

**EGI-Engage**

Deliverable Review Form

|  |
| --- |
| **Details of the document being reviewed** |
| *Title:* | **D4.1 CANFAR Integration Roadmap** | *Document identifier:* | **EGI-doc-2549** |
| *Project:* | **EGI-Engage** | *Document url:* | <https://documents.egi.eu/document/2549>  |
| *Author(s):* | **Giuliano Taffoni** | *Date:14/08/2015* |  |

|  |
| --- |
| **Identification of the reviewer** |
| *Reviewer:* | **André Schaaff** | *Activity:* | **Project manager at CDS / Observatoire astronomique de Strasbourg**  |

**General comments on the content**

|  |
| --- |
| **Comments from Reviewer:** |
| It is a nice challenge to establish a transatlantic interoperability between CANFAR and EGI cloud resources. The involved parties, CADC and INAF are in the same domain (A&A) and have participated to the definition and implementation of the IVOA standards used in the project. This is a good factor of success. The federation way and not the sharing of a cloud is also a factor of sustainability. The scale change management in the astronomical large projects, especially in terms of data volume will benefit from this type of effort. Indeed, maintaining both the expertise and the local data storage in multiple data centers will undoubtedly become problematic in the coming years. These data centers should benefit from this cloud infrastructure while maintaining a high level of expertise. The success of this effort and its adoption by the data centers will ultimately depend on the ability to ensure the sustainability. It would be interesting to add a paragraph explaining how it will be possible.The idea of moving the computation and not the data is an old paradigm, which is now mature due to all the new technologies (like Docker) which make it possible easily. It is very important for the A&A and it is a factor of adoption by its community.The project mentions data volume around a few hundreds of terabytes. In a near future petabytes will become common. It would be nice to explain how it will be possible to deal with this evolution.I agree with the risk analysis. The project backbone is based on open technologies and standards and will not be affected by a tierce party defection.This project should also be a good example for other communities. |
| **Response from Author:**  |
| Dear Andre` thanks for your useful suggestions. I integrate the text with a short discussion on sustainability (Section 2.1)The data volume treated in this project is less than one PB, and tests and verification will be done with 100TB of data, however the federation approach is suitable to manage PBs of data. I expand the introduction to Section 2 in order to discuss this point. |

**Additional comments**

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

|  |
| --- |
| **From reviewer:** |
|  |

**Detailed comments on the content**

| **N°** | **Page** | **§** | **Observations** | **Reply from author(correction / reject,  …)** |
| --- | --- | --- | --- | --- |
| **1** | **11** | **3.2.1 & 3.2.2** | **Is the Data Transfer Service limited to VOSpace ?** | **The data transfer service is used to transfer data from/to the VOSpace client. I specify it in the document.** |
| **2** | **17** | **4.1.4** | **Is VOSI-availability not a "too light" way to do the monitoring ?** | **The VOSI-availability is used in production by CANFAR infrastructure to monitor their services. It will be analysed and compared with EGI monitoring to verify if it is suitable also for the federation.** |
| **3** | **20** | **4.3.1** | **In table 8, "... to support data replication", which is the replication general policy ? It is not detailed in the document.** | **The replication policy was not detailed in the document. I agree that it is important to discuss how data is replicated, so I introduce a point in the data policy section (5) to discuss also replication policy.** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**English and other corrections:**

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

P.7: is based "on" two use cases; sharing "from" A&A;

p. 21: grant access to "his" data

P.24: a typo in CoSADIE; ICE and CoSADIE are now finished