Prioritised set of selected workflows, leading to the development of solutions based on composable EBRAINS services (OP6.31)

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	Prioritised set of selected workflows, leading to the development of solutions based on composable EBRAINS services
Document Filename:	HBP_SGA3_WP6_T6.14_OP6.31_v1.0.docx (1)
Output Number:	SGA3 OP6.31
Dissemination Level:	PU = Public(2)
Due Date:	SGA3 M6 / 30 Sep 2020
Achieved Date:	SGA3 M12 / 31 March 2020
Author(s):	Wouter KLIJN, JUELICH (P20)
Compiled by:	Wouter KLIJN, JUELICH (P20)
Contributor(s):	
WP QC Review:	Maren FRINGS , JUELICH (P20)
WP Leader / Deputy Leader Sign Off:	Thomas LIPPERT, JUELICH (P20)
PCO QC Review:	firstname FAMILYNAME, SHORT PARTNER NAME (P#) *(to be updated by PCO)*
Abstract:	This output details the current set of workflows driving the actions of the Scientific Liaison Unit (SLU) leading to the development composable integrated workflows based on EBRAINS services. Additionally, the document introduces the SLU, details the relationship of the SLU with other coordination tasks in EBRAINS, details the challenges encountered in formalizing science or engineering cases and finally provides the methods and procedures followed to address these challenges.
Keywords:	EBRAINS, Showcase, Scientific Liaison, Workflow, Requirements Analysis
Target Users/Readers:	computational neuroscience community, HBP-consortium members, funders, general public, platform users, policymakers, researchers, scientific community

Table of Contents

Table of Tables

Table of Figures

1. Introduction

This output report summarises the current set of workflows that drive the actions of the Scientific Liaison Unit (SLU) and lead to the development of composable integrated workflows based on EBRAINS services. Chapter 1 introduces the SLU's range of tasks, further describes the SLU's relationship with other EBRAINS coordination tasks, explains the challenges that have encountered in formalising scientific and/or technical cases, and finally describes the methods and procedures that have been used to overcome these challenges. A prioritised set of selected workflows can be found in chapter 2. Chapter 3 contains the template that is used to streamline the information gathering process.

The EBRAINS Scientific Liaison Unit (SLU) will bring together the Science and Infrastructure Directors' visions (top-down), the community's needs and the day-to-day technical activities (bottom-up) conducted within the EBRAINS development teams. It will act at the interfaces between different Work Packages (WP's) work plans, and those between Science and Infrastructure activities. It will structure and trigger cross-Service Category (SC) developments foreseen around self-contained projects, to enforce comprehensive solutions that respect the principle of service composability, i.e., to systematically bring the EBRAINS services together in a range of typical and specific contexts. These solutions will help the EBRAINS SLU to actively liaise with the scientific community, to encourage the acceptance of its services by the community, and to strongly promote the use of EBRAINS by the external community.

On the one hand, the selection and prioritisation of projects to create the solutions will reflect the HBP-wide scientific strategies defined by the scientific leadership of HBP/EBRAINS. On the other hand, projects to create the solutions must be selected based on their prospects to explicitly support major needs expressed by the scientific community, and to evolve into a solution illustrating the merits of the combined EBRAINS services. The selected projects must also provide a collaborative structure to frame software development activities across SC's. Such coordinated, concerted efforts are necessary to achieve decisive (and substantial) advances in the development of a cohesive EBRAINS. The selected projects will prioritise, in the early stages of SGA3, functional connections of key services with the aim of supporting the development of workflows spanning the breadth of neural data, models, and functions, highlighting the holistic perspective promoted by the Flagship.

The function served by the EBRAINS SLU is that of a key enabler, which places EBRAINS at the service of the scientific research community, and anchors technical developments within the community's needs, through the combined understanding of both the scientific nature of the problems to be addressed and the technical capacities or constraints (to overcome) of the Infrastructure, It will empower the scientific community to take ownership of the services offered, fulfilling EBRAINS' ambition of digitising neuroscience.

1.1 The SLU in the HBP: Relationship to other HBP activities

Establishing EBRAINS as a new digital research infrastructure, created by the EU-funded Human Brain Project (HBP) involves a large number of interdisciplinary research groups and scientific fields with a large number of researchers working in close collaboration towards novel neuroscientific- and AI infrastructure Services.

The Human Brain Project consists of ten Work Packages (WP) with WP1-6 being relevant for the work on the current document. To coordinate the shared work and efforts, a number of different tasks and units exist, each with their defined roles. Sections 1.1.1 - 1.1.5 details the different tasks and roles of: Service Categories (SC's), High-Level Support Team (HLST), EBRAINS Technical Coordination (TC), Scientific Integration (SI), and Calls for Expression of Interest (CEoIs). Figure 1 shows the relationship between the different coordination tasks.

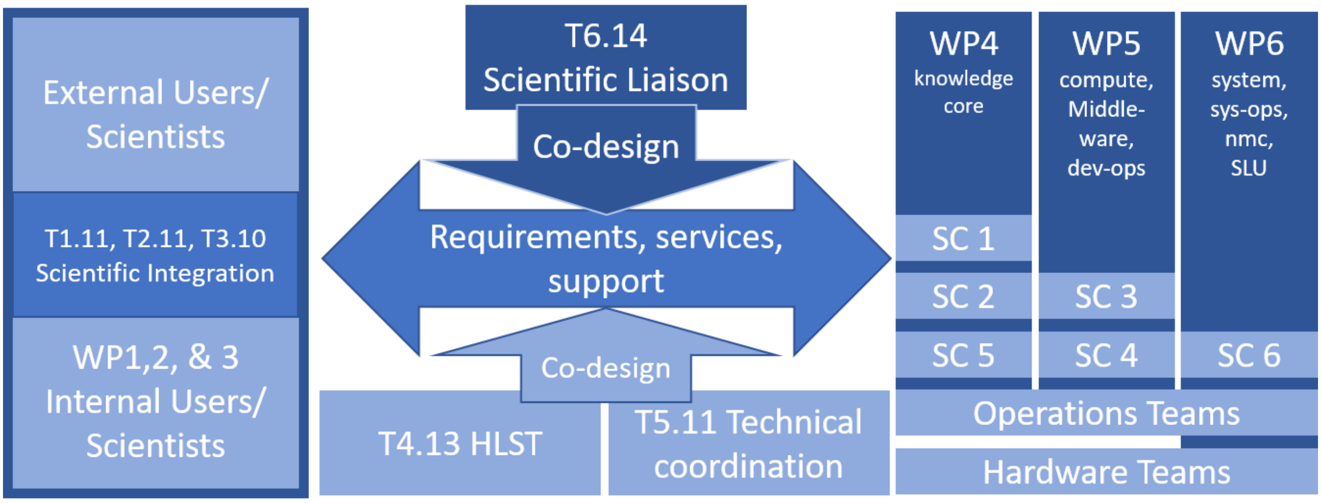

Figure 1: The relationship between the different coordination tasks

The relationship between the different coordination tasks can be seen in Figure 1. The science WP 1, 2 & 3 users and scientists will be the first source of science (show) cases to be formalized, which will be done in close collaboration with the Scientific Integration tasks embedded in the work packages. In the second phase these cases will shift to come primarily from external users and scientists. The collection of requirements and assessment of needed services and support will be done in a co-design process with support of both HLST and Technical coordination. The produce formalized workflow and accompanying requirements will be provided to the Service Categories (SC) in the infrastructure work packages 4-6.

1.1.1 Service Categories (SC 1-6)

Over the last of four funding phases, the project structure of the HBP has changed significantly. The new structure reflects a clear prioritisation towards being able to deliver functionalities more reliably, at greater scale and to a broader user base. As a result, the infrastructure Work Packages (WP 4-6) are primarily organised around the type of operation performed, referred to as Service categories, in effect, seeking efficiencies by grouping activities that are common to different scientific work streams.

1.1.2 High-Level Support Team HLST (T4.14)

The High-Level Support Team (HLST) was already, in a previous project phase, established to support all users in the best possible way and thus mitigate technical and cultural gaps. This HLST support helps scientists to achieve their research goals, facilitating their use of the most appropriate tools and services provided by the HBP. The scientists also help to enhance developed tools through the feedback they provide when using them. Regarding the transition to EBRAINS HLST will maximise EBRAINS user productivity, facilitating access to unique offerings through competent support, documentation, and training. The task will thus attract scientists to EBRAINS and retain them as long-term users, helping to build a broad, strong and diverse EBRAINS user community. HLST will work closely with the teams providing EBRAINS tools and services to deliver coherent expert user support, user-directed documentation and tutorial material. Providing support at five levels: frontline support, in-depth support for tools and services, operations support**, coordinating support for scientific use-case development** and integration, and support for model curation. In-depth and operations support will be provided in close collaboration with the teams that provide EBRAINS tools and services, while more complex use-case-driven development efforts will require funding through EBRAINS Infrastructure Vouchers.

1.1.3 Technical coordination (T5.11)

Technical Coordination will coordinate the planning and design of the EBRAINS infrastructure components, maximizing the efficiency of the work teams, the quality of the outcomes, and the satisfaction of their users. Overall, the efforts are organized into two subtasks: (a) the implementation planning and coordination of involved stakeholders, and (b) the establishment of the user requirements and architecture of EBRAINS.

1.1.4 Scientific Integration (T1.11, T2.11, T3.10)

Activities pursued within T1.11 include scientific and technical coordination across collaborating Tasks within WP1, between WP1 and other scientific work packages, as well as scientific activities in WP1 and technical developments in EBRAINS"

The objectives of T2.11 are to ensure the successful interaction of all scientists in WP2 with EBRAINS, providing a bridge between the platforms and the WP, and providing technical support, coordination and scientific integration.

Activities pursued within T3.10 include scientific and technical coordination efforts, across collaborating Tasks within the Work Package, between this Work Package and complementary activities in other scientific Work Packages and, crucially, mediating the relation between scientific activities in the Work Package and technical developments in the infrastructure"

1.1.5 Calls for Expression of Interest (CEoIs)

The Aim of Calls for Expression of interest is to strengthen the Flagship by integrating new Partners that provide required, missing expertise. CEoIS target institutions outside the current HBP Consortium or departments/groups within existing HBP Partners not currently involved in the HBP.

1.2 Challenge and solution

EBRAINS is supported by a large number of research groups working together towards novel neuroscience and AI infrastructure. The needs and expectations from the EBRAINS infrastructure are thus diverse as illustrated in Figure 2.

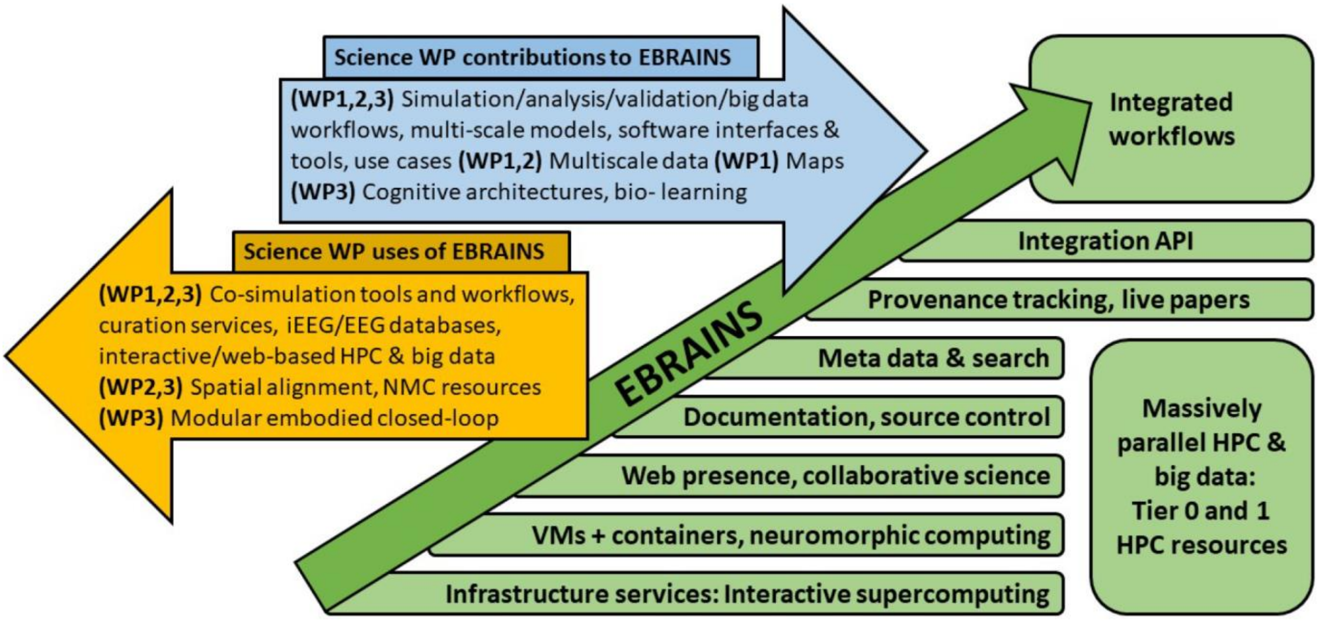

Figure 2: WP1, 2 and 3 contributions to EBRAINS and services needed from EBRAINS

Determination of these needs and expectations, the requirements, is complicated by a number of challenges:

1. Mismatch in the understanding of what concepts might entail. e.g., "workflow" is widely used in both, science and computer science but mean completely different things.

2. Lack of insight into what are the available components in EBRAINS to solve tasks in the science cases.
3. Limited insight in quantifiable technical needs as typically available in 'freeform' scientific descriptions of cases.

1.2.1 Template and process to formalize science cases into integrated EBRAINS workflows

To guide the collection of the information about the science and use cases, as well as the collection of quantifiable technical needs, a template has been created. This template is currently in version 3.0 with previous versions used in the use case specification of previous phases of the HBP and the requirements assessment done in the ICEI project (fenix-ri.eu). The template was developed in collaboration with technical and scientific experts of the SimLab Neuroscience at JUELICH-JSC and the HBP. The template with accompanying explanations can be found in section 4.

The templates will not be sent empty to potential science owners: Available information from known previous use-case formalization efforts will be included in the document. When possible a first diagram is added with a breakdown of the use case in smaller components, mapping on EBRAINS components. The pre-filled template will be sent to the owner of a use case, including a cover letter and a detailed explanation of the requested effort. In addition, these efforts will be explained further in face-to-face meetings or via Video conferencing meetings.

The specific task for the science owner is to complete the scientific part and, where possible, to complete the template with technical resource information. To expedite the answering of this technical part where possible the science case owner will also be matched with a technical expert to help guide the completion of the template. These experts will be selected from the scientific integration tasks of from the High Level Support Teams.

The different roles in the formalization process are illustrated in Figure 3.

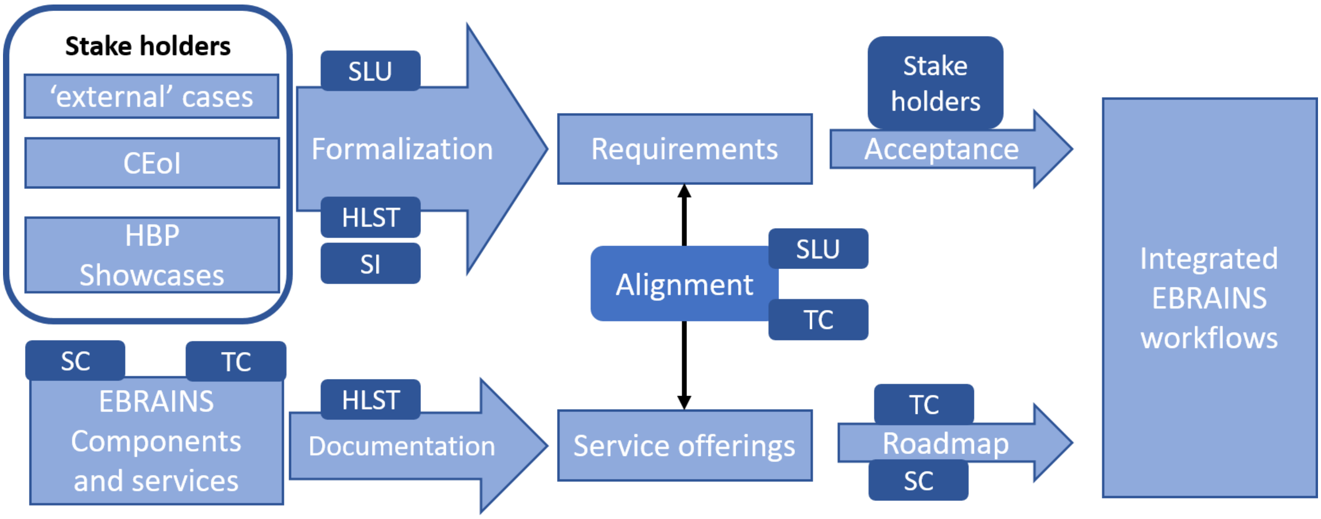

Figure 3: Steps and roles in the formalization process of the SLU

2. Prioritized set of selected workflows

There are different sources for the science cases managed by the SLU. Below a short overview of these sources followed by the prioritized set of workflows as current known to the SLU.

1) Work packages 1-3 have defined a set of showcases demonstrating progress in addressing the SGA3 scientific challenges and the empowerment EBRAINS brings.

2) Calls of expression of Interest brought in to strengthen the Flagship by integrating new Partners that provide required, missing expertise (further detailed in 1.1.6)

3) The community builders focus on building a lasting, large, diverse and highly collaborative EBRAINS Community with members who in a diversity of ways can gain from and contribute to projects and activities based on the infrastructure. Resulting in the generation of new EBRAINS users and collaborators with accompanying requirements.
4) WP8 address the intended impact of EBRAINS through communication and engagement efforts targeting scientific communities engaged in European brain research. Result in the generation of new EBRAINS users and collaborators with accompanying requirements.

2.1 Set of workflows

Due to potential confidentially issues the cases will not be detailed further in this document outside of their title.

Priority 1:

1) Showcase 1 (WP1): Degeneracy in neuroscience - when is Big Data big enough?
2) Showcase 2 (WP1): Improving epilepsy surgery with the Virtual BigBrain
3) Showcase 3 (WP2): Brain Complexity and Consciousness
4) Showcase 4 (WP2): Object Perception and Memory
5) Showcase 5 (WP3): Dextrous manipulation - how the brain coordinates hand movements

Priority 2:

1) Calls for Expression of Interest
2) External users of the EBRAINS infrastructure as generated by Outreach (WP8), the community building tasks in the science (WP1-3) and infrastructure (WP4-6) work packages.
3) Existing workflows with high potential for reuse e.g.
    a. Live papers
    b. Brain signal analysis workflow
    c. Full service MOOGs for EBRAINS components and science cases

3. Template

The following section contains the content of version 3.0 of the “Workflow Description and Specification” template used in the SLU to formalize workflows. The formatting has been adapted to match the rest of the document. The template in its original shape contains additional sections to list the content and formal information like version and change information. The reference in the text are towards the location in the original template document.

3.1 Introduction

This use case description and specification document provides a tool for developers and scientist to collaboratively transform a free form description of a science use-case into technical specifications. Specifications that guide the implementation of workflows fulfilling the science use-case. This document should help the science case in a number of ways: its structured methodology will help to find the essential parts, and it will assist in separation of the ***must*** haves and ***nice*** to haves [1]. The specifications should result in a standalone document that can be given to new partners of the project as introduction into the science and technical details of the project. On a more abstract level this document could be seen a contract formalizing of the expectations of both engineer and the scientist.

An important guideline when creating a use-case analysis document is the separation of user requirements and technical details. A user is ultimately only interested in the functionality of a software / hardware product and not in the underlying technical details of the implementation. Separating these concerns is a non-trivial matter: This document will therefore typically be written in an iterative manner, with the document bouncing from scientist to developer getting more detailed on each iteration. It will also be living document: Details of the project can and will change over time; Components might be hard to implement and trade-offs might be made depending on availability of manpower. The amount of work needed for this document might appears large, it is work that, for a typical software/science project, should be performed anyways.

The different elements/chapters in the template should be kept in order and contain the content described. This will allow comparison of use-cases and allow identification of shared / overlapping functionality. This document and accompanying PowerPoint introduce a set of visual components that can be used to describe the use-cases and systems (Section 1.2). The symbols should cover the majority of systems encountered, but if the need arises new element can be introduced. Do keep in mind that this will complicate comparison of the diagrams created. The main goal for collecting the information is to foster the reuse of efforts and components. Although the introductory chapters can be removed, it will limit the use as an introduction for new project partners.

In the next sections the goal of the individual parts of the template will be introduced. The first section (1.1) details the use-case description, it should provide the scientific reasoning behind the case. Section 1.2 explains the set of visual components that can be used to create the model diagrams. In section 1.3 we provide the typical data point that can be used to characterize the different components in more technical detail. In section 1.4 we explain list of potential infra structure requirements specific questions. High-level needs and services that can be cross-checked with the node characterizations. Section 1.5 gives a growing set of potential additional diagrams to be created in the process of formalizing your case. These diagrams will be asked to be created depending on actual need for this level of formalization.

Section 2 is the actual template, it contains just the titles and list of infrastructure questions. Other components can be copied from the introduction chapter 1. If you add multiple diagrams/systems it is best to copy the template multiple times, or, use different documents. This will improve coherence in the descriptions.

3.2 Use-case Description

The workflow description is a high-level description of the science flow of the use-case. It is typically written by the scientist and provides the reasons why to build or use a software or hardware system. Topics that might be encountered in this section are: How new (or better, bigger, faster) science is possible with this software. Problems and challenges encountered in current software.

Typically, the workflow is broken down in steps with partial goals for each step. It is advisable to keep implementation and technical details out of this section. Implementation details are not part of the description: An example of such and ***implementation detail*** would be: “The software must be fast, to allow fast turnover of experiments. ***We have to use GPUs***”. A complete separation of concerns is hard to arrive at. It is one of the more complicated exercises in system design. Having a starting point is more important that being completely correct. This is one of examples where the dialog with technical experts will help to arrive at a correct description.

An example of a science (and not technology) centric description:

“As a researcher I want to be able to perform a large scale computational experiment. This experiment cannot be performed on my local cluster due the size of parameter space I want to explore. The analysis of the results will need to be performed in my local institute due to A and B. The access of the results should be structured based on X and Y. “

Two widely different technical solutions would support this case:

1. Analysis of results on a virtual machine with data staying in a central location. Results selectable via a database, accessed via a web interface.
2. Transport of results to the local cluster with processing on the local machines with the data stored in clearly labelled directories.

Which of these solutions is implemented can now be made on available resources, software limitations, etc.

3.3 Annotated Use Case Diagrams

An annotated use case diagram is a relatively freeform graphical depiction of the textual description as detailed in section 1.1. We would suggest to use the diagram components as shown in Figure 4. As this will allow easy comparison between different use-case descriptions. The flowcharts in this document follow the practices as described in [2], [3].

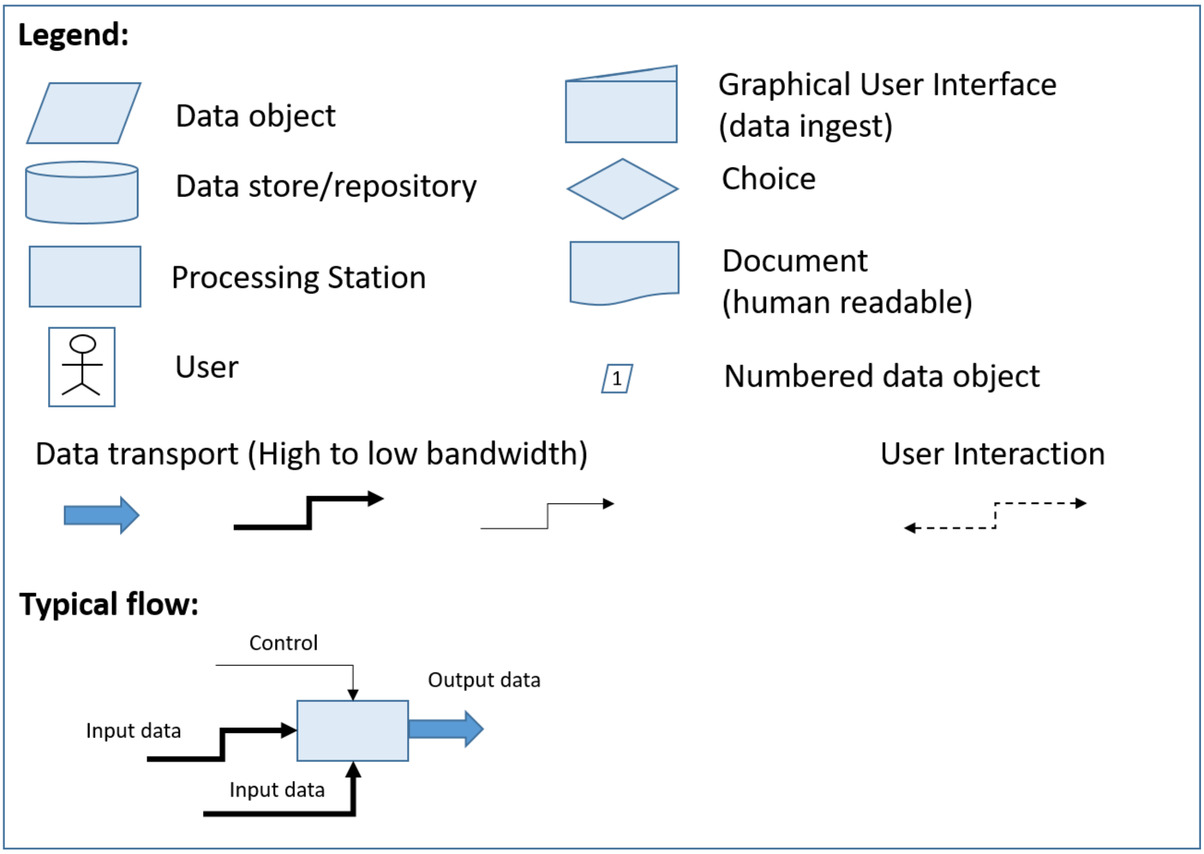

Figure 4: Overview of suggested symbols for a use-case diagram.

The symbols in figure 4 are based on [2], [3]. The symbol for GUI is a combination of processing station and data object. A suggested typical data and information flow is shown. Additionally, a simple bandwidth ranges us depicted. An editable version of the above diagram (a PowerPoint presentation) will accompany the current document.

To prevent cluttering of complicated workflow we suggest the following:

- Make use of specialize symbols to allow for a visual distinguishing of salient features (GUI would be an example).
- Use only a small pictogram for data objects annotated with a number.
- Use the suggested locations for the connectors: Control at the top. Inputs from the left or bottom. Outputs leave on the right side.

To reiterate: these are suggestions, the diagrams are in principle freeform and not all symbols might be used in your specific use-case.

3.4 Node Characterization

In this section a characterization of each component as depicted in the annotated use-case diagram. This is done in a table format with typical information points listed. The entries are typically split in different sets: The ***base*** information set without which an informed discussion might be complicated. The description is typically at a user / functional level. Secondly, **technical specifications** of the requirements. The Use-case is not yet solved thus this information will by necessity be added incrementally and optionally by a domain specialist. The third information set is regarding ***current solutions*** that one is aware of.

Not all information might be available. Fill in what is known at this stage. Having a start point for a dialog is more important that having perfect information, especially in the beginning stages

3.4.1 Data objects

Table 1: Data object

Data object: ***number in diagram***, name	
Base Information	General description of what data is stored. Potential addition information: Formats, Metadata, Database requirements
Technical specifications	Please select: • Transient (Temporary): Data discarded on simulation completion or when later processing steps are concluded. • Short-term (Campaign): Data used throughout the execution of the scientific workflow. • Permanent (Forever): Data outliving the machine used to generate it.
Current solution (optional)	Name
	URL to additional information
	Limitations

3.4.2 Data Transport

Table 2: Data transport

Data transport: ***number in diagram***, Name	
Base information	General description of what data is transported Potential addition information: Data access patterns (request rate, transfer sizes)
Technical specifications	(est.) Maximum required bandwidth
	(est.) Average required bandwidth
	Interface requirements

	Additional information
Current solution (optional)	Name
	URL to additional information
	Limitation

3.4.3 Data ingest / GUI

Table 3: Data ingest / GUI

Data ingest: ***number in diagram***, Name	
Base information	Description of input data source Potential addition information: Description of data introduction, e.g. upload? scanner characteristics? simulation characteristics?
Technical specifications	Characteristics of data: formats, loads, bandwidths, latencies, transports
	Additional information
Current solution (optional)	Name
	URL to additional information
	Limitation

3.4.4 Data repository

Table 4: Data repository

Data Repository: ***number in diagram***, Name	
Base Information	Classification of the data objects (see below)
	Access control requirements
	Access requirements
	Data availability requirements
Technical specifications	Maximum and average capacity requirements
	In case of repository for permanent data objects, i.e. repositories where data is accumulated, provide maximum capacity requirement as function over time.
	(est.) In terms of size & file number
	Additional information

Current solution	Name	
	URL to additional information	
	Limitation	

3.4.5 Processing stations

Table 5: processing station

Processing station: Name	
Base Information	General description of data processing
Technical specifications	Hardware architecture requirements
	(est.) computational needs
	Additional information
Current solution	Name
	URL to additional information
	Limitation

3.5 FENIX Infrastructure requirements

This section of the template will map from the FENIX HPC infrastructure to the use-case. Per envisioned infrastructure service we ask specific questions how this service might be used for your use-case. There will be overlap with information provided through annotated use case model diagrams. This duplication is **intended** it will allow consistency checks. This avoids the need of fixing the mapping between the case and specific infrastructure services at a later stage.

Information as provided on (https://fenix-ri.eu/infrastructure/services)

“Below is the list of available Fenix HPC and data infrastructure services for research communities and users. Click on each title for more details on the specific services.

Interactive Computing Services

https://fenix-ri.eu/infrastructure/services/interactive-computing-services

Quick access to single compute servers to analyse and visualise data interactively, or to connect to running simulations, which are using the scalable compute services.

Scalable Computing Services

https://fenix-ri.eu/infrastructure/services/scalable-computing-services

Massively parallel HPC systems that are suitable for highly parallel brain simulations or for high-throughput data analysis tasks.

Virtual Machine Services

https://fenix-ri.eu/infrastructure/services/virtual-machine-services

Service for deploying virtual machines (VMs) in a stable and controlled environment that is, for example, suitable for deploying platform services like the HBP Collaboratory, image services or neuromorphic computing front-end services.

Active Data Repositories

https://fenix-ri.eu/infrastructure/services/active-data-repositories

Site-local data repositories close to computational and/or visualization resources that are used for storing temporary replicas of data sets. In the near future they will typically be realised using parallel file systems.

Archival Data Repositories

https://fenix-ri.eu/infrastructure/services/archival-data-repositories

Federated data storage, optimized for capacity, reliability and availability that is used for long-term storage of large data sets which cannot be easily regenerated. These data stores allow the sharing of data with other researchers inside and outside of HBP"

Table 6: FENIX infrastructure requirements

Infrastructure service	
Interactive Computing Services	Which parts of the workflow require such services? What is the expected typical duration of interactive sessions? What software stacks need to be available? Is it possible to define memory capacity requirements?
(Elastic) Scalable Computing Services	Which parts of the workflow require such services?
Virtual Machine Services	Which parts of the workflow require such services?
Active Data Repositories	Which parts of the workflow require such services?
Archival Data Repositories	Which parts of the workflow require such services?

3.6 Additional diagrams to be completed when needed

3.6.1 Visualizing timescales in computation sciences

The HPB has as it goal the creation of an e-science infrastructure to support computational neuroscience in the broadest sense. Due to the inclusive, large and broad scope of this endeavor and multi-scale nature of the human brain the science, methods and computation solutions are equally varied:

- Measuring the current on a synapse, to tracking patient performance on mental tasks over time.
- Simulating ionic transport in a membrane junction, to exploring the neuroscientific foundation of lifelong learning.
- Deploying an analysis script on a supercomputer to laboriously writing a paper while collaborating with your colleagues.

The juxtaposition of these activities shows the scale of the challenges the HBP wants to provide a solution to.

In this section we acknowledge the importance of the science time scales involved in the HBP. Different time scales have different challenges and call for different solutions. The Figure 5 introduces the stacked time scale diagram central created help you formalize the different timescales in your science case

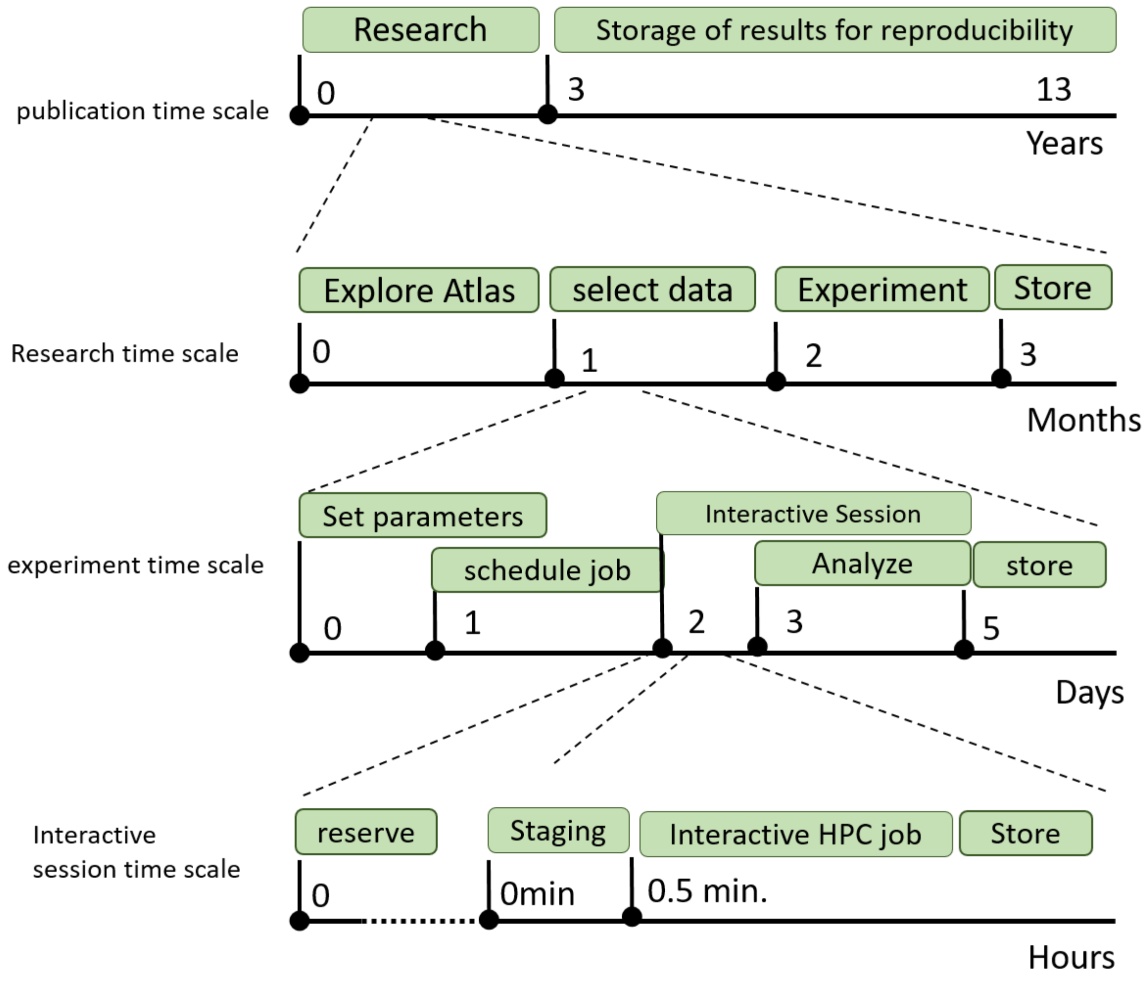

Figure 5: Stacked timescales.

The top time scale describes the longest timescale, here measured in years. At this this scale we expect to see creation of projects, and related project management actions. At this level the long term storage of result for reproducibility (in some instance up to 10 years) is found. On lower shorter time scale smaller, measured here in two months we find the research time scale. Where actions like exploration of available of HBP Atlases might be placed, selection of data, but also the storage of results. On scale smaller, looking at individual experiments more technical action and moments occur: setting of parameter, scheduling a job. At the smallest time scale we zoom in to the details of an interactive HPC session where we want to stage our jobs and interact with the HPC systems.

This breakdown of the complete science life cycle of the HBP, that this document aims to support, gives equal weight to challenges and solution at the different timescales. It relates the science phases at widely different conceptual scale. Additionally, allows the ordering of actions and it gives a natural grouping of actions and concepts.

The diagram as presented in Figure 5break down a complete end-to-end workflow in four different time scales, it is important to note that an instantiation for a specific use case might have a different number of scales or different units on its axis, or might not have information available on a specific time scale.

3.6.1.1 Publication time scale

The largest, publication time scale, spans multiple years. Items to be found at this levels could be creation of user-groups, sharing of research results between users and the access of provenance data for reproduction of data. The current incarnation of the collab performs a number of the functions of this time scale. This is one of the levels where outside scientist might typically first encounter functionality as provided by the HBP. The information at this level is typically summary and expected to be public in nature. At this level educational material and public relation material have their natural home.

The functionality of this layer is best compared to existing science portals, or even publication and data storage sites. The technological solutions should at the levels of websites, wikis, git repositories and data bases. Most activities in the HBP would share this level and thus the solutions provided.

3.6.1.2 The research time scale

At the timescale of months, we find the, possible incorrectly named, research time scale. Items at this levels would include user creation and management. It would need private data sharing inside a research groups. This is a collaborative time scale with data typically accessible for multiple users: access to data as produced in different experiments, exploration of results and reading of data for publication. The current incarnation of the collab performs a number of the functions at this level. Due to the private nature of this work user authentication is needed, although anonymous sessions should be allowed allowing potential users to interactive with the functionality of the system.

The actions on this level are becoming more specialized towards the vertical platforms as currently envisioned. The requirements of Atlas exploration are completely different to the needs of selecting algorithms for steering a robot. A number technical solutions at this level are expected to be shared. Specifically, the user authentication, data access and data transport solutions. Additionally, sharing within a research group is best supported with collab like functionality: wikis and git repositories.

3.6.1.3 The experiment time scale

At the time scale of days, we find the experiment time scale. Users at this level would typically enter via the higher time scale and are expected to be logged in. In contrast to the research time scale activities are single user centric, although access to shared data resources would be natural. The functionality at this level will be very task specific and a number of the vertical solutions have the majority of their functionality level: setting of application specific settings, creation of HPC jobs and interactions with GUI's like the Neurorobotics platform. It is at this level that a researcher is expected to spend the majority of this time. Because a majority of the functionality of the HBP will be located at this level anonymous access for potential users should be supported.

The actions at this level are not necessarily application specific: Jupyter notebook editing would have its logical home here. Analytical methods like machine learning or simple numpy operations should be supported. It is at this level that good computation workflows support is essential. Typically expected are creation of HPC job scripts, data transport to and from central.

A number of the technical solutions is currently performed by the collab but mature integrated workflow management support, from mature HPC workflow systems like Pegasus or Kepler would greatly increase the usability at this level. The usage of existing solutions allows capitalization on existing functionality like provenance tracking and HPC and workflow monitoring.

3.6.1.4 HPC Session time scale

The HBP has an aim to enable interactive science. Interactive access to real-time resources: HPC simulation, neuromorphic hardware, the neuro robotic platform and interactive exploration of large scale datasets in the HPC brain atlases. The NRP has an existing toolchain allowing interactive control of robotic experiments. The Brain atlas provides interactive access to petabytes of integrated data and the multiple simulators developed are increasingly integrated and will in the future allow scientist to adapt model parameters on the fly. Al these efforts are unique in the world but also provide unique challenges: Monitoring and provenance tracking of these multi components workflows where not a single component can fail is an unresolved challenge.

Partial technical solutions can be found in Pegasus but interactive computing is identified as an unsolved problem where the HBP might have a unique role.

3.7 References

[1] *MoSCoW Analysis (6.1.5.2)*. International Institute of Business Analysis, 2009.

[2] "Flowchart Symbols Meaning | Standard Flowchart symbol images and usage." [Online]. Available: https://creately.com/diagram-type/objects/flowcharts. [Accessed: 17-Aug-2017].

[3] "Flowchart Symbols and Notation | Lucidchart." [Online]. Available: https://www.lucidchart.com/pages/flowchart-symbols-meaning-explained. [Accessed: 17-Aug-2017].

[4] "UseCaseDescription_and_Specification_v1." .