

**EGI-Engage**

Deliverable Review Form

|  |
| --- |
| **Details of the document being reviewed** |
| *Title:* | **M4.1 “Open Data Platform: Requirements and Implementation Plans”** | *Document identifier:* | **EGI-doc-2547-v1** |
| *Project:* | **EGI-Engage** | *Document url:* | **https://documents.egi.eu/document/2547** |
| *Author(s):* | **Y. Chen, L. Dutka, B. Kryza, T. Ferrari** | *Date:* | **6 August 2015** |

|  |
| --- |
| **Identification of the reviewer** |
| *Reviewer:* | **Dr. Sandro Fiore** | *Activity:* | **Data Scientist****Director of the Advanced Scientific Computing of the Euro-Mediterranean Center on Climate Change (CMCC)** |

**General comments on the content**

|  |
| --- |
| **Comments from Reviewer:** |
| The document is very interesting and aims at proposing both (i) the requirements of a very challenging data platform joining open data and eScience and (ii) its implementation plans.The main comments are:1) “Executive Summary”It is too short and it is more an abstract than an Executive Summary. *I would suggest extending this section to provide a more comprehensive view of the key outcomes of the report.*2) Section 2 “Introduction”I understand that some of the objectives of the Open Data Platform are reported in the first bullet of Section 2.3. However, the Open Data Platform concept/vision could be provided (e.g. through a definition) at the beginning of the document to make clear its main focus and boundary. Moreover, in Section 2.3 the authors state *“...dissemination and exploitation of open data in cloud environments...*”. Would the Open Data Platform be working in cloud environments only?*I would suggest highlighting the general goal of the Open Data Platform in a dedicated section at the beginning of the document to make clear (through a sort of formal statement/definition) what this platform is aiming for.* 2) “Section 4: Research Communities and Their Use Cases”While the general description of each research community fulfils the expectations, the description of the “use cases” part should be improved. In particular while for some research communities high-level and data-oriented scenarios are described (e.g. HBP), for other ones (e.g. EMSO) no information is provided. *I would suggest increasing the “data-related” aspects in the description of each research community to better link this section to the main goal of the deliverable*. 3) “Section 5:Requirements Analysis and Findings”The “Summary of the Communities Requirements” (Section 5.1) does not fulfil the expectations. Two tables in Section 5.1 (“Open access policies” and “Metadata characteristics”) miss key information (for the design of the open data platform) from several research communities. *I would suggest improving this section by providing (i) more insights and feedback from the research communities and (ii) a brief summary (textual description) at the end of the section to highlight the key points.*4) M4.1: “Open Data Platform: Requirements and Implementation Plans”. While the “Requirements” part is well described and presented in the document, the “Implementation Plans” one is less clear. *I would suggest improving the “Implementation Plans” part of the document*5) From the requirements to the technological solution. I miss an architectural view of the Open Data Platform, before going into the infrastructural aspects (e.g. the adopted technology). *I would suggest adding a high-level view of the “Open Data Platform” architecture, before describing the technological aspects/solutions.*6) On a printed version of the document some of the text embedded in the figures (e.g. Fig 1) is not readable. *I would suggest improving the quality of the figures. In some cases increasing the size of the figures could be enough to address the issue.* |
| **Response from Author:**  |
|  |

**Additional comments**

*(not affecting the document content e.g. recommendations for the future)*

|  |
| --- |
| **From reviewer:** |
| The part related to the feedback/requirements from the research communities should be strengthened in the document (all the technological aspects depend on that). Presenting in detail some use cases (even from a restricted/selected set of research communities) could be very effective to describe/summarize the most relevant open data needs, issues and challenges. Adding some details from the research communities questionnaires could be also useful (e.g. in the annexes). |

**Detailed comments on the content**

| **N°** | **Page** | **§** | **Observations** | **Reply from author(correction / reject,  …)** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**English and other corrections:**

Note: English and typo corrections can be made directly in the document as comments.