Deliverable No. 3.3.2

Annual report on the Open Consultation Process and Architectural Refinements (update)

Grant Agreement No.:	270089
Deliverable No.:	D3.3.2
Deliverable Name:	Annual report on the Open Consultation Process and Architectural Refinements (update)
Contractual Submission Date:	31/01/2013
Actual Submission Date:	17/02/2014

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)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	***p-medicine***
Project Full Name:	From data sharing and integration via VPH models to personalized medicine
Deliverable No.:	D 3.3.2
Document name:	Annual report on the Open Consultation Process and Architectural Refinements (update)
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.0
Actual Submission Date:	17/02/2014
Editor: Institution: E-Mail:	Manolis Tsiknakis FORTH tsiknaki@ics.forth.gr

ABSTRACT:

This document reports on the open consultation efforts undertaken in the context of WP3 and presents the results and details of this task. This document reports for the third year of the project and focuses on a number of events and scientific conferences where the p-medicine platform was presented and discussed and opinions about its functionality and usage were exchanged with potential stakeholders and other interested groups.

KEYWORD LIST: standards, open consultation, architecture

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 270089.

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **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)

MODIFICATION CONTROL			
Version	Date	Status	Author
0.8	29/01/2014	Draft	Stelios Sfakianakis
0.9	13/02/2014	Draft	Stelios Sfakianakis
1.0	17/02/2014	Final	Stelios Sfakianakis

List of contributors

- Stelios Sfakianakis, FORTH
- Manolis Tsiknakis, FORTH
- Kostas Marias, FORTH
- George Zacharioudakis, FORTH
- Lefteris Koumakis, FORTH
- Vaggelis Sakkalis, FORTH
- Norbert Graf, USAAR
- Stefan Zasada, UCL (internal review)

Contents

Executive Summary

The consultation process in general involves seeking the opinions of interested and affected groups of the people. The objective is to gather information to facilitate the development of higher quality systems, methods, processes, or regulations. In task 3.3 of the project's work plan, the p-medicine platform is subjected to a similar process by seeking feedback from the user community, technology providers, IT engineers, and other stakeholders with respect to the current version of the architecture and the technology platform. To this end this document presents efforts for establishing communication channels with user communities, standardization organizations, and similar projects in a bidirectional and cross-pollinating way. This report further documents the details and results of these efforts.

Introduction

The p-medicine project aims to develop an innovative and integrated technological solution to facilitate personalized medicine. This endeavor is based on formulating an open, modular framework of tools and services supporting the efficient and secure sharing and handling of large personalized data sets. It is evident that such an ambitious effort should be based on common, accepted, and best-of-breed methodologies and standards. The use of open standards for interoperability, security, data formats, etc., in particular, is of the utmost importance.

But what is an *open standard*? According to the W3C the characteristic of open standards that sets them apart from the "closed" ones is the process by which the standard was defined[3], focusing on the transparency of the process, relevance to the user/market needs, openness in the participation, availability of the standard's text, etc. In agreement with these policies and the approach towards standardization, the project has dedicated tasks in its technical annex focusing on standards, both as a contributor and as a consumer. The activities involve monitoring of standards development, critical review and assessment of their applicability in the p-medicine framework, refining such standards based on domain-specific requirements. For this purpose annual dedicated scientific workshops are to be held (either internally to the project, or in the context of relevant international scientific events), with the participation of key international experts in respective domains. In parallel an open consultation process is foreseen, as an enabler for significant contributions to European and global standards.

[3] Daniel Dardailler (ed.), Definition of Open Standards, http://www.w3.org/2005/09/dd-osd.html World Wide Web Consortium, 2007

Report on the Annual Open Consultation process

The open consultation process is primarily for getting information about the use of standards that can be relevant for the p-medicine architecture. Therefore in this report we present some of the recent activities in the W3C that seem to be very pertinent to the p-medicine data architecture.

Additionally, during the previous year the p-medicine consortium members organized and participated in a number of sessions and workshops that we briefly describe in the following sections. The objectives of these events are threefold:

- To increase the familiarization of the community to the project's objectives and requirements and support its visibility and outreach.
- To gather feedback and experiences of the relevant stakeholders (users, IT developers, clinicians, bioinformaticians, etc)
- To influence and leverage the development and adoption of open standards in related domain areas.

1.1. Standards Update

In the previous year some new developments in semantic and data integration have emerged and been introduced by the World Wide Web Consortium (W3C). In the following paragraphs we present some of them that seem to be of great relevance in the p-medicine architecture and the development of its components

1.1.1.JSON-LD

A lot of p-medicine components offer an HTTP/ REST interface with the JSON data format[4] to serialize their messages. JSON is a simple format but so far there was not a generally accepted method (or schema) for expressing *links* between JSON documents. This linking facility is of utmost importance in the Semantic Web where the "Linked Data" principles[5] can be used for the sharing and the automatic discovering of new information. The "JSON-based serialization for Linked Data" (JSON-LD)[6] specification has recently been published as a W3C Recommendation and aims to fill this gap.

JSON-LD harmonizes the representation of Linked Data in JSON by describing a common JSON representation format for expressing directed graphs; mixing both Linked Data and non-Linked Data in a single document. The syntax is designed to not disturb already deployed systems running on JSON, but provide a smooth upgrade path from JSON to JSON-LD. It is primarily intended to be a way to use Linked Data in Web-based programming environments, to build interoperable Linked Data Web services, and to store Linked Data in JSON-based storage engines.

The JSON-LD 1.0 specification describes the JSON-LD language in a way that is useful to authors. It also provides the core grammar of the language for

[4] http://www.json.org/
[5] http://www.w3.org/DesignIssues/LinkedData.html
[6] http://www.w3.org/TR/json-ld/

implementers. The JSON-LD 1.0 Algorithms and API specification[7] describes useful Algorithms for working with JSON-LD data. It also specifies an Application Programming Interface that can be used to transform JSON-LD documents in order to make them easier to work with in programming environments like JavaScript, Python, and Ruby.

1.1.2. Link Data Platform 1.0

The W3C has published a Last Call Working Draft of the "Linked Data Platform 1.0"[8] This specification describes a set of best practices and simple approach for a read-write Linked Data architecture, based on HTTP access to web resources that describe their state using the RDF data model.

Some of the rules defined in this specification provide clarification and refinement of the base Linked Data rules while others address additional needs.

The rules for Linked Data Platform Resources address basic questions such as:

- What resource representations should be used?
- How is optimistic collision detection handled for updates?
- What should client expectations be for changes to linked-to resources, such as type changes?
- What can servers do to ease the burden of constraints for resource creation?
- How do I GET the entries of a large resources broken up into pages?

Additional informative guidance is available on the "Best Practices and Guidelines" document[9] that addresses deployment questions such as:

- What literal value types should be used?
- Are there some typical vocabularies that should be reused?

We consider these guidelines and recommendations to be of high interest to the p-medicine application developers and especially the Data Warehouse and other data repositories, which are denoted as "Linked Data Platform Container" in the context of the Linked Data Platform.

1.1.3."CSV on the Web" working group

Recently the "CSV on the Web" working group[10] has been formed inside the World Wide Web Consortium (W3C) to define a metadata vocabulary for CSV files, such as column headings, data types, and annotations, and, with it, making it easily possible to convert CSV into RDF (or other formats), easing data integration.

[7] http://www.w3.org/TR/json-ld-api/
[8] http://www.w3.org/TR/ldp/
[9] https://dvcs.w3.org/hg/ldpwg/raw-file/default/ldp-bp/ldp-bp.html
[10] http://www.w3.org/2013/csvw/

CSV is the primary data format for uploading data into the p-medicine Data Warehouse. Therefore the results of this working group will be very useful for the data upload and transformation services of the p-medicine platform.

1.2. Public Scientific Conferences

During the last year members of the consortium participated in a number of scientific conferences, summarized below:

- **35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13).** A special session entitled "Computational Solutions to Large-Scale Data Management and Analysis in Translational, Personalized and Predictive Medicine" was organized by Prof. Dr. Norber Graf (USAAR) during the Annual International Conference of EMBC. The workshop built on the experiences as well as technological and scientific developments stemming from some flagship projects funded by the EU under the FP7 framework programme. It aimed to bring together researches working in the fields of infrastructures and technologies for integrative biomedical research, ICT for predictive and translational medicine and the VPH at large, and in particular in all aspects of Cancer Informatics that support the vision of predictive and personalized medicine in the domain of cancer.
- The **13th IEEE International Conference on BioInformatics and BioEngineering (BIBE 2013)** took place in Chania, Greece. A special session called "The Digital Patient concept: Vision and Early Demonstrations" was co-organized by Manolis Tsiknakis (FORTH), Kostas Marias, and Norbert Graf (USAAR).

1.3. Conferences organized by the Project

The **2nd p-Medicine Summer School** took place in June 2013 in Schloss Dagstuhl in Wadern, Germany[11]. It was a great opportunity to present the project's achievements and also get feedback and information about what are the similar endeavors in the domain and their selected approaches. In particular the ObTiMA trial management system and its biobanking module were presented and demonstrated. Partners from ICCS presented the p-medicine Oncosimulator and its use as an integrative platform for in Silico Oncology in the CHIC EU project that was recently launched.

Participants from the VPH-Share project presented their approach for the use of public and private clouds for the development of the VPH scenarios and complex workflows. In particular the experience of the use of the Amazon Cloud was important for considerations inside p-medicine but the strong emphasis in the patient data confidentiality and privacy does not fit together with the use of public commercial clouds. Nevertheless the presented use of public and private clouds was quite impressive and it can be source for inspiration to future iterations of the p-medicine architectural design.

[11] http://www.dagstuhl.de/en/about-dagstuhl/

1.4. "Cross" Project workshops and events

1.4.1 ENCCA assembly event

ENCCA aims to establish a durable, European Virtual Institute clinical and translational research in childhood and adolescent cancers that will define and implement an integrated research strategy and will facilitate the necessary investigator-driven clinical trials to introduce the new generation of biologically targeted drugs into standard of care for children and adolescents with cancer. This will lead to more efficacious and less toxic therapies that will maximise the quality of life of the increasing number of survivors of cancer at a young age in Europe and allow them to assume their proper place in society. This biologically-driven research agenda will improve training of the clinical investigators and translational scientists of the future to spread excellence, increase capacity to participate in research and monitor outcomes across Europe. Patients and their families will be full partners and will be better informed about the need for and processes of clinical research. All of this will be achieved with respect for the highest ethical and patient safety standards. ENCCA will bring all stakeholders to the table in a timely and efficacious manner. It will address the needs of all the current multinational clinical trial groups for the benefit of children with cancer. It will provide them with common tools and approaches to solve the bottlenecks in testing new therapeutic strategies for those rare diseases in a vulnerable age group and in running a competitive clinical research agenda.

FORTH consulted with IT partners within ENCCA (AIT & CINECA) which are involved in biomedical research and related activities to get a comprehensive understanding on what are the needs and challenges of designing, developing and operating the envisaged electronic research infrastructure. Collateral activities have been put forward in order to support the interoperability and IT support to the ENCCA community.

1.1.4. Legal & security

The ENCCA project, amongst others, aims at implementing a secure Content and Knowledge Management Platform which will be based on the portal technology of ENCCA knowledge management system. It has to be stressed that a prospective collection clinical and genomics data needs a legal framework. This is especially true as these data are used by many different and sometimes multi-role end-users such as the ENCCA community. For that reason the legal framework of p-medicine has been presented at the 2nd and the 3rd General Assembly meetings of ENCCA from LUH (invited talks at GA of ENCCA).

1.1.5. Biobanking

In June 2013, an ENCCA Biobanking meeting held in Genoa. The meeting was held as part of the activities for Task 5.3 'Biobanking' within the ENCCA WP5 (Coordinator A. Eggert). The p-BioSPRE was proposed by FORTH as a solution for biobanks integration and the biobank work package of ENCCA expressed interest in the availability of a demo in order to demonstrate and verify the interest of biobanking integration and searching within the ENCCA community.

A demo, a thorough presentation and discussion for p-BioSPRE were led by FORTH with the support of FhG IBMT at the BioBanking pre-meeting of ENCCA on January 2014. p-BioSPRE was introduced as a fully integrated solution of two or more

biobanks. The Biology group of ENCCA verified the interest of such a solution and it was decided that p-BioSPRE would be nice to be setup and demonstrated for ENCCA, using two different biobanks (ALL and nephroblastoma).

The biobanking functionality of ObTiMA was also demonstrated at the biology pre-meeting of the 3rd GA ENCCA by USAAR.

Biobanks within ENCCA will also be used as sources to the ENCCA virtual institute. The proposed solution for the integration of biobanks with ABCD-4-E (the ENCCA virtual institute) requires a minimal dataset as a base to construct a CDA document that could then be registered in the ABCD-4-E.

1.1.6. ObTiMA

One of the objectives of ENCCA (WP11 task 11.3) is to implement prospective clinical, imaging and biological data collection on patients treated according to the standard arms of the SIOP Wilms tumour 2001. It was decided to use ObTiMA for the study and FORTH in collaboration with ENCCA WP11 has start a pilot of ObTiMA as a tool for data management on patients registered in the current SIOP WT 2001 trial and study. The study aims to define a database for the logistics of bio-banking of Wilms tumour material and clinical trial data throughout Europe to test the decentralised storage of Wilms tumour. The study is called IMPORT (Improving Population Outcomes for Renal Tumours of Childhood) and applies P-Medicine tools to improve treatment decisions for Wilms tumour and other renal tumours of childhood and young adults. Forms from Import Study (11 in total) have been created as eCRFs for use in ObTiMA (figure below).

Name	Version	Description
IMPORT - Bilateral Wilms Tumor (F12a)	1.0021	Bilateral Wilms Tumor (F12a)
IMPORT - Treatment of relapse/progression (F17)	1.0036	Treatment of relapse/progression (F17)
IMPORT - Bilateral Wilms Response Assessment (F12b)	1.0074	Bilateral Wilms Response Assessment (F12b) (1 form per assessment)
IMPORT - Bilateral Wilms Tumor (F12a)	1.0021	Bilateral Wilms Tumor (F12a)
IMPORT - Cardiotoxicity Monitoring Form (F11)	1.0014	All patients receiving doxorubicin need monitoring with ECHOs as part of standard practice.
IMPORT - End of treatment for bilateral cases (F20)	1.0033	This form should be completed after the patient's last course of chemotherapy
IMPORT - Follow-up/ Relapse/ Death (F18)	1.0058	Follow-up/ Relapse/ Death (F18)
IMPORT - Operative Findings (F3)	1.0052	PLEASE SEND TWO SEPARATE FORMS FOR BILATERAL CASES (indicate if Right or Left
IMPORT - Pathology Form (F4)	1.0036	Pathology Form (F4) - PLEASE CREATE TWO SEPARATE FORMS FOR BILATERAL CASES
IMPORT - Post Operative Treatment (F6)	1.0034	Post Operative Treatment (F6)
IMPORT - Pre-Operative Chemotherapy (F2)	1.0056	Unilateral Tumours ONLY! For Bilateral Cases Use Form 12b
IMPORT - Registration (F1)	1.0100	Registration (F1)

The system is already available and in use for the IMPORT study from UCL for retrospective and prospective clinical trial data. Integration of ObTiMA as a source to the ABCD-4-E architecture is currently under development from FORTH and it has been decided that CDISC–ODM compliant messages will deliver the core dataset information needed. The final core dataset which will be stored in ABCD-4-E for clinical trials is under discussion.

1.1.7. CHIC

The CHIC project[12] develops clinical trial driven tools, services and infrastructures that will support the creation of multiscale cancer hypermodels (integrative models). CHIC aspires to make a breakthrough in multiscale cancer modeling through greatly facilitating multi-modeller cancer hypermodelling and its clinical adaptation and validation. The CHIC project started on 2013 and there has been continuous interaction between the two projects in the last year and in particular in the following events:

- Kick-off meeting of CHIC on April 10-12, 2013 at Athens, Greece
- Progress meeting of CHIC on October 17-19 at Heraklion, Greece

Extensive collaboration and interaction between the two projects has been performed in all the previous mentioned areas, such as

- familiarization of the community with the objectives of the p-medicine project
- gather feedback from the relevant users and stakeholders
- influence the adoption of open technologies and standards in the related domain

Key technologies already being used by p-medicine have been identified as appropriate and candidate for adoption or even cross-collaboration between the two projects, to greatly enhance the interoperability and re-use of tools and services and to promote the usage of the same standards in the domain of personalized medicine.

Data warehouse

The data warehouse and the supporting technologies and protocols which have been implemented in p-medicine are candidate tools to be linked, where permitted under the existing legal framework and restrictions, as an external source of information in the CHIC platform. In the p-medicine project there has already been conducted elaborate work regarding the legal framework and the supporting security tools for hosting and sharing of sensitive medical data and this work can be further exploited in other projects, such as CHIC, to promote the usage of a homogeneous legal and security framework.

Cloud infrastructure

The OpenStack cloud platform, which is an open technology for building public or private cloud infrastructure and which is already being used in the context of p-medicine, is under consideration to be used also in CHIC. Usage of open tools such as OpenStack in p-mecicine has influenced the adoption of these open technologies also in CHIC.

Portal tools and interfaces

The Liferay platform, which is an open technology using open standards such as JSR168 and JSR286 Portlet specifications has been widely used in p-medicine. The

[12] http://chic-vph.eu/

same set of technologies, based on the feedback, evaluation and consultation of p-medicine has been chosen also in CHIC as an open technology for the CHIC portal implementation.