Deliverable D6.1

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 6 description	FP7-INFRASTRUCTURES-2012-1	
Deliverable title:	Document describing requirements for relevant biomedical infrastructures with regard to Metabolomics	
WP No.	6	
Lead Beneficiary:	10: CIRMMP	
WP Title	Coordination with BioMedBridges and biomedical ESFRI infrastructures	
Contractual delivery date:	01 10 2013	
Actual delivery date:		
WP leader:	Prof. Claudio Luchinat	CIRMMP
Contributing partner(s):	*Claudio Luchinat, Antonio Rosato, Leonardo Tenori, Paola Turano*	

Authors: Claudio Luchinat, Antonio Rosato, Leonardo Tenori, Paola Turano

Content

1. Executive summary

This deliverable aims at;

1. Maximizing communications with BioMedBridges and with its partner BMS infrastructures with an interest in metabolomics

2. Maximizing the usefulness of the COSMOS activity for the current large-scale EU biomedical infrastructures.

3. Improving participation in the concertation activities and meetings related with the e-Infrastructures area.

4. Optimizing synergies between e-Infrastructures by providing input and receiving feedback from working groups addressing activities of common interest.

2. Project objectives

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

No.	Objective	Yes	No
1	CIRMMP will coordinate the gathering of requirements regarding the use of metabolomics data as a molecular phenotyping technique with the above-mentioned e-infrastructures.	x	
2	Integration between e-infrastructures	x	

3. Detailed report on the deliverable

3.1 Background

The aim of WP6 is to foster the interaction between COSMOS and the biomedical infrastructures with a particular interest in metabolomics (BBMRI, Elixir, EU-Openscreen, EuroBioimaging and INSTRUCT) that are also participating in the BioMedBridges project. The idea is to obtain indications useful to focus and

prioritize the various activities in COSMOS in order to effectively respond to the needs of the current large scale EU biomedical infrastructures.

The present document provides an overview and analysis of the requirements of the aforementioned biomedical infrastructures with regard to metabolomics.

3.2 Description of Work

3.2.1 BioMedBridges

BioMedBridges (http://www.biomedbridges.eu/) is a joint effort of ten biomedical sciences research infrastructures on the ESFRI roadmap. Together, the project partners will develop the shared e-infrastructure—the technical bridges—to allow interoperability between data and services in the biological, medical, translational and clinical domains and thus strengthen biomedical resources in Europe.

Metabolomics is explicitly addressed in one of the use cases of BioMedBridges (Use case: PhenoBridge - crossing the species bridge between mouse and human, constituting Work Package 7). In this context, the usefulness of the technique is to provide data to assess the integration of ontologies from corresponding pathologies (diabetes, obesity) across different species (human and animal models). Achieving such anintegration and developing general methods to do this kind of activities is one of the principal aims of BioMedBridges.

The interaction with BioMedBridges occurred through the participation of COSMOS delegates at meetings/workshops organized by the BioMedBridges partnership. More in detail, the first such occasion was the Annual General Meeting of the project (Duesseldorf, 11-12 March 2013). Here, discussions were held concerning the whole of the BioMedBridges activities, in order to identify specific topics, and the corresponding project Workpackages, where there is overlap of interests with COSMOS. Key individuals to refer to were also identified, such as Dr. Nathalie Conte from EBI and Dr. Michael Raess from Helmholtz Zentrum München (who is also project manager of Infrafrontier). On this occasion, it was agreed that a more focused discussion and exchange of views could take place on the occasion of the upcoming PhenoBridge Ontology Mapping

Workshop, which would be held on April 15-16, 2013 in Hinxton, UK. This required a modest delay in the preparation of the present deliverable. The workshop allowed us not only to discuss about metabolomics data and their usefulness in the context of BioMedBridges but also to compare our approaches to ontology development. In turn, within the latter topic we could highlight commonalities/differences in the scopes.

The conclusions from the above discussions can be summarized as follows: the description of raw experimental data (and ultimately their storage) is not addressed by BioMedBridges, and actually constitutes one of the objectives of COSMOS; in this respect, the two projects are thus inherently complementary. Consequently, the requirements of BioMedBridges regarding metabolomics concern data that are at least partly interpreted. In particular, the identification of individual metabolites and how their levels in body fluids are differently affected by pathologies, life styles, aging, etc., also across species, would be of high added value to BioMedBridges. The identification of metabolites does not need to define exactly the chemical species involved, but rather should define metabolites that are distinctly associated to the pathologies of interest (diabetes, obesity for the time being). It was also agreed that data exchange between COSMOS and BioMedBridges could be practically, efficiently implemented via the MetaboLights database. The latter allows the deposition of both raw data and some interpreted data (e.g. metabolite levels), and thus it serves as a reference point for the communities of both projects that can be accessed in a fairly straightforward manner by all the involved researchers. Finally, it was noted that the development of an ontology or controlled vocabulary for metabolites is explicitly targeted by neither COSMOS nor BioMedBridges. The latter project indeed at present focuses mostly on disease terms. Nevertheless, COSMOS is tightly linked with the Human Metabolome Database (http://www.hmdb.ca/). Furthermore, the COSMOS partnership involves the developers of the Birmingham Metabolite Library (http://www.bml-nmr.org/), a collection of 3328 experimental 1D and 2D J-resolved NMR spectra of 208 metabolite standards. Taken together, these initiatives, in combination with the development within COSMOS of Data exchange format for metabolite identification and quantitation (WP2), address possible issues with metabolite terminology in an effective manner. It was reckoned that these

initiatives might result useful in supporting future BioMedBridges efforts involving metabolomics information.

As a result of discussions, there was a general consensus that, beyond the use cases defined for the ongoing projects, it would be important to design general strategies and procedures to make it possible to thoroughly compare metabolomic (derived) data obtained within relatively different experimental settings (as in the aforementioned case of metabolomics applied to different organisms). Indeed, as the goal of BioMedBridges is to build "bridges" across different biomedical disciplines, a further implicit requirement is that any methodology deployed can be applied in the various biomedical contexts with little or no modification.

3.2.2 BBMRI

BBMRI (Biobanking and Biomolecular Resources Research Infrastructure; http://www.bbmri.eu/) is based on a 54-member consortium with more than 225 associated organizations from over 30 countries. The essential aim of BBMRI is to create an interface between biological specimens and data (from patients and European populations) and top-level biological and medical research.

Human biobanks are structured resources that store: (a) human biological materials and/or information generated from the analysis of the same and (b) extensive associated information. Examples of different models of human biobanks are population biobanks, epidemiological collections, collections of carriers of specific genetic mutations/markers/ profiles; and collections of samples and data from individuals with a certain disease or taking specific medications. The resources of a human biobank may be used for a variety of research purposes to advance our understanding of human health and the life sciences, including in emerging "omics" fields such as proteomics, transcriptomics and metabolomics.

Metabolomics is the newest of the -omics sciences and there is a growing interest for its applications in prognosis, diagnosis, personalized medicine. Metabolomics studies rely on the availability of large sets of samples. The biological variation in individual metabolism and the dependence of metabolism on environmental factors necessitates large sample numbers to achieve the appropriate statistical power required for meaningful biological interpretation.

Biobanks are the ideal providers of large collections of high quality samples for metabolomics studies and of the associated data essential for the statistical analysis of such data. Analysis of metabolomic studies will greatly benefit from the issues of semantic interoperability through standardized message formats and controlled terminologies for gathering and storing genotype and phenotype information associated to stored biospecimen. Within BBMRI, there are ongoing activities on these topics within Work Package 5 (Database harmonization and IT-infrastructure) although not focused on metabolomics. Strong interactions between COSMOS and BBMRI have been established to coordinate efforts. For this reason, Kurt Zatloukal, coordinator of BBMRI has been nominated in the Advisory Board of COSMOS.

To facilitate and extend the use of metabolomics studies, important inputs can be provided to biobanks on the best way to create sample and data collections. This would contribute specifically to activities in BBMRI WP3 (Disease oriented biobanks) and to the scope of maximizing the opportunities for the discovery of new biomarkers and their validation and translation towards clinical applications

Besides the obvious link between metabolomics and biobanks as source of samples and associated data, a new aspect is emerging: metabolomics may be used as an efficient tool to monitor pre-analytical sample variations as metabolites proved be the most sensitive biomarkers of degradation phenomena among the various biomolecules (DNA, RNA, proteins, metabolites). Intactness of the metabolome is a good warrant of the intactness of the overall biomolecular profile. The first systematic use of metabolomics in this sense has been implemented with the SPIDIA project (Standardisation and improvement of generic pre-analytical tools and procedures for in-vitro diagnostics; http://www.spidia.eu), ended in March 2013, that has seen the participation of a partner of COSMOS. Biobanks cure procurement, collection, labeling, registration, processing, storage, tracking, retrieval, transfer, and use of human biological material. An obvious extension of the idea behind SPIDIA is the use of metabolomics to check the critical aspects of the overall workflow implemented at biobanks and also the performances of adopted procedures for sample storage.

Metabolomics and related issues may come in play at various levels. The following points have emerged from extensive discussions among COSMOS partners (in particular CIRMMP and EMBL-EBI) and two Members of the Advisory Board: David Wishart (University of Alberta) and Kurt Zatloukal (Medical University of Graz and BBMRI Coordinator).

Collection and handling: Metabolomics may be used to evaluate the procedures adopted at a given biobank in terms of their ability to maintain the original metabolome of the collected biospecimen. Comparison between different procedures/biobanks could provide hints to implement more efficient standard operating procedures. Intactness of biomolecular profiles in biofluids is often sought through the use of collection tubes containing stabilizing solution. Metabolomics may be used for a comparative evaluation of the various stabilizers and to provide hints for the new stabilizers.

Traceability: The overall quality of samples stored in biobanks is not only associated with their collection and storage conditions, but also to their traceability and the precision and reliability of the data associated with those samples. Emphasis should be given to the fact that successful downstream molecular analysis not only depends on the quality of the clinical data but also on the availability of the information about sample history: collection and handling times and temperatures should be carefully annotated and become part of the associated dataset.

Storage: Metabolomics provides a direct approach to monitor the performances of the different storage conditions on the molecular profiles of different types of samples. [1-7].

Nevertheless systematic studies on large sets of samples are still lacking and in particular no precise data exist on the “shelf-life time” of samples stored in biobanks. Metabolomics may become an efficient means to fill the void due to the present limited availability of data.

Inclusion of metabolomic profiles in the biobank databases would be useful to assess sample quality and history and COSMOS may play a key role in the development of standardized formats.

In terms of analytical platforms, NMR-based metabolomics may be preferred to mass spectrometry for this type of evaluation studies because it requires only

minimal if no sample handling, is highly reproducible and fast, although able to detect only the most concentrated metabolites, i.e. only a few tens of molecules. Studies evaluating the effect of pre-analytical procedures on the metabolome of some sample biofluids (urine, serum, plasma) have shown that the main changes in metabolites composition are related to a combination of residual enzymatic activities and oxidation processes due to exposure to atmospheric oxygen. In tissues apoptotic processes and altered uptake of some metabolites also play a role. Concentration of single metabolites or small subsets of metabolites cannot be used as reliable marker of the sample history because of the complexity of phenomena and large interpersonal variability in the initial concentration of the various species. However, when considering the entire NMR spectral profiles they change according to a similar trend in all samples undergoing the same type of treatment. Predictive models based on the NMR profile can be built on statistically relevant sets of samples exposed to a specific treatment. These observations are the result of collaboration among COSMOS partners and Biobank members within the Spidia project already mentioned above. We therefore propose the use of NMR profiles as further data to be associated to biological samples to assess their quality and to evaluate the impact of pre-analytical treatments.

From the above consideration and also within the frame of BBMRI WP4, which specifically aims at integrating existing biomolecular resources, technologies, standards and know-how, the idea of the creation of an Expert Center on metabolomics is being developed.

3.2.3 ELIXIR

The purpose of ELIXIR is to construct and operate a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society. In addition, ELIXIR coordinates the BioMedBridges project, which is a key initiative to bring together the ESFRI Research Infrastructures in the Bio Medical Sciences field. Together, the project will develop the shared e-infrastructure - the technical bridges - to allow interoperability between data and services in the biological, medical, translational and clinical domains and thus strengthen biomedical resources in Europe. In order to achieve its mission, ELIXIR will construct, operate and enhance a distributed research infrastructure in

accordance with the requirements of the scientific community and under the direction of the ELIXIR Board. The ELIXIR Hub will be connected to ELIXIR Nodes to provide infrastructure for data, compute tools and standards and training as well as support for the ESFRI biological and medical science infrastructures. The link from COSMOS to ELXIR is via MetaboLights; an open access repository housing metabolomics based experiments.

The metabolomics data from COSMOS partners at CERM, Max Planck Institute Golm Metabolome Database (GMD), Netherlands metabolomics centre, Birmingham Metabolite Library and potentially other will be shared and submitted to the MetaboLighs repository under standards proposed and developed within COSMOS initiative. The “metabolic datasets subset” hosted by CERM representing INSTRUCT and the “Universität Graz database” hosted by Medizinische Universität Graz representing BBMRI will be transferred to MetaboLights after which data linking will take place. MetaboLights will employ similar policy for the data sets stored as adapted by other services within EBI and EXLXIR.

3.2.4 Euro-BioImaging

Euro-BioImaging is a pan-European infrastructure project whose mission is to build a distributed imaging infrastructure across Europe that will provide open access to innovative biological and medical imaging technologies for European researchers.

In recent years, thanks to hyperpolarization techniques, real-time metabolic imaging with MRI has become a reality. Metabolomic imaging can map biomolecular profile values onto anatomical structures. Traditional metabolomics on biopsies, with higher sensitivity, allows a more detailed fingerprinting. Integration of MRI and NMR spectroscopy could improve the translation of basic science knowledge and information to the clinical practice. These two worlds are going to communicate more and more in the future. Consequently, there is a need for adopting common standards. Within this frame, a meeting has been held on January 26, 2013 between researchers working at CIRMMP and Prof. Silvio Aime (Chair of Euro-BioImaging WP 8: Molecular Imaging) and his co-workers. A potential role for metabolomics in the development of new imaging probes has

been identified, which involves two main aspects: i) synergy of metabolomics and bioimaging techniques for target identification through the use of metabolomics for biomarker identification, and ii) evaluation of the potential toxicity of imaging probes via metabolic profiling of treated cells/tissues/patients' body fluids. Aspects that may be relevant to facilitate contacts and collaborations between the two platforms/research environments are the following.

The first point is to develop a software infrastructure able to describe, process, store, convert and share the spectral data in the most common formats used in the MRI and in the NMR community. This in order to make faster and simpler the analysis of metabolomics data acquired with either techniques and to enhance the interoperability of both platforms.

The second point is to adapt the already developed statistical tools commonly employed for metabolomics analysis also to MRI data or vice versa.

The last point is to develop (or to agree on) common ontologies to refer to the metabolites identified through the different approaches that are adopted in the projects. This point is relevant also to BioMedBridges, as outlined in the corresponding section.

3.2.5 EU-OPENSCREEN

EU-OPENSCREEN is the European Infrastructure of Open Screening Platforms for Chemical Biology (http://www.eu-openscreen.de/) and integrates high-throughput screening platforms, chemical libraries, chemical resources for hit discovery and optimization, bio- and cheminformatics support, and a database containing screening results, assay protocols, and chemical information.

COSMOS activities are relevant to this Infrastructure in that they contribute to extending the nature of the information contained in screening databases. Metabolomic profiles of urine and/or serum and/or plasma of patients that have been treated with drugs in clinical trials may contribute to the prediction of liver and kidney toxicity of drug candidates. They can also provide hints about the biochemical mechanisms of toxicity. The approach can also be used to predict individual ability to metabolize a given drug and individual susceptibility to the side effects of that drug. The feasibility of this approach has been demonstrated on animals by the Consortium for Metabonomic Toxicology (COMET) [8]. EU-

OPENSCREEN would thus highly benefit from the development of consistent procedures to collect, store, interpret and report the interpretation results of metabolomics data, especially in the form of profiles or individual metabolite levels across multiple individuals and/or as a function of time and treatment.

3.2.6 INSTRUCT

INSTRUCT aims at creating a scientific environment that offers real opportunities to understand in atomic detail how the three-dimensional structure of a protein or pathogen interacts dynamically with its function, and with its environment within the cell. The relationships among these goals and the objectives of COSMOS have been discussed by the partners commonly involved in the two projects (CIRMMP, University of Oxford, and Leiden University which serves as a reference point to the Dutch scientific community, involving also the INSTRUCT center at the University of Utrecht), and the corresponding National INSTRUCT communities of Italy, United Kingdom and the Netherlands. Altogether, this network of organizations involves as many as nearly 400 research bodies (both public and private).

It has been pointed out that the interactomic view of INSTRUCT can be well complemented by the metabolomics data central to COSMOS as the latter tool to characterize and quantify all the small molecules in a biological sample. The two views will help connecting molecular events at the protein level to those occurring at the macrosystem level. INSTRUCT requirements thus mainly concern methods to characterize cellular metabolome variations in a time-dependent manner and as a function of stimuli.

References

1) Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, Lucio M, Zell A, Häring HU, Xu G, Lehmann R. Clin Chem. 2013 [Epub ahead of print]

2) Rist M, Muhle-Goll C, Görling B, Bub B, Heissler S, Watzl B, Luy B, Metabolites 2013; 3:243-258

3) Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. J Biomol NMR. 2011; 49(3-4):231-43

4) Rosenling T, Stoop MP, Smolinska A, Muilwijk B, Coulier L, Shi S, Dane A, Christin C, Suits F, Horvatovich PL, Wijmenga SS, Buydens LM, Vreeken R, Hankemeier T, van Gool AJ, Luider TM, Bischoff R. Clin Chem. 2011; 57(12):1703-11

5) Rasmussen L.G., Savorani F, Larsen TM, Dragsted LO, Astrup A, Engelsen SB. Metabolomics 2011; 7(1):71-83

6) Lauridsen M, Hansen SH, Jaroszewski JW, Cornett C. Anal Chem. 2007 ;79(3):1181-6.

7) Maher, A.D.; Zirah, S.F.M.; Holmes, E.; Nicholson, J.K. Anal. Chem. 2007; 79:5204–5211.

8) Lindon JC, Keun HC, Ebbels TMD, Pearce JMT, Holmes E, Nicholson JK. Pharmacogenomics 2005; 6(7):691-699

3.3 Next steps

An update of this document will be carried out if deemed necessary/appropriate by the WP partners, in response to significant events relevant to the contents.

4 Publications

N/A

5 Delivery and schedule

The delivery is delayed: ? Yes ☑ No

However the final extensive discussions with selected BioMedBridges partners took place in mid April.

6 Adjustments made

No

7 Efforts for this deliverable

Institute	Person-months (PM)	Period	
	actual	estimated	
CIRMMP	4	6	4
Total	4		

Appendices

None.

Background information

This deliverable relates to WP6; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP6 Title: Coordination with BioMedBridges and biomedical ESFRI infrastructures
Lead: Claudio Luchinat
Participants: EBI-EMBL, LU-NMLC, CIRMMP, UOXF

This work package aims at maximizing communications with BioMedBridges and with its partner BMS infrastructures with an interest in metabolomics (Elixir, EU-Openscreen, BBMRI and Instruct), and helping steer the work of the other work packages to maximize the usefulness of the COSMOS activity for the current large scale EU biomedical infrastructures. The COSMOS consortium will actively participate in the concertation activities and meetings related with the e-Infrastructures area. We will help to optimise synergies between projects by providing input and receiving feedback from working groups addressing activities of common interest (e.g. from clusters and projects). If requested we will offer advice and guidance and be receptive for any information relating to 7th Framework programme implementation, standardisation, policy and regulatory, EU Member States initiatives or relevant international initiative.
Description of work and role

Work package number	WP6	**Start date or starting event:**	month 1
Work package title	Coordination with BioMedBridges and biomedical ESFRI infrastructures		
Activity Type	COORD		

Participant number	1: EMBL/EBI	2: LU/NMC	10:CIRMMP	14: UOXF				
Person-months per participant	6.00	2.00	12.00	2.00				

Objectives

1. maximizing communications with BMS infrastructures with an interest in metabolomics
2. helping steer the work of the other work packages to maximize the usefulness of the COSMOS activity for the current large scale EU biomedical infrastructures.

Description of work and role of participants

Task 1: Gather metabolomics requirements for BioMedBridges, BBMRI, ELIXIR and EU-OPENSCREEN CIRMMP will coordinate the gathering of requirements regarding the use of metabolomics data as a molecular phenotyping technique with the above-mentioned e-infrastructures. The University of Florence as a third party of CIRMMP will contribute to this task. EBI as coordinator of ELIXIR, BioMedBridges and responsible for database and standards development in EU – OPENSCREEN will contribute use cases from its on-going integration efforts. UOX is leading the development of the ISA infrastructure, which assists in the annotation, and local management of experimental metadata from high-throughput studies employing one or a combination of omics and other technologies, and will work toward integrating the findings from Task 1 in the ISA development.

Task 2: Coordinate with the activities of BioMedBridges regarding the standardization of metabolomics data WP 7 of the BioMedBridges grant will work in particular on NMR metabolomics data and towards a standardized description of sample donors, sample collection; pre-processing, analysis and evaluation will be established as a prerequisite for the inter-species comparison of metabolomics results. In this task all contributors to this task will ensure the appropriate coordination of the developments in WP2 of COSMOS and WP7 of BioMedBridges.

Task 3: Coordinate the activities of COSMOS versus the needs of Biobanks with respect to the association of NMR profiles to stored samples. The primary objective of biobanks is not merely archiving, but also distributing conserved and documented biological samples for research, and so they represent an irreplaceable support for all those studies in which the impact of the results is linked to the large number of the collected samples. The quality of stored biological samples is crucial for the outcome of subsequent studies. The molecules constituting the metabolic fingerprint are generally very sensitive to handling procedures and storage conditions, so metabolomics is a useful tool for

checking and assessing the quality of stored samples. The NMR profile of a sample allows its evaluation in entrance (to decide its acceptance) and in exit (to decide if it is still good to be distributed), so it is important to associate each stored sample to the respective NMR metabolic profile. The aim of this task is to coordinate the activities of COSMOS, taking into account the requirements of Biobanks with respect to the association of NMR metabolic profiles to stored samples. BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) was one of the first European Research Infrastructure projects funded by the European Commission (EC). The EC-funded preparatory phase of BBMRI came to its end in January 2011. During the past 3 years BBMRI has grown into a 53-member consortium with over 280 associated organisations (largely biobanks) from over 30 countries, making it the largest research infrastructure project in Europe (http://www.bbmri.eu/). In this task we will interface with BBMRI and develop a strategy for the use of Metabolomics for Biobank sample monitoring and deposition of the sample status data in COSMOS partner databases.

Deliverables

No.	Name	Due month
D6.1	Document describing requirements for relevant biomedical infrastructures with regard to Metabolomics	6
D6.2	Establishment of an NMR metabolomics working group between COSMOS and BioMedBridges	12
D6.3	Joint consensus document between COSMOS and BioMedBridges	18
D6.4	Joint consensus document between COSMOS and BioMedBridges (Updated)	36
D6.5	Report on the recommendations of the use of Metabolomics of Biobank sample monitoring	24