Deliverable D4.4

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	
	Research Infrastructures, FP7 Capacities Specific Programme; [INFRA-2011-2.3.2.] “Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Life sciences"	
Deliverable title:	Consultation of the MSI implementation of the COSMOS data flow Publishers and International Society	
WP No.	WP4	
Lead Beneficiary:	THE UNIVERSITY OF MANCHESTER	
WP Title	Data Deposition	
Contractual delivery date:	31 December 2013	
Actual delivery date:	31 December 2013	
WP leader:	Roy Goodacre	UNIMAN

Contributing partner(s):	Elon Correa, Dirk Walter, (WP3) Thomas Hankemeier (WP5), Steffen Neumann, Daniel Schober (WP2), Reza Salek (WP1), Roy Goodacre (WP4)

Authors: Elon Correa, Reza Salek, Roy Goodacre

Contents

1 Executive summary

This deliverable aims to describe actions taken towards effective COSMOS and community consultation of the proper implementation of Metabolomics Standards Initiative (MSI) (Fiehn et al. 2007) to the COSMOS data flow. To achieve this goal we are actively encouraging partners and the community to provide constructive feedback and evaluation of the best practices for effective implementation of MSI standards to the COSMOS data flow.

2 Project objectives

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

No.	Objective	Yes	No
1	Consultation of the MSI implementation of the COSMOS data flow Publishers and International Society	X	

3 Detailed report on the deliverable

3.1 Background

COSMOS is establishing clear procedures for metabolomics data annotation, submission and deposition, results reporting and publishing requirements. This will ensure proper reporting of MSI compliant metabolomics data, metadata, annotation and that required minimum information is captured according to existing and newly adapted MSI guidelines. These new guidelines are currently being carefully discussed, elaborated and agreed by COSMOS partners, data curators and publishers. The consortium is also taking every opportunity to

engage with stakeholders and potential collaborators on planning, discussion and implementation of the MSI guidelines for data flow. For examples; several of the COSMOS partners are Members and Directors of the Metabolomics Society, also on the Board of other "omics" standardization initiatives, ensuring links and cross talks, and working with publishers. For example, new data publication platforms such *Nature Publishing Group's Scientific Data* and BioMedCentral/BGI's GigaScience already use the ISA framework adapted as a means of capture metabolomics metadata in MetaboLights.

3.2 Description of Work

In September 2012, the National Institutes of Health (NIH) Common Funds Metabolomics program awarded funding related to metabolomics research advancement, funding three Regional Comprehensive Metabolomics Research Cores (RCMRC) and a Data Repository and Coordination Centre (DRCC) to act as a North American hub for metabolomics related research.

As the Editor of the *Metabolomics* journal states on his 2014 editorial (Goodacre 2014) "Metabolomics 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 a MSI compliant framework, COSMOS is encouraging researchers to deposit their data to one of the above repositories. While we recognise that progress may be gradual as we develop mechanisms and resources consistent with these principles, we are working in concert to achieve all goals set by COSMOS. 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.

3.3 Next steps

Proper implementation of MSI across the COSMOS data flow will be heavily influenced by decisions made in WP2 and implemented in WP3 and WP5. It will

also be supported by a communally developed set of mechanisms designed to detect and verify MSI compliance. We are liaising with all parties interested and striving to make sure that COSMOS data flow and all data submitted adhere to the community agreed standards set by MSI. In addition, as deliverable 4.3 "MSI implementation of the COSMOS data flow" has been extended and is not fully operational we will need to postpone real extend or delay consultation of the MSI implementation. Nonetheless, once open access to the system is in place all interested parties will be consulted and corrective measures to the system will be implemented as necessary. Table 1 lists relevant URLs related to the work described above.

Table 1: List of relevant URLs.

COSMO consortium	http://www.cosmos-fp7.eu/
MetaboLights	http://www.ebi.ac.uk/metabolights/
Metabolomics journal	http://link.springer.com/journal/11306
Metabolomics Standards Initiative (MSI)	http://msi-workgroups.sourceforge.net/
Nature's Scientific Data	http://www.nature.com/scientificdata/
BioMedCentral's GigaScience	http://www.gigasciencejournal.com/
ISA Tools and community	http://isa-tools.org and http://isacommons.org
National Institutes of Health	http://www.nih.gov/

4 Publications

N/A

5 Delivery and schedule

The delivery is delayed: ☐ Yes ☑ No

6 Adjustments made

This work might change and updated as the requirements change by time in accordance to data work flow schema and negotiation with data repositories.

7 Efforts for this deliverable

Institute	Person-months (PM)		Period
	actual	estimated	
9: UNIMAN	6	10	10
2:LU/NMC	1		
4:IMPERIAL	0.37		
12:UB2	1		
Total	8.37		

8 References

- Fiehn O., Robertson D., Griffin J., van der Werf M., Nikolau B., Morrison N., Sumner L.W., Goodacre R., Hardy N.W., Taylor C., Fostel J., Kristal B., Kaddurah-Daouk R., Mendes P., van Ommen B., Lindon J.C., Sansone S.A. (2007). The metabolomics standards initiative (MSI). Metabolomics 3, 175–178.
- Goodacre R. (2014). Water, water, every where, but rarely any drop to drink. Editorial, Metabolomics 10, 5-7.
- Goodacre R. (2010). An overflow of… what else but metabolism! Metabolomics, 6, 1-2.

Background information

This deliverable relates to WP4; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP4 Title: Data Deposition
Lead: Roy Goodacre, UNIVERSITY OF MANCHESTER
Participants: WP1, WP2, WP3 and WP5

First, we will implement harmonized and compatible data deposition and annotation strategies across all partners, providing data producers involved in Metabolomics experiments with a single point of submission. The data deposition and exchange workflow in the COSMOS consortium will be formally defined, agreed, and documented in relation with WP3 and all partnering databases in Europe and world-wide that will be invited to participate.
As a second objective, we will work towards the generation of an annotation manual for submitted data and strive to make sure that all metabolomics data submitted to partner databases are annotated to this standard.
Since the adoption of minimal standards for metabolomics by the relevant journals is a major goal of this coordination action, we are going to consult with publication houses and ensure data annotation quality and consistency, according to the required standard level set by each journal.
In this activity the work by the BioSharing initiative (http://biosharing.org) will also be explored. Building on the effort of Minimum Information for Biological and Biomedical Investigations' (MIBBI) portal (http://mibbi.org), the BioSharing initiative works to strengthen collaborations between researchers, funders, industry and journals, and to discourage redundant (if unintentional) competition between standards-generating groups.

Work package number	WP 4	**Start date or starting event:**							month 1		
Work package title	Data Deposition										
Activity Type	Coord										
Participant number	1: EMBL-EBI	2: LU/NMC	3: MRC	4:mperial	6: VTT	7: UB	8:MPG	9: UNIMAN	11:IPB	12: UB2	13:UBHAM
Person-months per participant	9	6	6	6	2	2	2	14	1	2	2

Objectives

1. Define COSMOS metadata format, as formally agreed by the members of the COSMOS consortium

Description of work and role of participants

Task 1: Definition and implementation of deposition data flow in the COSMOS consortium. The value of metabolomics data without proper biological, technical and statistical background is really quite limited. This was recognized by the Metabolomics Standards Initiative (MSI) and this resulted in a series of guidelines for minimum reporting standards that should be used for metabolomics experimentation (published in *Metabolomics* **3(3**) in 2007).

In a close collaboration of all COSMOS participants, and after consultation with stakeholders (viz. MSI, Metabolomics Society, relevant Publishers, National and international funders), we will define the COSMOS data deposition workflow. MSI guidelines will be followed and we shall co-ordinate the representation of results and metadata in a relational database/XML representation, with data stored as WP2-compliant formats. We will define the joint COSMOS data format and submission requirements, likely a thin metadata wrapper around MSI data formats. On successful submission, a standard format file will be generated, containing a COSMOS accession number, metadata, and a private data access option for the use of the data owner and reviewers. The file will be sent to the data depositor, for him/her to pass on to the journal for review purposes.

On publication of a manuscript, the associated dataset will be released by publisher and/or corresponding author, and an updated version of the metadata will be issued via the COSMOS RSS notification system, allowing all interested parties to access, process, and import the relevant data. This will have tremendous benefit to the metabolomics community, allowing others to re-create statistical approaches, providing data for others to mine and allowing the peer review process to access the raw and processed data of an experiment.

The precise format of this has not yet been implemented and as discussed above we shall engage all stakeholders as well as publication houses. This task involves contributions from all COSMOS participants to deposit data and test the validity of the developed workflows, reflecting the central role of the data deposition workflow for all partners involved.

Task 2: Implementation of a MSI journal validation system

As discussed in Task 1 the value of metabolomics data without proper biological, technical and statistical background is really quite limited. This task will develop tools to validate compliance of the submitted metabolomics data with the MSI guidelines or specific journal requirements. This is not meant to tell people how to perform their analyses but to allow adequate reporting of what was performed so that others can repeat the work. As a result of the validation process, after COSMOS data deposition, a report about guideline compliancy of each submission will be generated automatically. This would

aid Reviewers of articles submitted for publication as well as Editors handling paper submissions.

Springer will pilot this initial system as the publisher of *Metabolomics* (http://www.springer.com/life+sciences/biochemistry+%26+biophysics/journal/11306) with the backing of the International Metabolomics Society (http://www.metabolomicssociety.org/) as this is their official journal. Several of the COSMOS consortium participants are Members and Directors of the Metabolomics Society. In addition many other journals are interested in developments in this area including *Nature Biotechnology* (Nature PG), *Genome Biology* (BMC), *Molecular Systems Biology* (RSC) and *Molecular BioSystems* (Nature PG and EMBO).

Deliverables

No.	Name	Due month
D4.1	COSMOS repository data flow definition	9
D4.2	COSMOS metadata format definition	9
D4.3	MSI implementation of the COSMOS data flow	15
D4.4	Consultation of the MSI implementation of the COSMOS data flow Publishers and International Society	15
D4.5	Implementation of MSI/journal validation system	15