Deliverable D7.3.2

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:	Report on annual stakeholder meetings: Second report	
WP No.	7	
Lead Beneficiary:	13:UBHam	
WP Title	Report on annual stakeholder meetings	
Contractual delivery date:	1 October 2014	
Actual delivery date:	1 April 2014	
WP leader:	Ulrich Günther	UBHam
Contributing partner(s):	UBHAM, EMBL-EBI, UB	

Authors: Reza Salek, Chris Steinbeck, Ulrich Günther.

Contents

1 Executive summary

The aim of this deliverable is to summarize outreach activity by the partners:

1. Report on COSMOS stake holder meetings - update

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 the COSMOS stakeholder meetings	X	

3 Detailed report on the deliverable

3.1 Background

We will initially employ the usual channels for the disseminations of COSMOS standards, including scientific publications, workshops and presentations at metabolomics conferences to reach the wider metabolomics community. The project will plan activities adequately resourced devoted to dissemination for specialised constituencies and the general public, in particular for awareness and educational purposes. The dissemination plan deliverable will consider adequate messages about the objectives of the project and its societal and economic impact.

3.2 Description of Work

The COSMOS second stakeholder 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.

3.2.1 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 memorandum of understanding

In subsequent and follow-up discussion, a memorandum of understanding was prepared and distributed among participant and was agreed upon. This document was not intended to be legally binding, but to define the mode of interaction among collaborating database providers. It was also agreed that initially only the core database providers were selected as members and to have an advisory board that we envision to have representatives of the NIH, of the Metabolomics Society and other selected individuals decided by the MetX members to participate.

Therefore, the initial 7 members of the consortium will be:

1. Leiden University (as the maintainer of metabolomexchange.org)
2. MetaboLights
3. MetabolomicsWorkbench
4. Golm Plant Metabolme Database
5. Bordeaux Metabolomics Database
6. HMDB
7. Massbank

Once finally approved we will update MetabolomeXchange.org with a documentation section listing all of the members, the advisory board as well as the collaborative agreement.

Agreement

MetabolomeXchange collaborative agreement, Version 1.2, updated at 26 September 2014

MetabolomeXchange (MetX) is an international collaboration of metabolomics data repositories that handle public submissions ('Database Providers').
This data exchange covers both the exchange of study data as well as reference data for individual metabolites and metabolomic profiles.
All Database Providers pledge to provide primary data and metadata according to the following guidelines. These guidelines are not intended to be legally binding, but to define the mode of interaction among collaborating Database Providers.

1. Aims

1.1. To provide a network of stable, coordinated, freely accessible metabolomics data from repositories that handle public submissions.
1.2. To jointly make all published metabolomics research data easily accessible for the scientific as well as commercial user community.

1.3. To provide easy access to metabolomic reference data and work together to close the gaps in reference data provision.

1.4 To work closely with publishers, instrument vendors, software developers, data generation facilities, MSI, Metabolomics Society and the user community in the field of metabolomics to promote data accessibility.

2. Membership

2.1. The MetabolomeXchange Consortium consists of database and infrastructure providers: The MetX partner commits to participate in the Consortium, by maintaining a major resource or infrastructure for metabolomics data that fulfils all of the data provisioning guidelines of the Consortium and the conditions of active Database Providership described below.

2.2. The MetX consortium is supported by an Advisory Board. A complete list of advisors is available on the MetabolomeXchange website.

3. Decision making process

3.1 Decisions about procedures and membership will be made by the database providers in the MetabolomeXchange, after consulting all members of the consortium.

3.2. Membership of the the Advisory Board is decided by the MetX partners in an annual consortium meeting.

4. Data provisioning

4.1. Database Providers must implement "public" and "private" data access mechanisms.

Private data access will usually be for pre-publication data, public data access for post-publication data, or if the authors permit pre-publication access. A typical use case for private data access is for reviewers of manuscripts referring to a data set.

4.2. On publication of a manuscript, the associated MetX dataset must become publicly accessible with minimal delay, but no more than 30 days.

4.3. Once released, all data must be and remain fully freely and publicly accessible to all potential user groups, without additional steps like user registration or limitation of access for example only to academic users.

4.4. Database Providers must implement mechanisms for download of datasets to enable re-use of data.

4.5. Database Providers must implement mechanisms to ensure that ethically sensitive data are handled according to national and international privacy protection law.

5. End of Database Providership

5.1. Database Providers may leave the Consortium at any time by notification to the other Database Providers.

5.2. Leaving Database Providers must make all their data records available for import by an active partner database, for a 12-month period following departure, such that they may continue to be made searchable via the MetabolomeXchange portal (changing the underlying URLs). The importing database will then actively maintain these records but will acknowledge the originating database within the record.

5.3. If an active partner is not fulfilling the conditions of active Database Providership, the other Database Providers may vote to issue a formal warning. If the issues remain unresolved after six months, the other active Database Providers may vote to terminate the MetX Database Providership of the partner in question.

6. Steps for joining the Consortium

6.1. The applicants must contact all MetX Database Providers expressing their desire to join the Consortium. They must submit a document including the description of the resource (including the data workflow(s) they would like to support), with URLs to live site, and available resources, in particular curators, as well as backup strategies.

6.2. The MetX Database Providers will review the document and the resource, and ask for further details if needed. The document will be refined through iterations until both the applicant and the MetX Database Providers agree in a final version. If the applicant cannot meet the criteria for joining the Consortium, the existing Database Providers may vote to decline the application.

6.3. In parallel, the applicant must be able to create a MetX XML file for their first planned submission and must send it to all current Database Providers for review.

6.4. Database Providers will discuss the MetX XML file with the applicant and iterate with the applicant to ensure that the document conforms to MetabolomeXchange agreed standards.

6.5. All current Database Providers vote to approve the application according to the voting guidelines. If a vote does not pass, the existing Database Providers must draft a response letter to the applicant that indicates what the path to passing might be or if there is no path to acceptance.

6.6. The applicant is given test access to the Metabolome Central web service and may perform test ID requests and test submissions.

6.7. In parallel, the external documentation of the MetX consortium must be updated to include the new resource. Once it is ready, the documentation will be made available at the MetX web page (http://www.metabolomexchange.org).

6.8. When the MetX Database Providers and applicant both agree in that the external documentation is ready and the software communication between Metabolome Central and the applicant are satisfactory, the new partner will be given full production status and production submissions may commence.

7. Revision of the document

7.1. This document may be revised at any time such that the approval of changes follows the voting guidelines described above.

4 Delivery and schedule

The delivery is delayed: ☐ Yes ☑ No

5 Adjustments made

N/A

6 Efforts for this deliverable

Institute	Person-months (PM)		Period
	actual	estimated	24
1:EMBL-EBI	0.5		
7:UB	0		
13:UBHAM	0		
Total	0.5	1	

Appendices

1. N/A

Background information

This deliverable relates to WP7; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP7 Title: Outreach and Training
Lead: Ulrich Günther, UBHam
Participants: Ulrich Günther

This work package will provide a close link between the COSMOS consortium and the wider metabolomics and the biomedical community. We will raise community awareness for the services provided by the COSMOS consortium, from data submission support to different views on

metabolomics data, but also for the increasingly stringent requirements for data availability as part of the metabolomics publication process. Through the existing framework of the Metabolomics Society, we will ensure broad community input into the services developed by the COSMOS consortium and the standards for Metabolomics data representation developed in this proposal. This deliverable relates to WP7; background information on this WP as originally indicated in the description of work (DoW) is included below.

Work package number	WP7	**Start date or starting event:**			Month 1			
Work package title	Outreach and Training							
Activity Type	COORD							
Participant number	1: EMBL-EBI	2: LU/NC	7: UB	13UBHam				
Person-months per participant	8	2	4	8				

Objectives

1. Provide link between the COSMOS consortium and the wider metabolomics and the biomedical community
2. Raise community awareness for the services provided by the COSMOS consortium
3. Collect broad community input into services developed by the COSMOS consortium

Description of work and role of participants

We will initially employ the usual channels for the disseminations of COSMOS standards, including scientific publications, and workshops and presentations at metabolomics conferences to reach the wider metabolomics community.

The project will plan activities adequately resourced devoted to dissemination for specialised constituencies and general public, in particular for awareness and educational purposes. The dissemination plan deliverable will consider adequate messages about the objectives of the project and its societal and economic impact. The tools we will use will include web-based communication, press releases, brochures, booklets, multimedia material,etc. The 'dissemination material' will be regularly updated to provide the latest version of the project status and objectives. Electronic and/or paper versions of this 'dissemination material' will be made available to the Project Officer beforehand for consultation and upon its final release. In all material produced in all dissemination activities we will properly acknowledge the source of funding by prominently placing the FP7 logo and the European Commission logo.

We will create and maintain the consortium website in an open source content management system. The website will have an EU domain such as www.cosmos-fp7.eu. This website will have a specific COSMOS branding with a professionally designed COSMOS logo to reflect the collaborative and international nature of the consortium. The COSMOS website will allow for content

management by the partners, additional component (intranet, calendar, web site search), advanced analytics, functional testing, and communication via mailing lists. COSMOS will also build close links between the COSMOS consortium and the European and International metabolomics community, and the wider biomedical community. For this COSMOS will build on existing links with other EU and International initiatives (e.g. EU projects including the ESFRI infrastructures ELIXIR, BioMedBridges, EU Openscreen, and the IRSES Word Wide NMR to build links international stakeholders in China, specifically the Wuhan metabolomics center, India, and South America). Moreover, COSMOS will build links metabolomics groups in Canada (HMDB, Wishart) and the US (BMRB; see letters of support), both running major metabolomics WEB portals. COSMOS will also build an intensive dialog between mass spectrometry and NMR instrument vendors, search engine providers, experimentalists, data resources, and journal offices. This will require travel funds for all COSMOS stakeholders. Stakeholders are key members and opinion makers of the wider metabolomics community worldwide, in particular the USA, Canada and Japan. These will be invited to workshops and the annual stakeholder meetings where COSMOS standards are disseminated. COSMOS will react dynamically to requests from participants and outside advisors. The stakeholder meetings will preferably be held attached to the annual Metabolomics society meeting or the MetaboMeeting, where the majority of the metabolomics community is present.

Two ELearning WEB tutorials on “Metabolomics Data Deposition” through COSMOS will be made available through partner websites, links from conferences, and announcements on the project web site. The tutorials will be step-by-step 20-30 minutes guides to “Metabolomics Data Deposition and Dissemination through COSMOS”.These tutorials need to be carefully scripted, rehearsed and produced in the Media Production room of the Wellcome Trust Genome Campus.We expect that COSMOS standards will be adapted quickly considering this wide-spread array of dissemination channels involved.

Task 1: Development of outreach material (Website, flyer, brochures, etc)

Task 2: Publication of results in scientific journals

Task 3: Presentation of work of the consortium at conferences, in particular the MetaboMeeting and the annual meeting of Metabolomics Society.

Task 4: Development and dissemination of a web tutorial about the workflows and standards developed in this consortium.

Task 5: Provide training workshops on Metabolomics data deposition, dissemination and access through the general EBI outreach department. (EMBL-EBI)

Deliverables

No.	Name	Due month
D7.1.1	Outreach activity plan (including publications)	2
D7.1.2	Updated Outreach activity plan (including publications)	12
D7.1.3	Updated Outreach activity plan (including publications)	24
D7.2	Report on the COSMOS consortium website	2
D7.3.1	Report on annual stakeholder meetings	12

D7.3.2	Report on annual stakeholder meetings	24
D7.4.1	Web-based tutorial	18
D7.4.2	Updated web-based tutorial	24
D7.5.1	Report about training workshops	24
D7.5.2	Report about training workshops	36