summation_projection

sample_introduction_method

water_suppression_method

post-acquisition_water_suppression_method

hankel_singular_value_decomposition

convolution

polynomial_fitting

WaveWat

presat

water_suppression_enhanced_through_T1_effects

nuclear_overhauser_and_exchange_spectroscopy-presat

watergate

flip-back_watergate

excitation_sculpting

coherence_pathway_rejection

jump_and_return

jump_and_return_1-1

jump_and_return_1-3-3-1

window_function_method

exponential_multiplication

gaussian_broadening

sine

sine_squared

quality_control_method

quality_check

NMR_quality_check

?result

quality_check_result

NMR_quality_check_result

Role

buffer

calibration_compound

chemical_shift_standard

quality_check_compound_signal

NMR_solvent

acquisition_nucleus

1H

2H

11B

13C

17O

19F

29Si

31P

15N

sample

NMR_sample