



Project no. : **200767**
Project acronym: **APO-SYS**
Project title: **Apoptosis systems biology applied to cancer and AIDS.**

Instrument: Collaborative project

Thematic Priority: Large-scale integrating project

Deliverable reference number and title

DL7: Data management workshop on consortium requirements for sharing of high-throughput data, definition of standards to be used for all types of data (WP5.1)

Due date of deliverable: **31/07/2008**
Actual submission date: **31/08/2008**

Start date of project:
1 of February 2008

Duration:
4 years

Organization name of lead contractor for this deliverable:
European Bioinformatics Institute (Beneficiary no 13)

Description of DL7 with completion date "Month 6"
Editor: Henning Hermjakob/Phil Jones

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Overall description of DL7

Editor: Henning Hermjakob/Phil Jones

The APO-SYS consortium will generate large sets of experimental data on cell death regulators, and will characterize in detail the roles of these genes and proteins in a variety of distinct systems and experimental conditions. These data will be generated in different forms like sequence mutations, gene inactivation, expression studies, phosphoproteomics, protein localization, interactomics, etc. Given the high quality of the consortium these results will be accurate, highly reliable and up to date. This wealth of knowledge must be integrated into a recognized and well-structured database environment. Moreover, the data must be disseminated within the consortium and to the scientific community.

Data extraction by from the literature by scientific curators is a highly inefficient method of data collection, and increasingly the scientific community develops guidelines and standards for the representation and dissemination of experimental results and conclusions. To efficiently disseminate APO-SYS project results, it is essential to take data management and dissemination requirements into consideration already at the experiment planning stage. Thus, data management and requirements for systematic data dissemination are an early focus of the consortium.

Description of work

This deliverable (DL7) is closely related to DL 6 “Data management workshop and workshop report on consortium requirements for sharing of high quality annotation of molecular entities”, because the distinction between annotation of individual biomolecular entities, for example genes and proteins (DL6), and the high throughput data that determine properties of the entities as well as their interactions, is not a sharp distinction. In collaboration between EBI and Mediceal Oy, we focussed on two major activities for DL6 and DL7:

1. Requirements document: EBI and Mediceal Oy jointly developed a guidance document “Recommended Data Exchange Formats, Ontologies and minimum Reporting Guidelines” for data types and technologies relevant to the APO-SYS consortium. This document provides an overview of current community practice and is provided as Appendix 1. It has been disseminated to all partners and it is available on apo-sys webpage.
2. Data management workshop: A consortium-wide data management workshop has been jointly organised by EBI and Mediceal Oy, and took place at the EBI June 24-27, 2008. To minimise travel for all partners, we decided to hold one longer data management workshop focussing on several aspects, rather than several separate events. All workshop material relevant to EBI resources is publicly available from the workshop website at <http://www.apo-sys.eu/aposys/WorkshopTraining/Workshop.htm> . All workshop material generated by Mediceal is, for commercial confidentiality reasons, available on request.

Deviations from planned work

None.

Organization of the workshop

The workshop was co-organized by Henning Hermjakob / Philip Jones of the EBI (Partner 13) and Christophe Roos / Timo Laitinen of MediceL Oy (Partner 23). The workshop was held at the European Bioinformatics Institute in Cambridge, UK between 24 and 27 June. The schedule of the workshop is available on the project web page at <http://www.apo-sys.eu/aposys/WorkshopTraining/EBI-workshop/Workshop.htm>

The first two days of this workshop focused on public data exchange formats and public data repositories covering the majority of the research interests and techniques being applied in the APO-SYS consortium. Particular emphasis was placed upon the available mechanisms for data submission to public data repositories to allow the sharing of high quality annotation with the wider scientific community with reference to the MediceL Integrator Software Platform. The utility of the resources provided by the EBI for data mining, analysis and comparison was also considered and described. The majority of the trainers focused their examples and tutorials upon apoptosis in the context of cancer and HIV infection.

The second two days focused on the MediceL Integrator Software Platform. Since a first training course is scheduled for the beginning of 2009, this section of the workshop was used to provide a very general introduction to functionalities that will be available and developed during the APO-SYS project.

Participants:

12 out of 23 participating organizations were represented by one or several researchers. The computer literacy of the participants varied significantly, from bioinformatics specialists to wet lab specialists (**Appendix 2**). The proper annotation and storage of measurement data will be critical to the successful systems biology approach to apoptosis. It is to be expected that the participants of the workshop will have an important role in their respective organizations managing measurement data and both EBI and MediceL will continue to provide support to the partners.

Feedback from the workshop:

The course delegates were given the opportunity to give feedback on the two sections of the data management workshop (i.e. two separate responses from each delegate), with consideration to various aspects: the overall organization of the course was considered excellent or good in 22 out of 23 responses, with 1 response describing it as satisfactory; the clarity of presentations was considered excellent or good by 20 out of 23 responses, with 3 responses describing this as satisfactory; the knowledge of the speakers was rated as excellent or good in all 23 responses; the content of practical sessions was considered excellent or good in 17 out of 23 responses, with the remaining 6 responses rating these sessions as satisfactory; 19 out of 23 responses rated the level of scientific content in the tutorial as 'about right' with 3 responses indicating this was 'a little specific' and 1 response indicating this was 'too general'.

The attendees made several comments in the feedback in which they expressed an appreciation of both the benefits of the complexity of the available resources and infrastructure, which

highlights the need for further training and support for the consortium partners throughout the duration of the project:

Selected general comments from attendees:

“I think in general the necessity of standardizing reports and provide controlled annotations are the most important message I had. Even if I am working in a field (phenotypic cell based assays and high content screenings) where this is still a chimera, it was a "wake up call" to start to think in a more (widespread) knowledge attitude”

“How foolish is not to take the time in using bioinformatics in my work since it will save me a lot of time and give me so much more information.”

“The (large) amount of tools that (the) EBI offers. I was surprised to see how many options there are that I don't know about.”

“...more [opportunities for] detailed training because a lot of things which looked important were not fully done for now”

“[the material was useful, though] I think this course should have concentrated more on the needs of the participants”

APPENDIX 1

APO-SYS: Recommended Data Exchange Formats, Ontologies and minimum Reporting Guidelines: http://www.apo-sys.eu/aposys/WorkshopTraining/APO-SYS-format-recommendations_2008-04b.pdf

APPENDIX 2

List of participants

NAME	Current Position	PARTNER	Main Technologies of participant	Data ^{aa}
Imreh, Gabriela	Project Manager	01 - KI (BZ)	siRNA, proteomics	yes
		01 - KI (RL)		yes
Galluzzi, Lorenzo*	Ph.D. Student	02 - INSERM	siRNA, transcriptomics	
Morozova, Nadya*	Researcher	02 - INSERM	siRNA, biostatistics	
▣		03 - IC		
▣		04 - INMI		
▣		05 - VIB		
▣		06 - Uulm		
▣		07 - UNIL		
Rohde, Mikkel	Senior Scientist	08 - DCS	siRNA, proteomics, array	yes
Ceschin, Danilo	Post-doctoral researcher	09 - CERBM - GIE	transcriptomics, CHIP-chip	
Azaria, Yaara	PhD student	10 - WIS	siRNA, proteomics	
▣		11 - VTT		yes
▣		12 - DKFZ		
Jones, Phil**	Senior Software Engineer	13 - EMBL-EBI	proteomics	
Warman, Naama	Phd student	14 - TAU	microarray	yes
Kamburov, Atanas	PhD Student	15 - MPG/Berlin	interaction data integration	yes
Barsacchi, Rico	Post doc	15 - MPG/Dresden	high content sceening	yes
Samusik, Nikolay	PhD Student	15 - MPG/Dresden	siRNA, biostatistics	
▣		16 - IMBA		
▣		17 - SDU		
Huber, Heinrich	Ajunct Senior Lecturer	18 - RCSI	in-silico	yes
▣		19 - KFU		
Lopez, Fabrice	Bioinformatics engineer	20 - UnivMed	micro-array, proteomics, java and C programming, web services	
▣		21 - UCBL		
Manzo, Fabio	PhD student	22 - SUNAP	microarray	
Dell'aversana, Carmela	PhD student	22 - SUNAP collaborator	microarray	
Roos, Christophe**	Principal analyst	23 - MediceI	software research infrastructure, data integration	-

(*) present at minor part of the workshop

(**) trainer

▣) no participant

▣▣) Answer to on-line data production questionnaire