Project Acronym: MuG

Project title: Multi-Scale Complex Genomics (MuG)

Call: H2020-EINFRA-2015-1

Topic: EINFRA-9-2015

Project Number: 676556

Project Coordinator: Institute for Research in Biomedicine (IRB Barcelona)

Project start date: 1/11/2015

Duration: 36 months

Deliverable 2.6: Initial Exploitation Plan based on Market Knowledge

Lead beneficiary: Institute for Research in Biomedicine (IRB Barcelona)

Dissemination level: PUBLIC

Due date: 30/04/2017

Actual submission date: 05/05/2017

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

Document history

Version	Author(s)	Partner	Date	Comments
0.1	Anna Montras	IRB Barcelona	28/04/2017	First draft
0.2	Josep Ll. Gelpí	BSC	02/05/2017	Inputs on sustainability
0.3	Andrew Yates	EMBL-EBI	03/05/2017	Minor corrections
0.4	Brian Jiménez	BSC	04/05/2017	Additions on PyDock tool
0.5	Mark McDowall	EMBL-EBI	04/05/2017	Minor corrections
0.6	Javier Conejero	BSC	04/05/2017	Minor additions and corrections on PyCOMPSs
0.7	Laia Codó	BSC	04/05/2017	Revised list of integrated tools
1.0	Anna Montras	IRB Barcelona	05/05/2017	Approved by Supervisory Board

Table of contents

Glossary

EOSC: European Open Science Cloud

FAIR: Findable, Accessible, Interoperable and Re-usable

Executive summary

The present report contains an assessment of the potential means of exploitation and sustainability for the MuG VRE within the context of the European Research Infrastructures.

The report builds on the needs of the 3D/4D genomics community, analyses the added value provided by the Virtual Research Environment developed by the project and how the different members of the community and the different stakeholders will be able to benefit from or contribute to the sustainability of the VRE.

The report analyses the measures that have been undertaken or that are foreseen to secure the technical sustainability of the platform and it also discusses the costs associated to its future maintenance and the possible sources of financing such costs. The importance of community building is emphasized and the measures to be implemented during the second half of the project to reach a larger and more diverse audience (including academia and industry) are discussed, as well as the right timings to do so to success.

The contents of this deliverable are often connected to D2.4, in which the implementation of the dissemination and training plan is discussed and corrective measures are described to improve community engagement.

The different elements of the business plan which have been outlined in D2.6 will be addressed in the coming months and the final outcome of the decisions, actions undertaken and achievements of the MuG consortium towards exploitation will be presented in D2.7 in M36.

1 INTRODUCTION

1.1 Virtual Research Environments

One of the essential aspects of a Virtual Research Environment is that its creation is driven by community needs. A VRE aimed at the research community, should assist the target community during the complete lifecycle of their research process in order to make it attractive to different market /community segments and achieve a broad uptake by the community[1].

Collaborative research has become essential to tackle complex research challenges. Managing collaboration is a complex challenge that requires extensive resources and proper infrastructure. Research projects increasingly involve researchers from **geographically distributed** research groups. In addition, research challenges require increasing **cross-disciplinary** collaboration. Such geographically distributed and multidisciplinary scenario demands tools to enhance collaboration among researchers in order to tackle shared challenges efficiently.

A VRE should enable collaborative research activities, helping researchers manage a complex range of tasks. In order to be useful to the user community, a VRE should learn from the needs of the target community and provide tools to efficiently address those needs.

Some key concepts that any successful VRE should include are[2]

1. Be integrated with current policies and infrastructure
2. Reflect its users and be driven by their needs
3. Facilitate research processes and collaborative research
4. Evolve over time to meet changing user needs

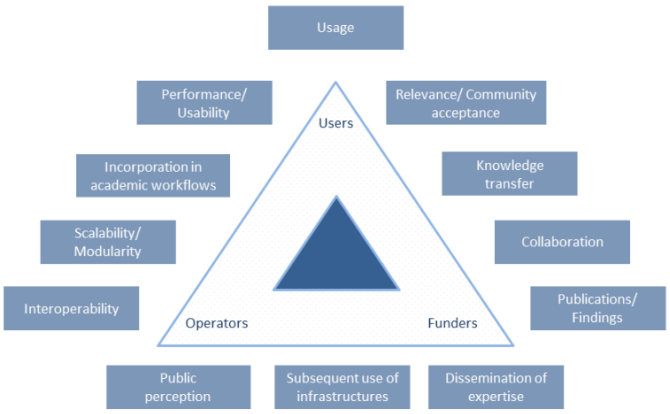


Figure 1: Success factors and measurable success criteria, weighted by relevance for the different stakeholder categories (Buddenbohm *et al.,* 2015)

The present document reflects on the sustainability and potential exploitation models for the MuG VRE. It starts with an assessment of the needs and evaluation of the impact made on the community during the first 18 months of development, as a key tool to pave the way for future uptake of the services and discusses potential business models and sustainability measures to be adopted.

[1] https://www.jisc.ac.uk/full-guide/implementing-a-virtual-research-environment-vre

[2] Frameworks for Virtual Research Communiities – A scoping study. A.P. Robbins. Branz Study Report 2009. ISSN: 1178- 4938. http://www.branz.co.nz/cms_show_download.php?id=8d6108e4becdcc34a65b0f7079f0e0d5e3ffde16

[3] Buddenbohm S, Enke H, Hofmann M, Klar J, Neuroth H and Schwiegelshohn U (2015). D Lib Magazine 21 (9/19). DOI: 10.1045/september2015-buddenbohm

1.2 Outlining an exploitation plan

Outlining a business model from scratch can be a difficult task in such a complex concept as a Virtual Research Environment with a user-base composed largely of academia researchers and mostly based on open access principles.

A useful tool to assess the exploitability of the MuG VRE, as a platform driven by community needs, is the "Lean Canvas" model, an adaptation of the "Business Model Canvas"[4], which focuses on the *customer-problem-solution* paradigm and which is a good tool to capture the unique value proposition of the product.

A template for the "Lean Canvas" model is provided in Annex I. The following sections (2 to 7) follow the structure of this business model tool and the contents are the results of discussion sessions with different members of the consortium. This is a good starting point that has allowed us to identify where the MuG VRE is adding value and what are the potential business models that would allow us to exploit this added value in the long-term.

2 THE PROBLEM: A research community with urgent needs

Scientific excellence triggers the need for more efficient tools to tackle the demands of a given research community. 3D/4D genomics is a clear example of a research community urgently calling for an e-infrastructure that can help handle the research outcomes in an efficient manner.

MuG is anticipating the problem. Our pilot projects are lead users of the MuG VRE, creating a pull for the technology with their cutting-edge research and increasingly demanding needs.

Community needs are indeed evolving fast, motivated by the continued advent of new experimental techniques (e.g. 3C- micro-C, Hi-C, Capture C, MNaseq, ATACseq, RICC-seq, FISH, etc.) to study chromosome folding. Data is locally stored using non-standardized formats and making post-processing close to impossible, and the disconnection between 1D and 3D/D data. This lack of consensus and disconnection, in turn, leads to a lack of standards in analyses and simulations and lack of connection between different resolution-level simulation tools. This lack of standards is threatening to become a bottleneck that slows down progress in the field.

2.1 Real biological problems and existing alternatives

MuG should develop solutions to the ***real problems biologists working in 3D/4D genomics are facing***, and the interaction between developers and experimentalists is crucial to fulfill this objective. To this end, MuG relies on 3 pilot projects that focus on selected use cases, conducted by high-profile research groups who generate new requirements and, implicitly, face problems that require new solutions.

An exercise was conducted with the pilot projects in which they identified a number of **problems** faced during their everyday research activity, the **existing alternatives** and how they would expect the VRE to improve their work. An example of inputs provided by pilot projects is provided in Figure 2.

[4] Osterwalder, Alexander; Pigneur, Yves. (2013). Business Model Generation. Hoboken, NJ: Wiley.

Figure 2: Exercise undertaken with the pilot projects to study the "problems", current alternatives and solutions that can be offered by the VRE

A way to sum up the exercise is that end users spend a lot of time looking for the most suitable tools using different sources. On the other hand, there is the need for local storage and disconnection.

Pilot projects, as end users, are key in defining the problems but also contribute actively to shaping the VRE by providing technical specifications, assessing the development of infrastructure and service portfolio and testing the resulting prototypes.

3 SOLUTIONS

3.1 Tools integration

In order to achieve long term sustainability of the platform, it is necessary that it has the capability to adapt to future needs. Based on the inputs from pilot projects, VRE developers will have clear hints on the needs for eventual scale-up and need for additional tools to extend the services to the wider community that will sustain the VRE.

In order to fulfill the evolving needs of the end-users, the tools offered by the VRE cannot be limited to those integrated in the platform during the 3-year grant duration. Even during the development stage of the VRE, MuG has already received inquiries from developers to integrate tools in the platform. The roadmap for integration of third-party tools (both technical and relating to legal framework) is being discussed internally at the moment.

3.2 Services offered by the VRE

The MuG has the capacity to offer a wide range of solutions to tackle the different problems identified by our end users. The service offer of the MuG VRE could be summarized as follows:

1. Source of information in 3D/4D genomics field
2. Multidisciplinary hub to interact with other 3D/4D genomics community members (forum, etc)
3. Sole access point to integrated (up-to-date) analysis tools and allow the execution of personalized analysis workflows through a user-friendly Graphical User Interface
4. Common data repository, compatible with other repositories and with personal workspace
5. Multi-resolution browser
6. Training activities (tailored to customer segments)
7. User support tools and tutorials
8. Consultancy (longer-term)

4 KEY METRICS

4.1 Uptake potential by the community

An indirect way to measure impact potential and estimate the size of the target user base is to use statistics of analysis tools, databases, and related platforms progressively being integrated and tailored to 3D-genomics community user-needs in the MuG VRE portal.

Below are statistics of key tools already integrated in the VRE allowing measurement of (i) the performance of the current solutions and (ii) the potential size of the user base:

- **NAFlex** is a web tool for Nucleic Acid flexibility study. It has an average traffic of **2,200 users/year**.
  BigNASim, a database and analysis portal for nucleic acids simulation data, already has over **550 users** since its publication in late 2015. **Europe (36%), USA (20%).**
- **PyDock** is the original protein-protein docking protocol which has been adapted to protein-DNA docking (pyDockDNA). PyDock is available as a very active web server, pyDockWEB, with more than **6000** jobs served and **1200** unique users since its publication on November 2011.
- **MDWeb** (Molecular Dynamics web-based simulation platform): online since March 2014, the average traffic on this platform is of **6,700 unique visitors /year**. **India (27%), USA (14%), Europe (15%)**
- **NucleR,** a package for non-parametric nucleosome positioning determination receives an average of **2,800 downloads per year**.
- **TADbit** is a computational package for the analysis of 3C-based experimental data. Its Github repository registers an average of **950 unique visitors per year.**
- **TADkit** is a 3D genome visualization tool that creates interactive 3D representations of chromatin conformations modeled from 3C-based interaction matrices. The download site registers **~480 unique visitors yearly**.

A key condition to achieve long-term sustainability and successfully exploit the platform is to ensure that user needs are met and to identify the unique value proposition that distinguishes the MuG VRE proposed model from other existing (or under development) platforms or initiatives related to the field.

5 UNIQUE VALUE PROPOSITION

The following value proposition derives from the inputs received from the pilot projects in the consultation presented in section 2.1.

The MuG Virtual Research Environment supports the expanding 3D/4D genomics community by developing tools to **integrate** the navigation in genomics data **from sequence to 3D/4D** chromatin dynamics data.

The value proposition complies with the condition of being a simple and compelling text that remarks why the MuG VRE is different from other existing solutions. The added value of the VRE for the pilot projects lies in the integration of different steps required in the evaluation of their research results (e.g. the possibility to **integrate analysis and browsing**).

5.1 High level concept

A high-level analogy to which the concept offered by the VRE could be compared is the navigation concept offered by Google Earth, in which the user can navigate across scales, applying the suitable resolution for each scale.

5.2 "Unfair" advantage

The MuG VRE provides a single (virtual) access point to both data and tools allowing users to configure their own workspace.

MuG's main competitive advantage is the combination of **development time invested** and the need for a **unique combination of expertise** from world leading institutions working closely together towards a common objective to fulfill the needs of this specific community.

The offer described by the value proposition encompasses, a single point of access to tools and community tailored workflows, interoperability, user support and a stable sustainable infrastructure.

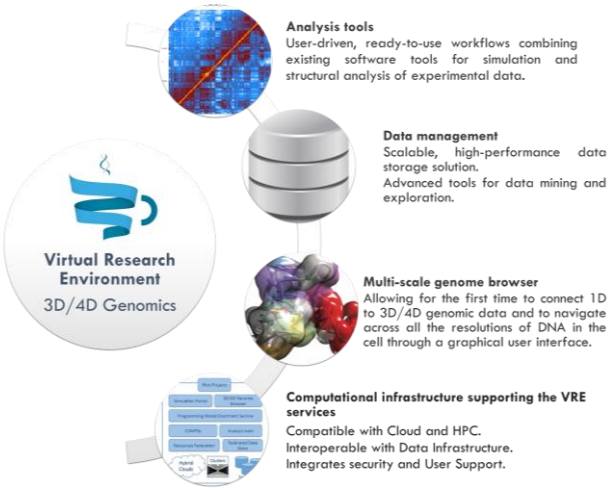


Figure 3: Detail of features offered to the community

6 CUSTOMER SEGMENTS AND CHANNELS

Although different types of "customers" can be identified who may be interested in different features offered by the VRE portal, the dissemination and training activities have been first and foremost focused on addressing the Unique Value proposition, i.e. engaging end-users who will benefit from the integration of tools and the use of tailored workflows. This is the main strength of the VRE and it should be enhanced and properly communicated to the right audience.

One of the key tools to achieve this is to organize hands-on training activities that allow users to work on real use cases. The strategy is being applied successfully, having achieved great satisfaction results on the first training course offered in April 2017 at EMBL-EBI (see D2.4 for details).

6.1 Beneficiaries of the developed solution

We have identified different types of beneficiaries among (i) the research community, (ii) Education and (iii) Industry.

CUSTOMER	BENEFIT	CHANNELS
EDUCATION		
Education (Biology)	Adopt MuG VRE as a tool to enhance teaching at undergraduate level (bioinformatics, etc.). User support and training	Social media General media Training targeting university teachers.
RESEARCH (END-USERS)		
Genomics and personalized medicine (incl. ICGC-PAWG, Blueprint, EGA, Genomics England)	Tools to connect data from different levels delivering a realistic model of chromatin User-friendly interface User support and training	Dissemination in scientific conferences. Publication of scientific results by pilot projects. Social media Offer of (customized) training courses
Biosimulation and 3D Genomics (e.g. ABC, ExTASY, HEC-BioSim, BioExcel CoE)	Platform with tools covering the whole scale. Connect different sources of data and provide data storage facilitating re-use and reproducibility. User support and training	Dissemination in scientific conferences. Publication of scientific results by pilot projects. Social media Offer of (customized) training courses
TOOL DEVELOPERS		
Developers of analysis /visualization tools	Integration in VRE platform, gaining visibility and users.	Bioinformatics conferences Training events
INDUSTRY		
Sequencing Instrument Vendors	MuG will help processing output data and generate valuable information through the integration of analysis and visualization tools.	Partners (CNAG) Innovation events / industry fairs Published results Social media (LinkedIn, Twitter)
Biotech / Pharma Industry	Data and processed information generated by MuG Consultancy services	Published results Innovation events /industry fairs Social media (Twitter, LinkedIn)
OTHER STAKEHOLDERS		

HPC for Life Sciences initiatives (e.g. BioExcel, EESI2)	MuG contributions to Exascale roadmap, facilitating link with target community.	Shared project partners e-infrastructure events
Public health	Data and processed information generated by MuG. Consultancy services.	Results published and/or presented in specialized events

6.2 Engagement

Traffic in the MuG portal is monitored through Google Analytics, which is implemented since May 2016 ***(~150 new users/month)***. The results are analysed in detail in D2.4, as well as the performance in social media like Twitter ***(102 followers)***. As discussed in that report, traffic is progressively increasing and a clear relationship is observed between key events (high-impact publications from pilot projects, training events, VRE releases) and number of new users. In order to keep track of users and their profiles and increase the user experience while the VRE is under development, a list of registered users was established with the aim to create the foundations for a 3D genomics hub. In the short term, the so-called "MuG community" has been conceived as an interest group whose members will be informed when new features are implemented and will be able to interact through the MuG forum.

So far the community counts with ***~40 registered members***, including potential end users, tools developers, etc. some of which have expressed their interest in becoming more actively involved in the VRE definition or in offering their tools (developers) through the MuG VRE. Terms of collaboration and integration timeline with interested community members will be defined in the near future.

The following figure symbolizes the expected timeline relating VRE development and release and adoption by the community.

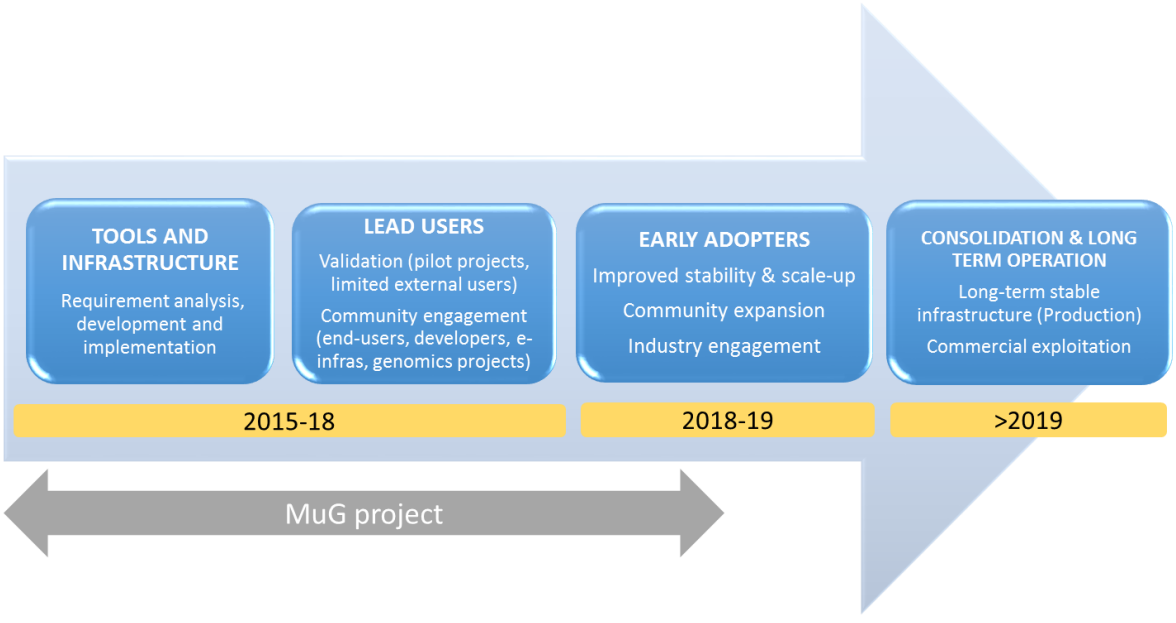


Figure 4: Expected timeline for the development and uptake of MuG VRE services by the community

6.3 User feedback

In addition to feedback from our pilot projects, feedback from external users is highly useful to prevent any bias in understanding the community needs.

The feedback collected in training events, where users have had a chance to work hands-on with the VRE portal and computing platform is a very valuable input.

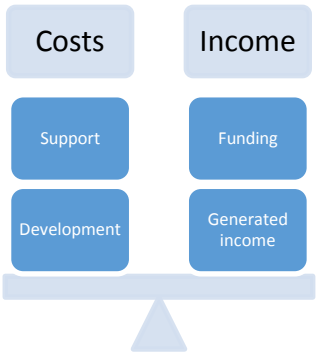
Surveys completed by the community are more useful when users have interacted with the services of the portal. As a future measure, it would be useful to invite VRE registered users to fill in a satisfaction survey after they use it.

7 COSTS AND REVENUE STREAMS

7.1 Costs

For an open source project/infrastructure to become sustainable, it must reach a point in which it meets its own costs, including[5]:

- Hosting and supporting services
- Development, updating and maintenance of software (incl. integration of new tools)
- User support and training
- Marketing and communications costs
- Costs associated with governance



While project funding (3 years in the case of MuG) will allow to develop the VRE and reach production, at the end of the grant it will be necessary not only to provide support for the current tools and infrastructure but also to cope with the increasing needs of the end-user community, which is sure to generate a demand for new tools and data formats rather soon.

A key task to undertake during the next few months is to evaluate the cost of maintaining the developed platform in a window of 10-years from present.

The VRE is being developed in such a way that integration of new tools and upgrades can be performed in the simplest possible way, thus minimizing future costs.

7.2 Potential Financial sustainability measures

To achieve sustainability it is important to secure financial means.

Public funding to sustain the non-commercial exploitation side:

- EU funding, either alone or as use case/community within large EU e-infrastructure initiatives (see section 9.4 for already taken actions).
- Apply for funding in countries where a broader user-base is established at the end of the project

Co-funded /in kind-contributions by users /e-infrastructures/ developers

[5] OSS Watch. Sustainable open source (April 2014). http://oss-watch.ac.uk/resources/sustainableopensource

- Support from large e-infrastructures: services and resources provided through agreements with national or international initiatives (PRACE; EUDAT, BioExcel, ELIXIR).
- Support from international genomics initiatives: MuG aims to be the reference platform for chromatin analysis in the field of genomics and personalized medicine.
- Contribution from third party developers in integrating software tools (development costs – manpower - of integration would be shared).
- Training sessions with tailor-made content for different user segments).

Services to industry:

- Licenses to industry for commercial use of MuG VRE services (adapted VRE versions with specific features for industry users) (customer segmentation).
- Potential commercial products may arise from the developed tools.
- Consultancy for data exploitation: data analysis or access to data offered to industry (e.g. pharma). MuG will work towards establishing a public-private collaboration framework to apply 3D genomics to study epigenomics and non-coding DNA as a new generation of drug targets. Activities in this direction will strategically be started once the VRE is ready for production and has an established user-base (during the last year of the project) to prevent losing the momentum.

7.3 Exploitable results

Below is a preliminary list of the main innovations expected from the project which are subject to evaluation for exploitation potential. Related consultancy services are being negotiated with a third-party provider to assist the consortium in the second half of the project in which results start to materialize.

Table 1: Potential outcomes of the project and exploitation routes. *Information in this table is provisional. External consultancy services will be sought to evaluate the most suitable means of protection.

Related Innovation	Project WP	Result type & description	Potential Product/ Exploitation route
Visualizer of genomic machinery with 4D resolution	**WP3**	**Software**	Tool: visualization software
Optimized software modules and workflows	**WP3** **WP6**	**Software**	Service: Compute Platform
Models of chromatin reorganizations / Analysis of raw data and existing annotations	**WP3** **WP6** **WP7**	**Long tail of data**	Consultancy Service
e-Infrastructure supporting the MuG VRE	**WP4** **WP5**	**Computational infrastructure**	Service: Compute Platform
VRE API	**WP6**	**Software**	Service: Compute Platform
VRE Graphical User Interface concept	**WP5** **WP2**	**Software** **Design**	Service: Compute Platform

8 GOVERNANCE and LEGAL STRUCTURE

The **governance structure** should include at least the core developers of the MuG e-infrsatrucutre and potentially also some of the tools developers. It is also important to continue to have representation from the user community (role undertaken by pilot projects in the consortium).

- WP4: Data Management (EMBL-EBI)
- WP5: Computational Infrastructure (BSC, IRB)
- WP6/WP3: tools developers (UNOT, CNAG, BSC)
- User community representation
- External advisors

In order to facilitate governance and being operational, it is necessary to constitute a **legal entity** that has the capacity to sign formal legally binding agreements with third parties, apply for funding, hire personnel, etc.

During the second half of the project, the 6 consortium partners will enhance discussions on their interest/feasibility/contribution to a potential future organization for long-term sustainability and exploitation of the MuG VRE.

The different options available for constituting a legal entity are being explored and discussed with other e-infrastructure projects. Among options that have been explored so far are:

- **ERIC: European Research Infrastructure Consortium**. A legal entity specific for European Research Infrastructures. The principal task of ERIC is to establish and operate new or existing research infrastructures on a non-economic basis. A much more complex path than an EEIG, requiring approval at national level and not considered the most suitable option for MuG[6].
- **European Economic Interest Group (EEIG)**: A type of legal entity designed to allow companies in different countries to do business together in a simpler way. In the case of the MuG consortium, it would facilitate access to future funding as a consortium unified under a single entity. This figure was discussed with BioExcel, West-Life and MaX and it is regarded as a possible outcome, although the pros and cons need to be evaluated further. In the case of MuG, it may not be an option if consultancy services against payment are to coexist with services to academia as part of the business model.
- Other forms of Joint Venture between the partners may also be explored once the end-product is more defined and balance between commercial and non-commercial activities is established during the second half of the project.

9 SUSTAINABILITY PLAN

As stated in the e-IRG Roadmap 2016[7] , one of the key recommendations for user communities is to "organize themselves" in order to formulate their e-Infrastructure requirements, actively participate in the innovation of e-infrastructure services and contribute to standards. The MuG project encompasses the needs of the 3D/4D genomics community and provides the necessary framework and infrastructure to fulfill those needs. MuG aims at nucleating the 3D/4D genomics community and provides tools to facilitate its access to HPC. The MuG VRE is developing a cloud-based computational

[6] https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric
[7] E-IRG Roadmap 2016. Paving the way towards a general purpose European e-Infrastructure. December 22, 2016. http://e-irg.eu

infrastructure to support the deployment of a series of software tools addressing several levels of analysis of the genome. The needs of such analysis tools, including computationally demanding molecular dynamics simulations to tools for NGS and Hi-C data analysis, where the stress is on data management, and high throughput data analysis.

On the other hand, key recommendations derived from the e-IRG report for e-infrastructure providers are the need to *work closely together* and to speed up convergence.

9.1 Technical sustainability challenges

MuG counts in the consortium with key partners, including BSC, CNAG-CRG, IRB Barcelona and EMBL-EBI:

- BSC is a major resource provider in **PRACE, and Spanish National Supercomputing Network,** key member of EU Computational initiatives (**ELIXIR**, **BioExcel CoE**, etc.) and partner in large omics initiatives (e.g. Human Brain, BluePrint, ICGC-PanCancer).
- CNAG-CRG is one of the largest sequencing centers in Europe, and has a strong programme in genomics consultancy services.
- IRB Barcelona, one of the largest users of HPC resources in Europe, is an active participant in the ELIXIR project and BioExcel CoE.
- EMBL-EBI is the major center in Europe for supporting bioinformatics tools and data and is the main player in all the big omics EU consortia. It contributes the necessary experience in setting BioData infrastructure, becoming key for the long-term sustainability of the MuG VRE.

The MuG infrastructure is currently (M18) based at IRB-BSC clouds (Barcelona), with a satellite installation at EMBL-EBI's Embassy cloud. The main pillars of the MuG infrastructure are:

- The MuG central VRE infrastructure at IRB, including MuG's main data repository
- INB-BSC Cloud at the Barcelona Supercomputing Center
- HPC computers (MareNostrum and MinoTauro), and active archive at BSC
- Embassy Cloud at EMBL-EBI

Software infrastructure design will assure the interoperability of MuG tools, and the computational infrastructure will provide execution environment for tools and workflows, with access to HPC facilities.

MuG will also hold a data and metadata repository, and access to several public data repositories (ArrayExpress, ENA, EGA, PDB, BIGNASim).

The VRE aims to provide a single (virtual) access point to both data and tools allowing users to configure their own workspace.

The current set-up is powerful enough to service the initial user community, but additional actions are necessary in order to guarantee scalability and address the needs of a wider user-base. As a measure for long-term sustainability of the VRE, infrastructure needs to be interfaced to European e-infrastructures (Figure 5).

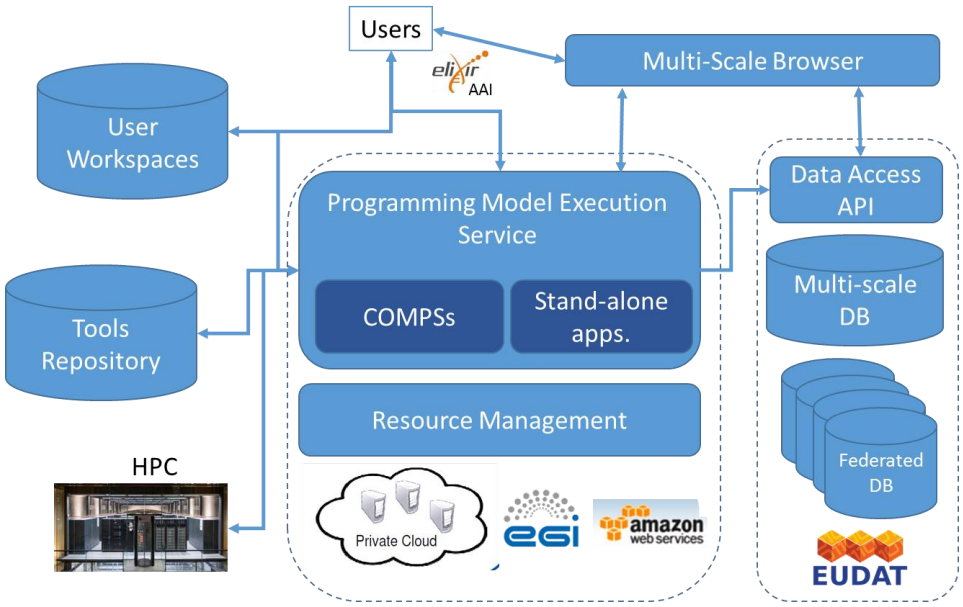


Figure 5: Projected layout of the MuG VRE infrastructure with a view to long-term sustainability (Adapted from D5.1)

9.1.1 Risks and measures for long-term sustainability

The major **bottlenecks** that the initial infrastructure will face, in order to complete the development stage and move fully into production (Figure 6) are summarized below:

1. Inefficient data transmission, especially relevant when analysis requires voluminous data that is not co-located.
2. Lack of computational resources to scale the required analyses as the user community grows.

Figure 6: Technology Readiness Levels (Adapted from **© EGI Foundation**, Personal communication)

Some key limitations that need to be tackled through securing external services or resources are listed in Table 2.

Table 2: Required services and type of resources required for VRE scalability and sustainability

Purpose	Limitation	Required external service	Quantity of resources	Target /Potential providers
Homogeneous data space	Data is distributed in a series of repositories	Setup based in B2Share or equivalent	TB to PB (depending on local resources available – tbd)	EUDAT, Zenodo
Data Integration HPC	Data management is restricted in HPC environment	Extend integration HPC & EUDAT at BSC	Hundreds of TB for simulation raw data	EUDAT
Integration of MuG tools in	No possibility of open workflow management	Global access to Virtual Appliances	Highly depending on the growth of	EGI

general purpose Workflows	within MuG infrastructure	for MuG designed tools	the user community.	
Scalability	MuG clusters may not be sufficient to handle required capacity	Scaled-up PaaS	Highly depending on the growth of the user community.	IndigoDataCloud, Embassy Cloud; EOSC
Long-term sustainability	MuG local equipment are tied to project usage.	Full integration in the European e-infrastructure landscape	Highly depending on the growth of the user community.	EUDAT, EGI, EOSC

9.2 Technology development Roadmap

After releasing the first prototype of the computational infrastructure in August 2016 (M10), the development of the MuG VRE has moved fast along the development phases and in early April 2017 (M18), a first release of the actual MuG workspace integrating the first set of tools ready for testing with an initial set of users was already in place.

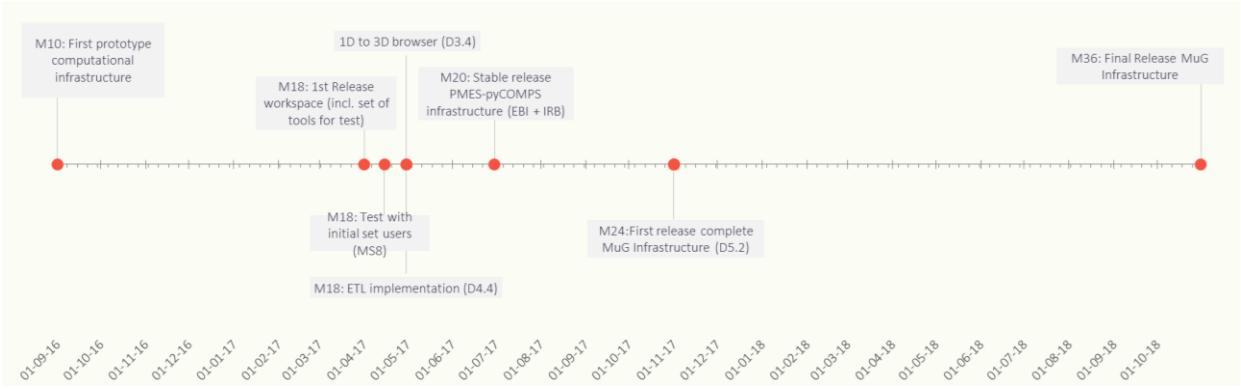


Figure 7: Development roadmap for the MuG VRE infrastructure

Key Milestones for the infrastructure release plan can be summarized as follows:

- November 2015: Project Start
- August 2016: First prototype of computational infrastructure
- **April 2017** (M18)
  - 10/04:First release of MuG Workspace, integrating already available set of tools: TADbit, NAFLex/BigNASIM, pyDockDNA , Nucleosome Dynamics, Chromatin Dynamics, MC-DNA, DNA Shape Scan, MD Energy Refinement & 3D consensus .
    Dual deployment, IRB and Embassy Cloud at EMBL-EBI (processing data from EMBL-EBI hosted repositories).
    Tested at 1st user Workshop at EMBL-EBI.
  - 30/04: 1D to 3D browser (D3.4)
  - 30/04: ETL implementation (D4.4)
- **June 2017.** Stable release of PMES-PyCOMPSs infrastructure at both IRB/BSC and EMBL-EBI sites. Galaxy interface (PoC)

- **November 2017**. First Release of the complete MuG infrastructure (D5.2), Data access stable and fully documented API (D4.5). To be presented at IRB's BioMed Conference on 3D-4D genomics (where all the 3D 4D genomics community is going to be present).
- **October 2018**: Final Release of MuG infrastructure.

9.3 User support and training challenges

During the second half of the project, interaction with big consortia such as ELIXIR, EGA, BioExcel, ICGC-PanCancer or BLUEPRINT will be enhanced. The recommendations from these organizations will be implemented.

Some of the pillars that sustain the MuG user experience are:

- Multidisciplinary support enabled by MuG partners' balanced background that tackles the whole range of analysis tools.
- VRE portal and user support maintenance

9.4 European e-infrastructures context and positioning

In order to address the sustainability needs of the developed platform, the need to find synergies with other initiatives e-infrastructures is unquestionable.

In addition to the involvement of MuG partners in some key computational projects, which facilitates formal communication, MuG has actively participated as a project in several joint activities with other projects aimed at assessing long-term sustainability and finding synergies:

9.4.1 Interaction with other e-infrastructure initiatives

Interaction with other projects and initiatives in the e-infrastructure landscape becomes essential to harmonize strategy in terms of optimizing resources for a sustainable exploitation model.

- **e-concertation activities**: The 11th e-concertation meeting offered a unique opportunity to the MuG coordinating and infrastructure development team for networking in the context of the e-infrastructure landscape. Synergies with projects in the areas of Life Sciences and Transversal services were identified. Among the potential collaborations identified in November 2015, specific collaborations are already under discussion with **BioExcel CoE** and **West-Life VRE**.
- **BioExcel CoE**. 3 of MuG's partners (IRB, BSC, EMBL-EBI) are partners of BioExcel. MuG will benefit from BioExcel in the organization of training activities (e.g. webinar on nucleic acids flexibility aimed at entry level users is scheduled and a joint 3-day training workshop is under discussion at the moment). The terms for long term collaboration and taking advantage of complementary services offered by both initiatives are being explored.
- MuG is fully committed to the adoption of **ELIXIR** recommendations and tools - e.g. ELIXIR's Authorization and Authentication Infrastructure (AAI). BSC, IRB and EMBL-EBI are active members of ELIXIR.
- **Consultancy with mature e-infrastructures:** MuG counts with **Dr. Tiziana Ferrari**, EGI Foundation technical director, as a member of the Scientific Advisory Board, who contributes essential advice in terms of the measures to be undertaken to enable long-term technical sustainability.
- **Sustainability workshops:** in the context of European e-infrastructures, enhancing collaboration and discussing sustainability strategy becomes essential to facilitate joining forces and facilitating the realization of the EOSC. MuG has participated in two workshops:

- **Design your e-infrastructure** (Krakow, Poland, 27/09/2016**):** The workshop was organized by 4 e-infrastructures (EGI, EUDAT, GÉANT, OpenAIRE) in the context of the Digital Infrastructures for Research (DI4R) in Krakow, Poland.

  MuG was presented as a use case among other e-infrastructures and the needs and potential collaborations required to achieve long-term technical sustainability were addressed.

  The event also served to foster discussion with **EUDAT** for long term storage of MuG data and **EGI** for provision of additional computational resources.

- **BioExcel sustainability Workshop** (Schipol, NL, 23/03/2017): This event organized by BioExcel CoE brought together two CoE projects (BioExcel and MaX) and two VRE projects (MuG and West-Life) to discuss the different models being developed by the these 4 e-infrastructures. Besides discussing alternative business models applicable to each case, an agreement was reached on the need to collaborate in terms of joint dissemination (e.g. cross-linking websites). The discussion was very productive in terms of governance models and financing options.

- **European Open Science Cloud:** Integrating MuG services into the EOSC would be instrumental to facilitate interaction and enhance competitiveness with international initiatives, thanks to the holistic character of the EOSC at EU level.
  MuG has submitted a proposal to be considered as a "Science Demonstrator" in the context of the *EOSC Pilot*. This collaboration would be beneficial for MuG as it would provide the VRE with an opportunity to position itself within the EOSC e-infrastructures at an early stage, while making a strong case to demonstrate the relevance and usefulness of the **EOSC Services** and the capacity to compete with strongly-funded international initiatives such as NIH's 4DNucleome.org in the USA.

9.4.2 Market and technology watch

WP6 and WP3 carry out technology watch in the respective areas of the tools they develop. Complementary tools that might need to be integrated in the VRE are discussed with pilot projects and VRE developers. In addition, technical progress (e.g. advances in technology changing input data formats), which is one of the drivers for the development of such a portal as MuG, could also mean that developed solutions become obsolete (e.g. loss of efficiency of data models) and need to be addressed (WP4).

As an example of this need for continuous assessment of the state of the art (essential in this fast evolving field) and technology watch, technical WPs and Pilot projects collaborated in the generation comparators for 3C analysis tools and for Image analysis tools (http://www.multiscalegenomics.eu/MuGVRE/#) which will be periodically updated.

On the other hand, the consortium needs to keep an eye on parallel initiatives being developed worldwide (e.g. **4DNucleome.org**). MuG PI Marc Martí-Renom is among the leaders of the **4DNucleome.eu** initiative and co-PI of the ERC-Synergy Grant 4DGenome, having a key role in technology watch. On the other hand, technical WPs work hand in hand with the coordinator to position the project as an e-infrastructure that is taken into account in the roadmap towards the realization of the **European Open Science Cloud (EOSC)**, making sure that project developments are done according to all guidelines and recommendations.

As described in the first Report of the High Level Expert group on the EOSC[8], the EOSC (Figure 8) is not conceived as a major localized and centrally governed initiative, but it should rather build on reached agreements on minimal standards and be technically conceived as an Internet of FAIR data and Services. Only under the umbrella of the EOSC as a Europe-wide initiative shall it be possible for the MuG VRE to compete with other strongly-funded international initiatives and take the lead in the 3D/4D genomics field.

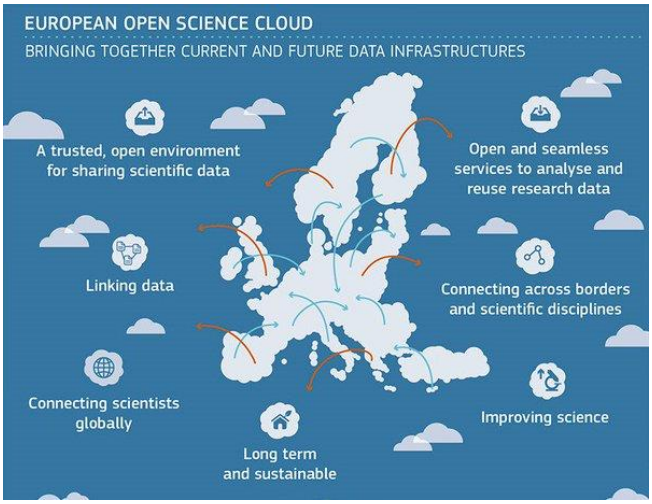


Figure 8: Description of the European Open Science Cloud, supporting transition to more effective Open Science and Open Innovation in the Digital Single Market**.**

9.5 Community uptake

As a user-driven infrastructure, the existence of a critical mass that justifies the maintenance of the service and the provision of further resources to be committed is going to be essential to secure long-term sustainability in the European e-infrastructure landscape.

A proper training and dissemination strategy and the implementation of corrective actions along the way (D2.4) is essential to ensure the relevant stakeholders are reached. The key actions undertaken so far to enhance community engagement are:

- **Dissemination strategy**: MuG partners span different levels of expertise and have a huge engagement potential among end users. Our pilot projects (lead users) participate in the VRE definition and following initial testing of VRE services with their own research results, will have a key role in engaging the wider capacity. CNAG-CRG, IGH-CNRS have a huge potential to reach the target community of experimentalists generating data. IRB, BSC and UNOT, have a key role in reaching out to the users working on the atomistic level resolutions.
- A **position paper** is under preparation about standardization in relation to data. This will be the first step for MuG towards being recognized as a reference and increasing the user community that uses MuG tools.
- **Training** becomes instrumental to reach VRE end-users and engage audiences in the VRE testing from the early stages of development. Together with lead user inputs, feedback from sets of external users will contribute to shape the VRE according to users' needs.

[8] Realising the European Open Science Cloud. First Report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. European Commission, Directorate General For Research and Innovation (2016) doi:10.2777/940154.

- **Registration of community interest** in the website (MuG community, Forum, Newsletter, social media), in addition to monitoring of website metrics.

10 CONCLUSIONS AND FUTURE WORK

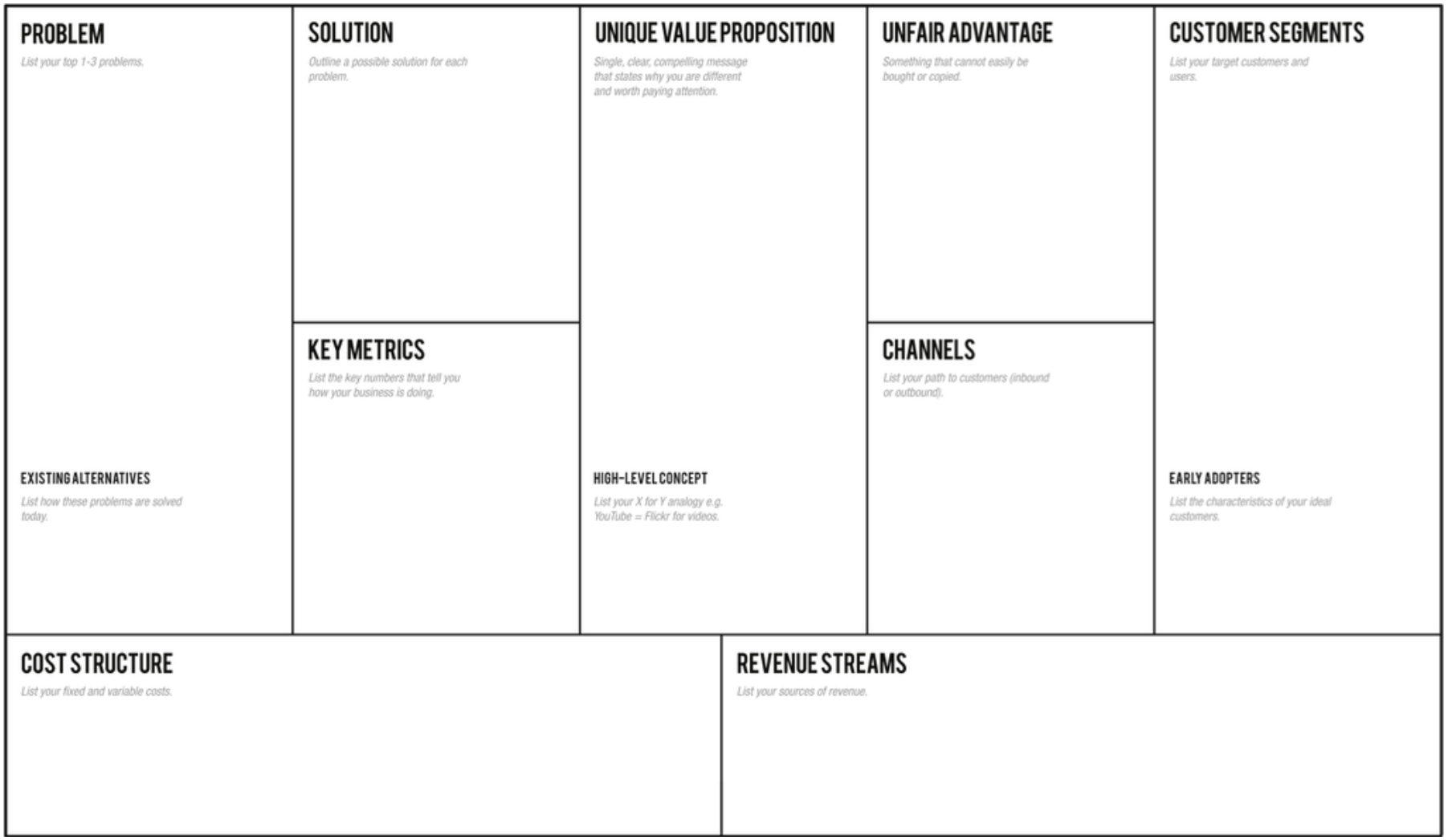
The foundations of the sustainability and exploitation plan for the MuG VRE have been set, identifying the key points of intervention for the next months in order to pave the way for exploitation. Measures for sustainability, including positioning the VRE within the European e-infrastructure landscape have been outlined and possible routes to commercial exploitation have also been explored.

Key actions that need to be addressed in order to establish the definitive path to long-term sustainability and exploitation include:

- Enhance training activities in collaboration with other European projects to increase community engagement in academia (M18).
- Foster links with industry and further develop the business model based on commercial exploitation. To this end, the subcontracting of external consultancy services is already being formalized (M24).
- Define terms and formalize agreements with different stakeholders:
    - Agreement with third parties for incorporation of tools (M24)
    - Formalize collaboration agreements with e-infrastructures and international genomics initiatives (starting M24).
    - Agreements with users for data reuse and verification (M36).

A revised exploitation plan will be submitted in M36 (D2.7).

11 ANNEX I: Lean canvas business model



MuG strategy discussion board based on Lean Canvas Business model