

**Support for Genome Analysis and Protein Folding**

**EGI VIRTUAL TEAM**

**PROJECT FINAL REPPORT**

|  |  |
| --- | --- |
| **Date** | 13 April 2015 |
| **Version and Status** | Final |
| **Document Link** | <https://documents.egi.eu/document/2586> |
| **Project Initiation Document** | <https://documents.egi.eu/document/2149> |
| **Virtual Team** | Support for genome analysis and protein folding |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Project Title:** | Support for genome analysis and protein folding (GAPF) | | | |
|  |  |  |  |  |
| **Project Leader:** | Afonso Duarte |  | **Area:** | Outreach to user communities |
|  |  |  |  |  |
| **Project EGI.eu Support:** | Gergely Sipos |  | **Beneficiary:** | EGI |
|  |  |  |  |  |
| **Start date:** | 2014-04-23 |  | **End date:** | 2015-05-20 |

|  |  |
| --- | --- |
| **Short Project Description:**   |  | | --- | | This EGI Virtual Team (VT) will investigate and report on applications required and used by the Protein Structural Biology and Protein/DNA/RNA Sequencing communities within the context of the European Grid Infrastructure. The VT will comprise membership from NILs, Champions and users from the different communities involved. |   **Objectives:**  The main objective of the VT is to increase awareness amongst the community of the existing services and applications of EGI that fall inside their expertise area, and to increase the attractiveness of EGI for researchers of these fields by the further development of the e-infrastructure. The following supporting aims will help to achieve the major goal.  1. Identify tools available in the EGI e-infrastructure relevant for the VT interested community.  2. Identify reusable tools and scientific applications relevant for the VT interested community not yet supported by EGI, and make these available on the EGI production infrastructure.  3. Develop outreach materials to disseminate relevant applications to the target community.  4. Identify synergies (knowledge networks) within the users in order to increase the EGI usage experience and increase the number of users.  5. Organize training and promotional sessions to disseminate the services, tools and applications to potential users.  **Initial Scope:**  **Listing of the working status of applications in the areas of Genetics and Structural Biology present in AppDB.**  **Integration of new Use cases in the fields of Genetics and Structural Biology in the EGI FedCloud**  **Final Scope:** |
| **Listing of the working status of applications in the areas of Genetics and Structural Biology present in AppDB.**  **Integration of new Use cases in the fields of Genetics and Structural Biology in the EGI FedCloud** |

| **Summary of Findings:** |
| --- |
| The GAPF VT produced two main achievements:   1. Setup of four new use cases to use the EGI FedCloud   **READemption**  READemption is a pipeline for the computational evaluation of RNA-Seq data. The use case consists in running the analysis workflow on the EGI Cloud Federation.    Application:<http://pythonhosted.org/READemption/>  (Outreach stage during the VT execution period)  **Trufa**  TRUFA (Transcriptomes User-Friendly Analysis) is a webserver designed to help researchers in genomics to perform de novo RNA-seq analysis. The goal is to exploit Cloud Federation resources from the TRUFA portal.  Application: <https://trufa.ifca.es/web>  FedCloud Wiki: <https://wiki.egi.eu/wiki/FedCloudTRUFA>  (the use case is in assessed stage and now we are waiting for the community to start the integration on the EGI FedCloud)    **Chipster**  Chipster is a user-friendly analysis software for high-throughput data. It contains over 300 analysis tools for next generation sequencing (NGS), microarray, proteomics and sequence data. Chipster's client software uses Java Web Start to install itself automatically, and it connects to computing servers for the actual analysis. Chipster is open source and the server environment is available as a virtual machine image.  FedCloud Wiki: <https://wiki.egi.eu/wiki/FedCloudChipster>  Application <http://chipster.csc.fi/>  (the use case is fully integrated in the EGI FedCloud. A first pilot is just started with a first real user. More pilots are planned for the next months)  **RSAT**  RSAT provides a series of modular computer programs specifically designed for the detection of regulatory signals in non-coding sequences. (Ongoing Use Case)  Application: http://rsat.ulb.ac.be/  Fedcloud Wiki: https://wiki.egi.eu/wiki/FedCloudRSAT  (currently discussing technical details with EGI FedCloud ) (the use case is in assessed stage and now we are waiting for the community to start the integration on the EGI FedCloud)   1. Production of a resume table describing the applications deposited in the EGI AppDB in the fields of genomics and structural Biology. |

**Accomplishments:**

|  |  |
| --- | --- |
| **Milestones** | **Comment (including variances)** |
| Table of the Genomics and Structural Biology applications from the AppDB | This table includes all the applications from Structural Biology and Genomics deposited in the EGI AppDB. The information gathered includes: working status, Grid/Cloud application, running status of the application in the EGI infrastructure, link to developers. |
| New Use Cases | Four Use Cases currently in different implementation stages in the FedCloud. |
| READemption EGI webinar | As an outcome of the READemption Use Case an EGI webinar was organized. Webinar presented on the 27th November 2014 (https://indico.egi.eu/indico/conferenceDisplay.py?confId=2345 |

**Issues:**

|  |
| --- |
| None. |

| **Root Cause of Variances:**  *If any* |
| --- |
| During the project development new use cases were identified, this identification was done in an extended period leading to an extension of the project implementation. This extension could be expected as the identification of new use cases can take place at any given moment of the project time line. |

**Documentation:**

*Documentation produced by the project*

|  |
| --- |
| **Wiki of the VT:**  https://wiki.egi.eu/wiki/VT\_GAPF  **Table of applications:** [https://docs.google.com/spreadsheet/ccc?key=0Ama69JoAAogvdHFzQi1UamwxN0MtVS1GUEV 4ZmVGWXc&](https://docs.google.com/spreadsheet/ccc?key=0Ama69JoAAogvdHFzQi1UamwxN0MtVS1GUEV%204ZmVGWXc&)usp=sharing  **Meeting minutes:**  <https://docs.google.com/document/d/1rDJ-NhMOBR0WORt8Pic4Db-YCqHlBPxyey0xJGPT2aI/edit?usp=sharing> |

**Recommendations:**

*Lessons learned – what went well, what should be done differently*

|  |
| --- |
| **The GAPF VT achieved the main objectives.**  **The engagement with the Life Sciences communities indicated that the bridge between EGI and LS communities is getting shorter as LS is looking for new solutions and partners for their projects. This is an excellent opportunity for all EGI and EG members to engage new Use Cases and users.**  **Recommendation for the future: setup of funded projects (i.e. pilots with small funding) to support bridging between EGI and LS users that could allow support of developers in LS groups for a better Grid/Cloud user experience.** |

**The “full name”VT membership**

*list all those involved in the team and doing the work*

|  |  |  |  |
| --- | --- | --- | --- |
| Country Code: | Name: | e-mail: | Role: |
| PT | Afonso Duarte | aduarte@itqb.unl.pt | ITQB-UNL |
| GR | Fotis E. Psomopoulos | fpsom@issel.ee.auth.gr | Aristotle University of Thessaloniki |
| DE | Konrad Förster | konrad.foerstner@uni-wuerzburg.de | University of Wuerzburg |
| IT | Daniele Cesini | daniele.cesini@cnaf.infn.it | NGI\_IT |
| PT | João Pina | jpina@lip.pt | IBERGRID |
| GR | Kostas Koumantaros | kkoum@grnet.gr | NGI\_GR - GRNet |
| NL | Nuno Ferreira | nuno.ferreira@egi.eu | EGI.eu |
| NL | Sara Coelho | sara.coelho@egi.eu | EGI.eu |
| NL | Gergely Sipos | gergely.sipos@egi.eu | EGI.eu |
| NL | Diego Scardaci | diego.scardaci@egi.eu | EGI.eu |
| NL | Neasan O'Neill | neasan.oneill@egi.eu | EGI.eu |
| ES | Jesus Marco de Lucas | marco@ifca.unican.es | CSIC |
| FR | Tiphaine Martin | tm483@medschl.cam.ac.uk | CNRS |
| FI | Kimmo Mattila | kimmo.mattila@csc.fi | CSC |
| ES | Beatriz Ranz Ribeiro | branz@ifca.unican.es | CSIC |
| FI | Aleksi Kallio | aleksi.kallio@csc.fi | CSC |
| ES | Fernando Aguilar | aguilarf@ifca.unican.es | CSIC |
| FR | Jacques van Helden | Jacques.van-Helden@univ-amu.fr | Aix-Marseille Université |

The “full name”VT Stakeholders

*list all the stakeholders who had an interest in your work, i.e. all those you consulted or wanted to consult*

|  |  |  |  |
| --- | --- | --- | --- |
| Country Code: | Name: | e-mail: | Role/Organisation: |
| FR | Johan Montagnat | johan.montagnat@cnrs.fr | CNRS, LSGC |
| NL | Alexandre Bonvin | a.m.j.j.bonvin@uu.nl | eNMR/WeNMR (Dutch NGI) |
| ES | Ignacio Blanquer | iblanque@dsic.upv.es | UPVLC |
| NL | Enol Fernandez | enol.fernandez@egi.eu | EGI.eu |
| UK | Rafael Jimenez | rafael.jimenez@elixir-europe.org | EMBL |
| FI | Tommi Nyrönen | tommi.nyronen@csc.fi | CSC |