Deliverable D1.2.3

Project Title:	Developing an efficient e-infrastructure, standards and data-flow for metabolomics and its interface to biomedical and life science e-infrastructures in Europe and world-wide	
Project Acronym:	COSMOS	
Grant agreement no.:	312941	
Work Package 7 description	FP7-INFRASTRUCTURES-2012-1	
Deliverable title:	COSMOS Project Report	
WP No.	1	
Lead Beneficiary:	1:EMBL-EBI	
WP Title	Management	
Contractual delivery date:	01 04 2014	
Actual delivery date:	01 04 2014	
WP leader:	Christoph Steinbeck	EMBL-EBI
Contributing partner(s):	*Reza Salek and Christoph Steinbeck*	

Authors: Reza Salek, Christoph Steinbeck

Contents

1 Executive summary

The aim of this deliverable is to summarise up to 18 months activity of WP1 on management and coordination with COSMOS partners:

1. Report on management activity
2. Report on coordination activity
3. Report outreach and dissemination activity mainly on workshops, external meetings and conferences related to COSMOS
4. COSMOS partner meetings and workshops

2 Project objectives

With this deliverable, the project has reached, or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	Report on COSMOS management activities	X	
2	Report on COSMOS coordination activities	X	
3	Report on COSMOS workshops and conferences	X	
4	Report on COSMOS outreach and Dissemination	X	

3 Detailed report on the deliverable

3.1 Background

This work package will provide the management infrastructure for the proposed work and coordination activities. It will make use of the existing electronic communication platforms of the Metabolomics Standards Initiative and the Metabolomics Society, and further develop them, in order to be used by the COSMOS consortium. We will also organize the annual COSMOS consortium and stakeholder meetings, as well as regular workshops and staff exchanges between the COSMOS partners. We will systematically document the decision-making process and decisions made in teleconferences, meetings and by email exchange. This will be compiled regularly into COSMOS consortium documentation.

3.2 Description of Work

3.2.1 Report on COSMOS management activities

To date, as part of this deliverable, we have organized monthly teleconference meetings using Google hangout with the COSMOS WP leaders. Discussions and decisions were minuted using a Google Document that could easily be shared within the SC participants with a final and a second copy of outcomes stored on the COSMOS website Internal pages (COSMOS-FP7.EU), distributed among WP partners and announced via social media.

The second COSMOS partner annual meeting has been scheduled for 25-26 of September in Leucorea Wittenberg, Germany to be hosted by Drs Steffen Neuman (IPB-Halle) and Dirk Walter (Glom, MPG). The draft Schedule is as follow:

Wednesday, 24.9.2014

Afternoon/Evening: Participant arrival

Thursday, 25.9.2014

8:45 Welcome
9:00 Morning session I: COSMOS - what happened in the past months
10:30 (Coffee)
11:00 Morning session II: Stakeholder voices
12:30 (Lunch + coffee + hallway-session)
14:00 Afternoon session I:
15:30 (Coffee)
16:00 Afternoon session II: ToDos and discussions from the WP leads
18:30 End of official part day I
19:00 Workshop dinner

Friday, 26.9.2014

9:00 Morning session I: ToDos and discussions from the WP leads
10:30 (Coffee)
11:00 Morning session II: Next Periodic Reporting preperation
12:00 (Lunch)
13:00 Afternoon session I:
15:00 End of official workshop

Meeting Aims: Within COSMOS consortium we would plane the next 12-24 months of work activity within different WP and different subgroups, discuss the requirements towards the deliverables and general overview of the work. In addition, we will asses results of workshops carried out and discuss about current or future planed staff exchange between partners. We will carry out discussions on past and upcoming stakeholder meetings, finalizing the potential attendees and starting the invitation process, while creating a tentative plan for the stakeholders meetings. Other discussions would be agreement on time and location of the annual meeting in 2015.

The COSMOS stakeholder 2014 meeting was held on April 2nd at EBL-EBI, UK as planned, under the name of “Meeting on International Data Exchange in Metabolomics”. This meeting was also Co-sponsored by the Metabolomics society, particularly data standards task group.

Scope of the meeting:

- Review current and planned database infrastructure across globe

- Agree on scope for data exchange between Metabolomics Workbench, MetaboLights, MassBank, HMDB, mzCloud and other metabolomics DBs. Agree on data formats, notification mechanisms and distribution workflows
- Agree on distributing the load on curation. (see below)
- Decide the name of the new network, and plan a formal consortium agreement or memorandum of understanding
- Plan future activities

Meeting Agenda

April 1st:	
19:00	Drinks and Dinner (Red Lion, 2 minutes walk from Campus)
April 2nd:	**Courtyard Room, EBI main building**
09:00	Welcome on behalf of Metabolomics Society & COSMOS, and round table introduction (Chris Steinbeck and Mark Viant)
	Introductions by "major players" in international database provision with a focus on geographic coverage, scope, services and types data? **Which services, what data?**
09:10	Saravanan Dayalan, representing Australia and New Zealand
09:35	Masanori Arita, representing Japan
10:00	Shankar Subramanian, representing US
10:25	David Wishart, representing Canada
10:50	**Coffee Break**
11:10	Christoph Steinbeck, representing Europe
11:35	Mechanisms, protocols and scope of data exchange between the partner databases. Import and export formats, metadata, federated data vs duplication.
12:30	**Lunch Break and Tour of the Campus**
14:00	Reza Salek, Update on format standards
14:20	Distribution of curation tasks (plenty of reference data on

	metabolites need to collected and annotated. We should not duplicate those tasks across the continents and coordinate the curation)
15:00	**Coffee Break**
15:30	General Discussion. Naming the consortium. Composition of a consortium agreement/MoU (draft prepared and circulated beforehand by Chris).
17:30	Roundtable planning for future meetings (face-to-face, teleconf etc.) and future funding for our network (research on funding opportunities by participants beforehand)
18:30	End of Work
19:00	**Drinks in Bar at Hinxton Hall**
19:30	**Dinner in Pompeian Room at Hinxton Hall**
April 3rd	Breakfast and Departure

The participant list: Merlijn van Rijswijk (NMC), David Wishart (Univ. Alberta), Dirk Walter (MPMPI), Joachim Kopka (MPiMPI), Lloyd Sumner (Noble Fnd), Susanna Sansone (Un. of Oxford, Nature Publishing Group), Leslie Derr (NIH), Christoph Steinbeck (EBI), Masanori Arita (Nat. Inst. Genetics/Mass Bank), Phil Smith (NIH), Shankar Subramanian (UCSD), Oliver Fiehn (UC Davis), Rick Dunn (Univ. Birmingham), Mark Viant (Univ. Birmingham), Roy Goodacre, (Univ. Manchester), Art Castle (NIH), Ken Haug (EBI), Reza Salek (EBI), James Smith (MRC-HNR). Saravanan Metabolomics (Metabolomics Australia).

Presentation: Mark Viant gave an introduction to recent developments in the Metabolomics Society. Various task groups formed, Society now supported by ASK, a professional organization company. Funding and manpower is now available for running meetings. Metabolomics Society is also keen to facilitate the international coordination of data standard efforts.

Saravanan Dayalan - Metabolomics Australia: Gave an overview of the Australia metabolomics network. Presented ANZMN a community interchange platform for metabolomics and lipidomics. Presented an overview of the current

infrastructure: MASTR-MS (LIMS solution), MAMBO-MS (central metabolite database), SHELF pipeline (GC-MS data processing tool,) MAR (statistical analysis toolkit) and iQC (instrument QC monitor) within the Australian metabolomics community. Saravanan pointed out the challenges in their curation capacity for MAMBO-MS database and need for community effort. The future plans are to integrate ISA-TAB into MASTR-MS, adoption of mzTab file format into MAMBO-MS, PyMS and MASTR-MS. MAMBO-MS database contents will also to be added and to become part of MassBank database.

Masanori Arita - Japan: Gave an overview of MassBank database - http://www.massbank.jp/ or http://www.massbank.eu/ - and its future. MassBank started in 2006 and now has about 40000 standard spectra for 15000 compounds from 27 research groups. Massbank has two servers, one in Japan and one in Germany. At policy level there are issues with the consents and limitation on who can use which data set and for what purpose has to be defined, before using the data. **Bio-massbank** houses bios-samples (such as; cells, tissues, plants microbes) with spectral annotation based on wiki data entry currently at http://bio.massbank.jp. Also Massanori gave overview of other metabolomics resource in Japan such as **metabolonote** for metadata (http://metabolonote.kazusa.or.jp/Main_Page; Kazusa and RIKEN) and **Prime** for raw data (Kazusa and RIKEN).

Shankar Subramanian representing the US: Gave an overview of the current state of NIH common funds effort, including descriptions for each 6 Regional Comprehensive Metabolomics Resource Cores (RCMRC). NIH currently has two funded initiatives for Metabolomics Standards Synthesis. Metabolomics Workbench uses a cloud-based resource at the San Diego Supercomputer Centre. They have successfully used the Aspera protocol for bulk uploads of large data sets. For curation and metadata capturing they are using Excel-Spread sheet templates and store the results as Tab-delimited files similar format to ISA-tab. Metabolomics Workbench now open to upload and access for the whole metabolomics community. Checklists, for upload criteria to Metabolomics Workbench are now available online at http://metabolomicsworkbench.org/.

David Wishart representing Canada: Gave an overview starting from DrugBank in 2004 and HMDB up to more recent one for Biofluids such as; saliva, blood plasma and urine metabolomics based resources in Canada. Drugbank, HMDB and YMDA (yeast) are ranked among highly accessed resources from Canada. HMDB includes data from chemical concentrations, mostly quantitative concentrations as well as their dynamic range, collected based on experiments. David also presented tools for chemical taxonomy (automated web based tool Classyfire) matching the largest substructure with a compound used to classify it. Also joint effort between COSMOS and his group on developments on "nmrML" file format was presented.

Reza Salek Data Exchange Standards: Including COSMOS WP2, overview about ISA-Tab format used to capture biological metadata. State of fileformat development for mzML and mzTab. The mzTab format is for processed data, biologist friendly and that there are potential to link mzTab to libraries and methods used to annotate or identify metabolites. Metabolite annotation/identification methods and processes that need to be captured via the proposed format with link to databases and identifiers were discussed. Overview of GC-MS vendor File format support and format converters (ProteoWizard and Maltcms) supporting now major vendors. State of NMR data exchange standards: nmrML, nmrTab. Particularly recent developments in nmrML open NMR format.

Outcomes: There were set of agreements and proposal out of meeting discussion in summary the most important ones are:

- Participants agree that a federated model, where partner databases export a minimal (to be agreed upon) set of metadata to a meta-database called **MetabolomeXchange**
    - Setup MetabolomeXchange website.
    - Reserve MetabolomeXchange.org (.net, .eu)
    - Community area for documentation (Wiki-style)
    - Create a section for metabolomes that certain groups (species or disease based) that are interested in curating
    - Actual search area for the meta-data exchange

- Partner databases still will import trimmed-down study data from each other in joint formats and methods for making their data accessible in community-agreed ways
    - To Come up with a dozen user cases, documented on the MetabolomeXchange website, with challenges highlighted
    - Member to come up with agreed minimal set of data (and meta-data) that each partner database should provide to the end user, and to agree upon the formats
    - Plan to develop tools toward database interoperability
    - Explore potential issues with Clinical Data and propose solutions. Perhaps check ethical implications during study design and historic data issues.
- Export of study data for end users in agreed format (i.e. mzTab format use case)
- To promote usage of open standard exchange formats for information on individual metabolites
- **Participants agree that we need a concerted effort for exhaustive characterization of compounds with NMR and MS data under an agreed set of conditions.**
    - This require further agreements on how to manage the physical data collection and distribution
    - Create a list of compounds that are sufficiently characterised.
    - Create a list of compounds available for measurement (commercially available: eMolecules, compounds in freezers or in labs)
    - Create a list of instrumentation available at the partner sites
    - Plans to distribute and measure compounds
    - Do a webinar on the deposition of putative compound identification in MetaboLights archive layer

Finally a draft proposal for a MetabolomeXchange collaborative agreement was proposed with further discussed to be agreed upon. MetabolomeXchange (MX) is an international collaboration of metabolomics data repositories that handles public submissions ('Database Providers'). This data exchange covers both the exchange of study data as well as reference data for individual metabolites. All

Database Providers pledge to provide primary data and metadata according to the set guidelines. These guidelines are not intended to be legally binding, but to define the mode of interaction among collaborating Database Providers.

To future continue with the discussion and effect coordination a Google user group was setup:

https://groups.google.com/forum/#!forum/MetabolomeXchange/

3.2.2 Report on Coordination effort with other WP

Lead by **WP4** the COSMOS consortium, with feedback from the metabolomics community have worked towards MSI implementation of data flow. The usage of existing MSI minimum information reporting for data sharing and experimental protocols was summarized in an editorial by WP4 (Goodacre 2014) which listed all community-agreed MI papers relevant in the metabolomics domain (D4.2). Some of these MI standards have also been solicited by major journals to enforce policies as part of publishing *best practices* including now *Metabolomics journal (D4.3).* **WP1, WP2, WP3, WP5** and **WP4** including many partners in the initiative, that also serve as members of the editorial boards of relevant metabolomics publications have now agreed to make MetaboLights recommended by these journals as a data repository for metabolomics experiments. Examples of these journals are; *Metabolomics, Metabolite, Nature Scientific Data (also accepting ISA-Tab format)* and *Gigascience journal (also accepting ISA-Tab format).*

An example text from *Metabolite: "Deposition at MetaboLights Database:* Metabolites supports the deposition of metabolomic data to public, open access databases such as EMBL-EBI's MetaboLights (http://www.ebi.ac.uk/metabolights). Authors are strongly encouraged to submit all supporting data, where practicable, to such a database. Data should comply with the relevant communities annotation agreements such as the Metabolomics Standards Initiative (MSI) and COSMOS (http://cosmos-fp7.eu/) recommendations prior to manuscript submission. Authors are encouraged to include database accession numbers or a statement of data communication to the repositories in their cover letter and manuscript."

As the editor of the *Metabolomics* journal states in his editorial "*Metabolomics* [the journal] is already on record in saying that it wishes studies that it publishes and data therein to be as MSI compliant as possible (Goodacre 2010)". Whilst established procedures are not yet in place for metabolite data upload within an MSI compliant decentralised framework, COSMOS is encouraging researchers to deposit their data to one of the above repositories; MetaboLights or Metabolomics Workbench. Meanwhile, MSI compliant standards of data annotation, reporting, management and flow are being developed, promoted and entrenched so that those can be shared routinely and re-used effectively (D4.3).

As part of the resubmitted deliverable 4.1 after agreements with international data exchange meeting in April held at EMBL-EBI, we have now proposed new guidelines for data deposition workflow between participating and potential metabolomics databases and repositories. This will ensure a coherent metabolomics workflow to run to its full potential, capturing agreed sets of metadata across different resources. The workflow definitions will prioritise simplicity, usability, annotation quality and the plurality of metabolomics resources and databases to ensure a coherent connectivity between similar studies and to provide rapid matching results to end users. In collaboration with stakeholders, member of metabolomics society, publishers and partners, appropriate strategies for the sustainability of the data deposition workflow are also being discussed (Details in Deliverable 4.4).

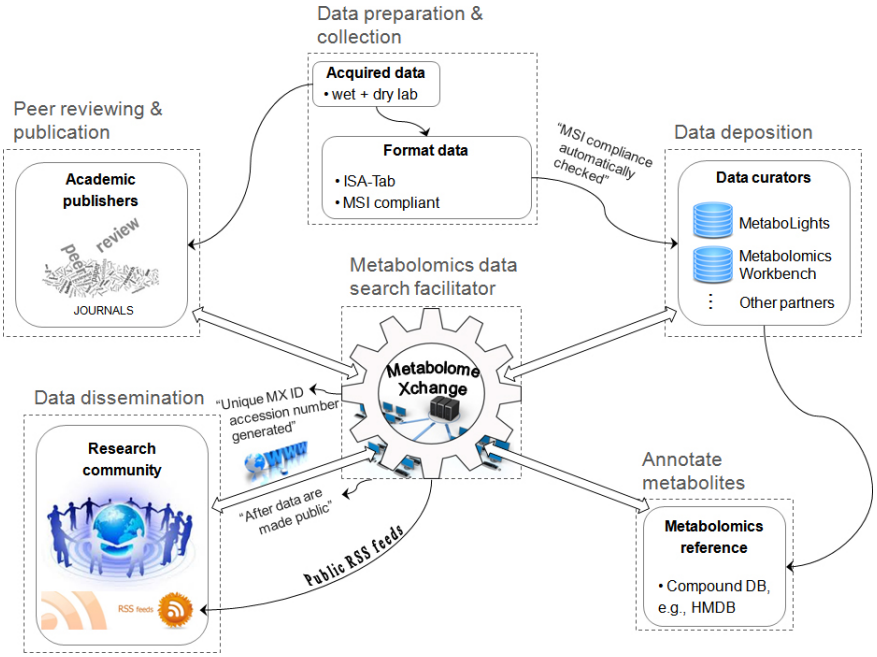

Figure 1: Current COSMOS data deposition workflow model.

In addition, in COSMOS data sustainability will be based on the EBI data sustainability policies and the ELIXIR project (http://www.elixir-europe.org/) whose objective is to provide the facilities necessary to store and share data for life by building a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.

WP2: WP2 with feedback from partners and stakeholder involved in developing nmrML format continue to progress. We have now coordinated efforts from multiple international groups who are working in NMR and metabolomics related software to design, adopt and establish the nmrML data format, based on the experience with the PSI mzML format for mass spectrometry. The nmrCV.owl ontology momentarily contains ~ 600 classes under NMR namespace. Around 2000 terms are imported from the units ontology and BFO top level ontology. The nmrML group also maintains a pure taxonomy without use of axiomatic definitions. Example

NMR data file in nmrML format are continuously added to the website http://nmrml.org/examples/.

WP3 and **University of Bordeaux** together improved and established software infrastructure that was able to conveniently capture standardize experimental metadata. This software infrastructure was built based on the existing XEML-lab software suite that has graphics-supported description of experimental designs along with standardized description of experiment conditions via ontologies. XEML-labs was expanded to facilitate data import and export as well as integration to existing databases. In particular, the software suite is not capable of export in ISA-tab format, which makes seamless submission of metabolomics data to MetaboLights data repository possible.

WP5: With the consortium partners we have proposed a guideline for data deposition workflow between participating metabolomics databases and repositories, having established a collaboration link with the NIH common fund databases. This work ensures a coherent metabolomics workflow to run to its full potential, capturing agreed sets of metadata across different resources. The technical side for such web portal is now being dvelped under the name of MetabolomeXchange. The core of “MetabolomeXchange” is MongoDB, a Document database (https://www.mongodb.org/). In this database there are 2 collections:
"providers": to store the data repositories providing the data
"datasets": with a subset of the meta data of all the individual datasets stored as a document. Currently, the test database has been setup and hosted at https://www.mongosoup.de for free. This is WP is progress well towards its goals.

WP6: Through working with WP6 we were able to optimize communications with BioMedBridges and with its partner BMS

infrastructures with an interest in metabolomics. Within this working group we have now established a joint consensus document between COSMOS and BioMedBridges. First such metabolomics data set are now in process of submission to MetaboLights.

3.3. Report COSMOS out reach and dissemination activity

3.3.1 Report on COSMOS Workshops and Partner Exchange

The Conferences attended or organised by EMBL-EBI

- COSMOS and MetaboLights workshop Imperial College 22 January Including presentation and hand on submission to MetaboLights. Attendees from Imperial College metabolomics group about 10 people.

- Cambridge University, Department of Genetics; Bioinformatics: Introduction to Metabolomics, metabolomics standard and databases.
    - Graduate School of Life Sciences, Mon 20 Jan 2014

- Participation in NIH common fund Data Sharing Work Group Meeting 5-8th of February, San Diego USA.

Agenda:

- Overview of the current metabolomics data/metadata infrastructure.
- Discussion of the key metadata items required for translation of experimental studies to online formats, which may be searched and analyzed
- Discussion of conversion of different data types to formats suitable for transfer to the DRCC
    - Raw data (NMR,MS, etc)
    - Analytical metadata (including protocols),
    - Associated biological and clinical data
- Final result datasets of quantitative/semi-quantitative metabolite values and appropriate substance identifiers.
- Discussion on data upload issues
- Discussion on online data presentation, search tools and statistical methods. (http://www.metabolomicsworkbench.org)

During this meeting we were able to have one-on-one discussions regarding COSMOS formats and data sharing as well as understanding the bioinformatics infrastructure of both project. We also agreed on next steps of collaburation in future.

- EMBL-EBI Agricultural-Omics course 17-21 February: This course was setup to provide an overview of data resources, tools and analysis pipelines for 'omics datasets within the agricultural sciences. It will be delivered using a mixture of lectures, practical computer sessions and discussions based on agricultural case studies and public domain data. Additionally there will be opportunity to review the key opportunities and challenges within this field. On day five we provide and introduction and hands on metabolomics databases, resource and standards, course was lacated and EMBI-RBI Cambridge UK. http://www.ebi.ac.uk/training/course/agricultural-omics

- **RIKEN-MPI Joint Workshop On Forefront of Metabolomics and Future Prospect RIKEN Japan 26 Feb 1 March** .Presenting MetaboLights: Metabolomics data repository and the role of COSMOS" - COordination Of Standards In MetabOlomicS.

- **A workshop and hand on MetaboLights and COSMOS and Tokyo University 1 March** presentation to participant and discussions on data standards.

- **First mzTab for metabolomics workshop, March 6-7 2014 -** COSMOS members together with the PSI community are holding the first “mzTab for metabolomics” workshop in Tübingen, Germany. Our aim is to drive reporting of metabolomic results further using a standardized, open, easy accessible and human readable tabular format. MzTab (version 1.0) already provides basic support for reporting small molecules that we plan to extend and harmonize with the more advanced reporting scheme for proteins and peptides available in MzTab.

- EMBO Practical Course on Metabolomics Bioinformatics for Life Scientists 17- 21 of March , 2014 http://www.ebi.ac.uk/training/course/embo-practical-course-metabolomics-bioinformatics-life-scientists

COSMOS Workshop and partner exchange

First mzTab for metabolomics workshop, March 6-7 2014

Agenda:

Date & Time: Thursday, March 6th 2014 from 9:00-17:00
Location: University of Tübingen, Sand 1

Agenda Thursday, March 6th 2014
9:00 Welcome and introductions
A brief history of mzTab [Juan Antonio, EBI]
Morning session I: Challenges and requirements for mzTab in metabolomics
[Chairs: Christoph Steinbeck, Oliver Kohlbacher]
Requirements from a quantification point of view
Requirements from an identification point of view
Requirements from a repository point of view
10:30 Coffee break
11:00 Morning session II: mzTab structure for metabolomics
[Chairs: Juan Antonio, Timo Sachsenberg]
Discussion on small molecule sections similar to proteomics
Discussion on summary vs. complete identification and quantification mzTab files
Discussion of required/optional meta values / columns in the four different files
12:30 Lunch + coffee + hallway session
13:30 Parallel afternoon sessions I

Quantification and Ion species annotation [Emma Schymanski and] Room 311	Structure Identification reporting [Steffen Neumann and] Here
Discussion of quantification Discussion of ion species annotation Columns “Modification/derivatisation” and/or “Adduct/Ion species”	Discussion of structure/database reporting of identification results

15:30 Coffee break
16:00 Afternoon session II: Software support and standardisation process
[Steffen N. + Timo S. + Oliver K.]
What’s required in OpenMS, XCMS, maltcms, Metabolights, ISAcreator plugin,
support libraries ?, ...

Further standardisation process and publication roadmap
Wrapup and assignment of Todos
17:30 End of official workshop
19:00 Workshop dinner downtown (not included)

List of participant:

Reza	Salek	EMBL-EBI
Christoph	Steinbeck	EMBL-EBI
Sven	Nahnsen	Quantitative Biology Center, Tübingen
Michael	Witting	Helmholtz Zentrum München
Jackson	Pope	Nonlinear Dynamics Ltd.
Ian	Morns	Nonlinear Dynamics (Waters)
Jürgen	Hartler	Institute for Genomics and Bioinformatics, TU Graz
Emma	Schymanski	Eawag
Nils	Hoffmann	Genome Informatics, Bielefeld University
Timo	Sachsenberg	Universität Tübingen
Simon	Perkins	University of Liverpool
Michael	van Vliet	LACDR/Leiden University
Kenneth	Haug	EMBL-EBI
Fabian	Aicheler	Universität Tübingen
Mathias	Walzer	Universität Tübingen
Marc	Rurik	Universität Tübingen
Steffen	Neumann	IPB Halle
Sebastian	Kusch	Thermo Fisher Scientific
Mika	Hilvo	VTT Technical Research Centre of Finland
Heli	Nygren	VTT Technical Research Centre of Finland
Juan Antonio	Vizcaíno	EMBL-EBI
Oliver	Kohlbacher	Universität Tübingen

3.3.2 Report on COSMOS in news and media

Information on COSMOS that has appeared in news articles:

MetaboNEWS

This newsletter is published in partnership between The Metabolomics Innovation Centre (TMIC, http://www.metabolomicscentre.ca) and the international Metabolomics Society (http://www.metabolomicssociety.org), and is intended to keep metabolomics researchers and other professionals informed about new technologies, software, databases, events, job postings, conferences, training opportunities, interviews, publications, awards, and other newsworthy items concerning metabolomics.

We now have a dedicated section in MetaboNews for monthly updates on Status of Data Standards: This new section within the Metabolomics Society News will be contributed regularly by Christoph Steinbeck (Chair of the Society's Data Standards Task Group) and Reza Salek from the EMBL-EBI, Cambridge UK.

- MetaboNew Status of Data Standards Archivew

    - **December issue: Status of Data Standards**: The **CO**ordination of **S**tandards in **M**etab**O**lomic**S** (COSMOS) consortium (http://www.cosmos-fp7.eu) together with Proteomics Standards Initiative (PSI) is working towards completing the existing **mzTab** format (http://www.psidev.info/) to better capture small molecules data and meta-data. mzTab can be used for **reporting** both **metabolite identification** and **metabolomics quantification**. After extensive research and discussions within the consortium and members of the MSI and PSI community, we concluded that mzTab is the medium of choice for capturing and reporting such metabolomics results. To test and evaluate the standard, the mzTab development is accompanied by early implementations, e.g., in development versions of the OpenMS (http://open-ms.sourceforge.net/), or separate export functions for XCMS and CAMERA (http://www.bioconductor.org/) software tools. In addition, the MetaboLights metabolomics database (http://www.ebi.ac.uk/metabolights) **accepts the quantification and identification of metabolites in a subset of mzTab,** soon to be updated to the full compatible version, once the final discussions on the mzTab format for small molecules are completed. We are planning to have an **"mzTab for metabolomics" workshop in Tübingen** (in March 2014) **for discussion and work on metabolomics extensions for mzTab**. We hope to adopt and finalize the format and eventually submit it to the community for review. The workshop is organised jointly by Steffen Neumann (Leibniz Institute of Plant Biochemistry, IPB Halle) and Oliver Kohlbacher (Tübingen University), one of the leading members in the mzTab development in the PSI. For details, please visit http://cosmos-fp7.eu/mzTab.
    - **January issue: Status of Data Standards:** One of the aims of the COordination of Standards in MetabOlomicS (COSMOS) consortium (http://www.cosmos-fp7.eu) is to gather metabolomics requirements for other major e-infrastructures such as BioMedBridges (http://www.biomedbridges.eu/), BBMRI (http://bbmri.eu/), ELIXIR (http://www.elixir-europe.org/) and EU-OPENSCREEN (http://www.eu-openscreen.de/). The University of Florence as a third party of CIRMMP (http://www.cerm.unifi.it/about-cerm/cirmmp) coordinates the gathering and requirements of metabolomics data with the above-mentioned e-infrastructures. Coordination with BioMedBridges and biomedical ESFRI infrastructures (the European Strategy Forum on Research Infrastructures), aims to foster their co-operation and interaction with COSMOS. The interaction with BioMedBridges

occurred through the participation of COSMOS delegates at meetings and workshops organized by the BioMedBridges. Strong interactions between COSMOS and BBMRI have been established to coordinate efforts. For this reason, Kurt Zatloukal, coordinator of BBMRI has been nominated among the Advisory Board of COSMOS. The ELIXIR Hub will be connected to ELIXIR Nodes to provide infrastructure for data, computing tools and standards as well as training and support for the ESFRI biological and medical science infrastructures. The link from COSMOS to ELIXIR is via MetaboLights (http://www.ebi.ac.uk/metabolights/); an open access repository housing metabolomics-based experiments. The coordination outcome will be published on the COSMOS (http://www.cosmos-fp7.eu) website.

- **February and March issue: Status of Data Standards:** Development of nmrML format: Currently, the most widely used data exchange format for NMR data is JCAMP-DX version 6.0 by the Joint Committee on Atomic and Molecular Physical Data (Davies and Lampen, 1993), but the specification is not very rigorous and many different flavors exist in the wild, which can lead to incompatibilities between different software packages. It is also not easily extendable to capture supplementary information. The MSI workgroups have provided detailed suggestions about the minimum information metadata to be captured for a NMR experiment. In particular, the MSI, had put forth recommendations to report instrument descriptions and configurations, instrument-specific sample preparation and data acquisition parameters (Rubtsov, Jenkins *et al.*, 2007), which resulted in a first round of NMR XML data standard development, focusing on raw and processed one- and two-dimensional NMR experiments and associated metadata (Ludwig, Easton *et al.*, 2012). Inspired by the huge success of mzML in mass spectrometry, the COSMOS COordination Of Standards In MetabOlomicS (http://cosmos-fp7.eu) consortium has joined forces with other groups and has now merged and adopted existing schemata into a new nmrML format (http://nmrml.org). The format consists of the XML schema that defines the structure of an nmrML file. This structure is deliberately kept simple to ease the task of implementation, and avoid the need for frequent changes when the terminology needs to accommodate upcoming new technologies and parameters. Instead, these will be annotated in the nmrML file using the second component of nmrML, the controlled vocabulary terms from the nmrCV ontology. The nmrCV is based on earlier work at the EMBL-EBI (Sansone, Schober *et al.*, 2007) and efforts at The Metabolomics Innovation Centre (David Wishart Group). The nmrCV contains nearly 600 terms and partly relies on external sources like ChEBI for chemical information, thus making it an integrative resource. Term request can be channeled through the issue tracker/mailing list. We also provide early prototypes for file converters from vendor formats to nmrML, as well as parser

libraries for Java, R and python, which can be used by open NMR processing and analysis software. The development of nmrML is taking place on Github (https://github.com/nmrML/nmrml), where the specification documents, more detailed descriptions of our use cases, examples files and the parser libraries can be found. We are now providing a first nmrML release candidate at http://nmrml.org for public consultation and feedback.

- **April issue: Status of Data Standards: Joint COSMOS and HUPO PSI Meeting:** In April 13-16 the COSMOS (COordination Of Standards In MetabOlomicS) is planning to participate and to have a joint meeting with the proteomics HUPO Proteomics Standards Initiative community. This meeting will take place in Schloss Reinhartshausen Kempinski, Nr Frankfurt, Germany. HUPO-PSI has defined community standards for data representation in proteomics and has facilitated data comparison, exchange and verification within the proteomics community. Many open source MS formats including: mzML, mzTab, mzIdentML and mzQuantML as well as guidelines for minimum information reporting for proteomic and peptide identification have been developed within this initiative. Working closely with the HUPO-PSI community should benefit the metabolomics community, particularly the COSMOS effort in development of open MS exchange formats for metabolomics. We also hope to contribute to the development of the MS-based controlled vocabulary by PSI-MS by adding the metabolomics community ontology requirement and terminology artifacts. More details can be found at: http://cosmos-fp7.eu/PSI and http://www.psidev.info/psi2014

We have regularly updated and promoted COSMOS via our several social media sites, including:

- Blogger - http://metabolights.blogspot.co.uk

- Twitter - #cosmosfp7

- Facebook - http://www.facebook.com/cosmosfp7

3.4 Next steps

1. Submission of technical manuscript on the new nmrML open standard file format.
2. Workshop plan for the EMBO practical course on Metabolomics 2015 in, EMBL-EBI
3. Schedule Stakeholders meeting at Metabolomics 2015

4. Organising the annual meeting of partner for 2015
5. Carrying out several COSMOS workshops nationally and internationally to promote standards in metabolomics
6. Coordination of development and dissemination of MS XML formats; mzML, mzIdentML, mzQuantML and mzTab
7. Coordination of development and dissemination of NMR XML formats; nmrML, nmrIdentML, nmrQuantML and nmrTab
8. Coordination on developments of tools, convertors and API for nmrML and mzML file formats
9. Coordination and dissemination of ontology and CV development
10. Interaction with vendors, software developers, Journals and databases to make COSMOS more inclusive and economically viable for them to participate in the development of the file formats, essential for the success of the initiative.

4 Publications

1. Salek RM, Steinbeck C, Viant MR, Goodacre R, Dunn WB. **The role of reporting standards for metabolite annotation and identification in metabolomic studies**. Gigascience. 2013 Oct 16;2(1):13. [Epub ahead of print] PubMed PMID: 24131531.

2. Kirsten Gracie, Elon Correa, Samuel Mabbott, Jennifer A. Dougan, Duncan Graham, Royston Goodacre and Karen Faulds. "**Simultaneous detection and quantification of three bacterial meningitis pathogens by SERS**". *Chem. Sci.*, 2014, 5, 1030.

3. Royston Goodacre, "**Water, water, every where, but rarely any drop to drin**k" Metabolomics February 2014, Volume 10, Issue 1, pp 5-7

5 Delivery and schedule

The delivery is delayed: ☐Yes ☑No

6 Adjustments made

N/A

7 Efforts for this deliverable

Institute	Person-months (PM)		Period
	actual	estimated	
1:EMBL-EBI	2	2	18
2:LU/NMC	1		18
Total	3		

Appendices

1. N/A

Background information

This deliverable relates to WP1; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP1 Title: Management
Lead: Christoph Steinbeck, EMBL-EBI
Participants: Christoph Steinbeck

This work package will provide the management infrastructure for the proposed work. It will make use of the existing electronic communication platforms of the Metabolomics Standards Initiative and the Metabolomics Society, and further develop them, in order to be used by the COSMOS consortium. We will also organize the annual COSMOS consortium and stakeholder meetings, as well as regular staff exchanges between the COSMOS partners.

We will systematically document the decision-making process and decisions made in teleconferences, meetings and by mail exchange. This will be compiled regularly into COSMOS consortium documentation.

Work package number	WP1	**Start date or starting event:**			Month 1			
Work package title	Management							
Activity Type	COORD							
Participant number	1: EMBL-EBI	2: LU/NC						
Person-months per participant	12	6						

Objectives
The consortium management activities will include

1. Coordination at consortium level of the 'technical' activities of the project.
2. The overall legal, contractual, ethical, financial and administrative management of the consortium.
3. Co-ordination of knowledge management, IPS and other innovation-related activities.
4. Preparing, updating and managing the consortium agreement between participants.
5. Maintaining communications with the Commission.
6. Overseeing the promotion of gender equality in the project.
7. Overseeing science and society issues related to the activities conducted within the project.

Description of work and role of participants

It is in the very nature of a coordination action to focus on communication between the participants for the sake of policy making, to document the outcome and spread the word to promote widespread community adoption.
We therefore wish to highlight the following:

Personal Communication
As part of this work package, we will organize monthly tele-meetings (Skype, phone, webex) of the COSMOS steering committee. Discussions and decisions will be minuted. We will invite international collaborating PI's to participate if needed.
Technical teleconferences of the work package participants will be held more frequently and likewise carefully documented.

Formal Communication
The policies, standards and workflows developed in this endeavour will be formally documented and published in the form of manuals, white papers and

recommendations. Any document created under this umbrella will be released under Creative Commons License to allow for barrier-free dissemination.
At the beginning of the project in month 2 we will deliver a project plan which will include a list of success indicators to monitor during the whole project, as well as the data we will gather that will help in assessing its impact. These indicators and metrics will be subject to change during the first review meeting and they will be reported at least in the annual reports.

Participants
The management work package will be coordinated by the EMBL-EBI, building on EMBL-EBI experience in the management of large consortia, for example in the BioSapiens, Embrace, and Felics projects. The Netherlands Metabolomics Center (LU/NMC) will be co-coordinator with their extensive experience in maintaining the largest national Metabolomics initiative in Europe and networking with an extensive set of international partners. In addition to EMBL-EBI and LU/NMC as the coordinators, all work package leaders are formal participants of this deliverable, due to the higher communication and reporting effort. EMBL-EBI have included the cost for an audit certificate under management subcontracting.

Deliverables

No.	Name	Due month
D 1.1	Project Plan	2
D1.2.1	COSMOS Project Report	6
D1.2.2	COSMOS Project Report	12
D1.2.3	COSMOS Project Report	18
D1.2.4	COSMOS Project Report	24
D1.2.5	COSMOS Project Report	30
D1.2.6	COSMOS Project Report	36