

**EGI-Engage**

Deliverable/Milestone review form

|  |
| --- |
| **Details of the document being reviewed** |
| *Title:* | **CANFAR federation roadmap** | *Document identifier:* | **D4.1** |
| *Project:* | **EGI-Engage** | *Document url:* [https://documents.egi.eu/document/2549](https://documents.egi.eu/document/2549%22%20%5Ct%20%22browserView) |  |
| *Author(s):* | **G. Taffoni** | *Date:* | **21/08/2015** |

|  |
| --- |
| **Identification of the reviewer** |
| *Reviewer:* | **R. F. Pizzo** | *Activity:* |  **Head of Science Support at the Radio Observatory of ASTRON** |

**General comments on the content**

|  |
| --- |
| **Comments from Reviewer:** |
| The document provides a comprehensive description of the implementation plan for CANFAR and EGI federated cloud in support of collaborative astronomy research. Obviously, this is a very important milestone for Astrophysical studies and research in general. Scientific collaborations are crucial to advance this field and make discoveries happen. Cloud resources will streamline these aspects, as there is currently an ever-growing demand for compute power and storage to properly handle the very large amounts of data delivered by current astronomical facilities.In general, I have no major comments about the content of the document. It seems that all aspects are covered in proper depth. A few minor comments are reported below:* I would try to be as quantitative as possible in the text. For example, at pg. 7, it says that this activity will allow users to process large data volumes – I would suggest citing the order of magnitude of the typical data that users will want to work on (TB?). This will give a very clear idea to future readers of why these federations are so important. A similar level of detail could be added at pg. 9 as well – ‘…and sharing of very large astronomical datasets…’.
* The document does not seem to provide numbers for typical required processing resources for current astronomical data – possibly this is not relevant at this stage, but it might be good to mention anyway.
* When discussing the Federation roadmap, it is stated that requirements will be gathered through the federation activities (pg. 13). Under the assumption that these are users requirements, is it clear already how these would be collected in practice? Both CANFAR and EGI would collect feedback from users through given fora? Maybe it would be good to briefly mention how this would be done, to show to the readers that this aspect is covered.
* As stated in the document, accounting and monitoring of the federated cloud will be crucial. I think it would be useful to mention at this stage whether you envision to have a mechanism in place that let the user actively know in case the resources he/she requested are underused.
* I would suggest adding the item ‘X.509 certificate’ to the glossary or explain its meaning in a note, as a general reader might not be familiar with it.
* The final comment is about testing at the end of the activity. Although a final test will be performed to make sure that things work properly, bugs might be experienced anyway at a later stage when more users start working with the system. How would the user report this? Is there a bug reporting system in place? Maybe it would be good to mention this post-final test phase as well in the document.
 |
| **Response from Author:**  |
| Dear Roberto, thanks for your useful comments. I update the document following your suggestions, in particular* I try to be more quantitative as regards both data and computing resources. The data volume managed by CANFAR is of the order of magnitude of PBs distributed in about 900 Million of files. I specify those values in the text.
* When presenting CANFAR infrastructure I also introduce some numbers regarding computing resources, in particular the amount of jobs processed per year and the amount of cores/year requested to process and analyse the data.
* The “Requirements gathered through the federation activities” are not user requirements but the federation requirements that may arise during the technical discussions or software implementation. I agree that this statement is misleading so I modify it in document. The A&A user requirements are not collected during this federation activities, however it is one of the key activities EGI-Engage project. User requirements are collected from different actors: EGI-Engage competence centers, research infrastructures, NGIs, resource providers, technology providers, user communities etc. The requirement gathering process is accomplished by the WPs which are in charge of the communication with users and key stakeholders. The WP5/SA1 is collecting requirements from EGI Operations team, NGIs and Resource Providers. The requirements from the eight EGI-Engage competence centers and, in general, from the EGI users are gathered in WP6/SA2 that coordinates the provisioning of services for scientific communities. Then, the WP3 and WP6 activity managers are working closely to identify the most relevant e-Infrastructure commons requirements for users. Technically all the requirements should be stored in the EGI ticket system based on RT and periodically prioritised.
* EGI accounting portal collects some statistics including the total CPU time consumed by a user (or by a Virtual Organization). In this way a user may actively monitor the use of its resources. I improve the document mentioning explicitly this information.
* I add both a footnote with a description and a reference to the RFC on X.509 certificates.
* I agree with you that bug tracking important, and not only during the development activities. EGI and CANFAR have a bug tracking system. In the section describing the federation roadmap I add a sentence specific to bug tracking.
 |

**Additional comments**

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

|  |
| --- |
| **From reviewer:** |
| No additional comments. |

**Detailed comments on the content**

| **N°** | **Page** | **§** | **Observations** | **Reply from author(correction / reject,  …)** |
| --- | --- | --- | --- | --- |
| **1** | **7** | **2** | **‘…high data volumes.’ -> I would try to quantify this** | **Done** |
| **2** | **9** | **3** | **‘…of very large astronomical datasets.’ -> I would quantify the size level** | **Done** |
| **3** | **10** | **3.1** | **X.509 certificate -> mention in Glossary or in note?** | **I add a note and a reference.** |
| **4** | **24** | **6** | **‘The archive houses…collections’ -> mention expected growth rate per year?** | **I add the last year growth rate.** |
|  |  |  |  |  |
|  |  |  |  |  |

**English and other corrections:**

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

*English and typo corrections have been made directly in the document*