Project Number:	720270	Project Title:	Human Brain Project

Document Title:	IT Architecture of the HBP Integrated System of Platforms
Document Filename[(1)]:	SP5 D5.6.2_FINAL_Resubmission
Deliverable Number:	SGA1 D5.6.2
Deliverable Type:	Platform Specification
Work Package(s):	WP 5.5, WP5.6, WP7.5, WP11.2, WP11.3
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA1 Month 6 / 30 Sep 2016
Actual Delivery Date:	SGA1 Month 7 / 12 Oct 2016, Resubmission 31 May 2017

Authors:	Jeffrey MULLER (EPFL-P1), Thomas LIPPERT (JUELICH-P2, DPIT), Thomas SCHULTHESS (ETHZ-P18), Jan BJAALIE (UIO-P81)
Compiling Editor(s):	EPFL (P1): Jeffrey MULLER
Contributors:	EPFL (P1): Selim ARSEVER, Jean-Denis COURCOL, Mike GEVAERT, Daniel PEPPICELLI, Luis RIQUALME, Felix SCHÜRMANN, Werner VAN GEIT, Stefano ZANINETTA, Martin TELEFONT, Guy WILLIS, Marc-Oliver GEWALTIG JUELICH(20): Thomas LIPPERT, Anna LÜHRS, Timo DICKSCHEID ETHZ(18): Thomas SCHULTHESS, Colin MCMURTRIE UMAN(63): David LESTER UIO(81): Jan BJAALIE, Dmitri DARINE
Reviewers:	EPFL (P1): Guy WILLIS, Colin MCKINNON

Abstract:	
Keywords:	

Document Status

Version	Date	Status	Comments
1.0	11.10.2016	Draft	Split Draft, Jeff Muller
1.1	11.10.2016	Draft	Review, TL
1.2	11.10.2016	Draft	Revisions to address TL's latest review
1.3	11.10.2016	Draft	Updated author list, with improved document linkages
1.4	11.10.2016	Draft	Minor update
1.5	11.10.2016	Draft	Final content review

Table of Contents

List of Figures and Tables

Executive Summary

The Human Brain Project (HBP) is an EU Flagship deeply founded on ICT. Its Neuroscience Subprojects (SP1 to SP4) as well as its six Platforms (SP5 to SP10) are all rooted in Information and Communications Technology (ICT). The immense amount, complexity and heterogeneity of HBP internal and external data, the need for high-end simulation and for data analytics capabilities, and the HBP's paramount collaborative ambition demand the development and operation of highly reliable underlying IT-based technologies and services. In other words, a world-class IT Architecture for the HBP Integrated System of Platforms is required. The HBP IT Architecture is the prerequisite to make the HBP platforms accessible in a most effective manner to researchers from neuroscience, medicine, computer engineering and beyond. This is provided over the internet, via the HBP Collaboratory (HBP-COLL or COLL).

The Neuroinformatics Platform (HBP-NIP or NIP in short) is the central *Chef d'orchestre* which guarantees traceability and discoverability of the Platforms via the Collaboratory, manages seamless access to all varieties of curated data, as well as guaranteeing coherency of ontologies and present and future data types. The revised work plan of the NIP is provided in the HBP Deliverable SGA1 D5.6.1.

The NIP, on the one hand, will be deeply integrated with the COLL in SGA1 in order to ensure an effective ecosystem for data sharing and software, enabling application of that data to scientific problems. On the other hand, SP7's High-Performance Analytics and Computing (HPAC) Platform provides the computational resources, storage and networking necessary for both the primary archive platform for the HBP as well as the management, analysis, transport and storage capabilities along with the federation of very large data sets required for key data and modelling use cases. The HPAC Platform thus needs to be deeply integrated with the NIP and the COLL as well. The HPAC has formed a pilot initiative named FeDaPP (short for Federated Data aPplication Pilot), that explored the potential for federated infrastructure of HBP's supercomputer centers. The pilot will become FENIX (short for Federated ENgine for Information eXchange) at end of October 2016.

The purpose of D5.6.2 is to set out the IT Architecture of the HBP Integrated System of Platforms. Together with the new structure and infrastructure focus of SP5's NIP as laid out in D5.6.1, the close integration will bring together data, software, compute resources and their users, while maintaining sufficient simplicity to encourage use by less technically adept users. This integration will increase the ability of the COLL and NIP to track data provenance in order to provide pragmatic solutions for reproducible neuroscience. By facilitating the sharing of expertise, data and results, the COLL and NIP will generate network effects throughout the HBP community and help to maximise the research impact of the entire HBP.

1. Introduction

The Human Brain Project (HBP) is a 10-year research project, funded by the European Commission (EC), which aims to lay the foundations for a new approach to brain research. Neuroscience, medicine and information technology each have important roles to play in addressing this challenge, but their contributions so far have remained rather fragmented. The HBP aims to integrate of all of these inputs and catalyse a community effort.

It is a central tenet of the HBP strategy that a comprehensive understanding of the brain requires knowledge of structure and function across all levels of brain organisation; this understanding cannot be achieved at any one level alone. To achieve this understanding, interdisciplinary expertise joining neuroscience, computer science, physics and mathematics is key. A massive scientific collaboration is required to reconstruct such multi-level models. The social internet and open source software communities have shown that modern Information and Communications Technology (ICT) permits the massive collaborative efforts needed.

The central element in this endeavour is the HBP Neuroinformatics Platform (HBP-NIP or NIP in short). It is the central *Chef d'orchestre* which guarantees traceability and discoverability of the Platforms via the Collaboratory, manages seamless access to all varieties of curated data, as well as guaranteeing coherency of ontologies and present and future data types. The revised work plan of the NIP is provided in the HBP Deliverable SGA1 D5.6.1.

The HBP Collaboratory (or simply COLL) is key to enable the scientific community's access to the very complex data sets and hardware and software facilities of HBP Infrastructure Platforms and, in a broader sense, to make large-scale collaborations possible in neuroscience, see the detailed description in Deliverable SGA1 D11.3.1. This web-based collaborative scientific platform provides access to the HBP's research, community and administrative activities, as well as to its Infrastructure Platforms. A tool within the COLL of particular importance is a deeply integrated search developed in the Neuroinformatics Platform (NIP) (described in T5.5.2 in D5.6.2).

A second key element is the High-Performance Analytics and Computing (HPAC) Platform in SP7. It provides the computational resources, storage and networking necessary for both the primary archive platform for the HBP as well as the management, analysis, transport and storage capabilities along with the federation of very large data sets required for key data and modelling use cases. The HPAC has formed a pilot initiative named FeDaPP (short for FeDaPP (short for <u>Fe</u>derated <u>Da</u>ta a<u>P</u>plication <u>P</u>ilot), that fathomed the potential for federated infrastructure of HBP's supercomputer centers. The pilot will become FENIX (short for <u>F</u>ederated <u>EN</u>gine for <u>I</u>nformation e<u>X</u>change). FENIX described in the accompanying document named FeDaPP Architecture Report to be found as accompanying documents.

Both instruments need to be deeply integrated with the NIP, building the basis of the IT Architecture of the HBP Integrated System of Platforms. It is the purpose of this document to describe this architecture. It is intended for a technical and scientific readership. It describes the ways in which web-based platform components will serve key NIP use cases. It provides details of how the COLL and HPAC Platform will be integrated in the SGA1.

This document is intended to answer review feedback by providing an "… 'overall' design of the services available, the software components that implement the services, and the IT platforms that serve to actually execute and store the data" while make it clear how this will "be presented to the user". In addition, this document serves to clarify "the role of SP7 in providing the project-wide IT infrastructure and the data management, at all levels".

2. Architecture

2.1 Overview

The diagram below shows an overview of relationships between the components produced by the various platforms, the Collaboratory (COLL) and the architecture proposed by FeDaPP (which will continue as FENIX).

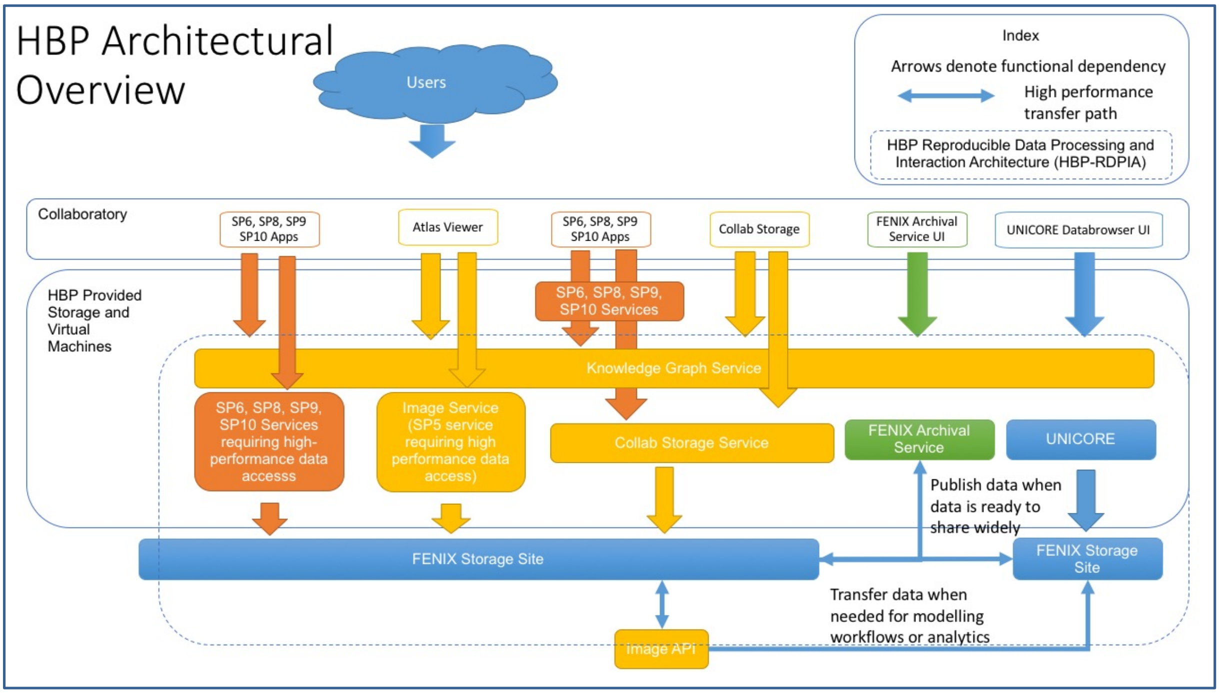


Figure 1: HBP Architecture Overview

2.2 Collaboratory Architecture Overview

- Service-Oriented Architecture – the COLL provides standard APIs to support a Service-Oriented Architecture ecosystem. These are typically versioned REST APIs to support COLL integration and the handling of common Authentication, Authorisation, metadata and data operations.
- Components – The COLL can be extended with the following components
  - Apps - web GUIs for certain services integrated into the COLL. In most cases these components will also use COLL services.
  - Services - web services, network file systems, SSH, source control (git), continuous integration, databases, configuration and deployment services. Software-as-a-Service, Platform-as-a-Service and Infrastructure-as-a-Service offerings.
  - Software - analysis libraries, simulators, and data access libraries, as well as thick client applications (desktop visualisation tools). These standard desktop scientific software tools will eventually (with operations support for FENIX) will be deployed for remote access via the Collaboratory.

The COLL architecture is such that operators of services can deploy their services wherever they like while still preserving key integration characteristics inside the COLL ecosystem.

This is primarily a consequence of architectural decisions implicit in the HBP Identity service authentication service provided by the COLL as well as the decision to have implement the COLL as a microservice-based Service Oriented Architecture

(https://en.wikipedia.org/wiki/Microservices).

As a result, there is no coupling between deployment sites for the Apps accessible in the COLL. This ensures that services and Apps can evolve independently and it allows a federation of App providers across Europe. This can already be seen in the Platform release demonstrated by the HBP at the end of the Ramp-Up Phase.

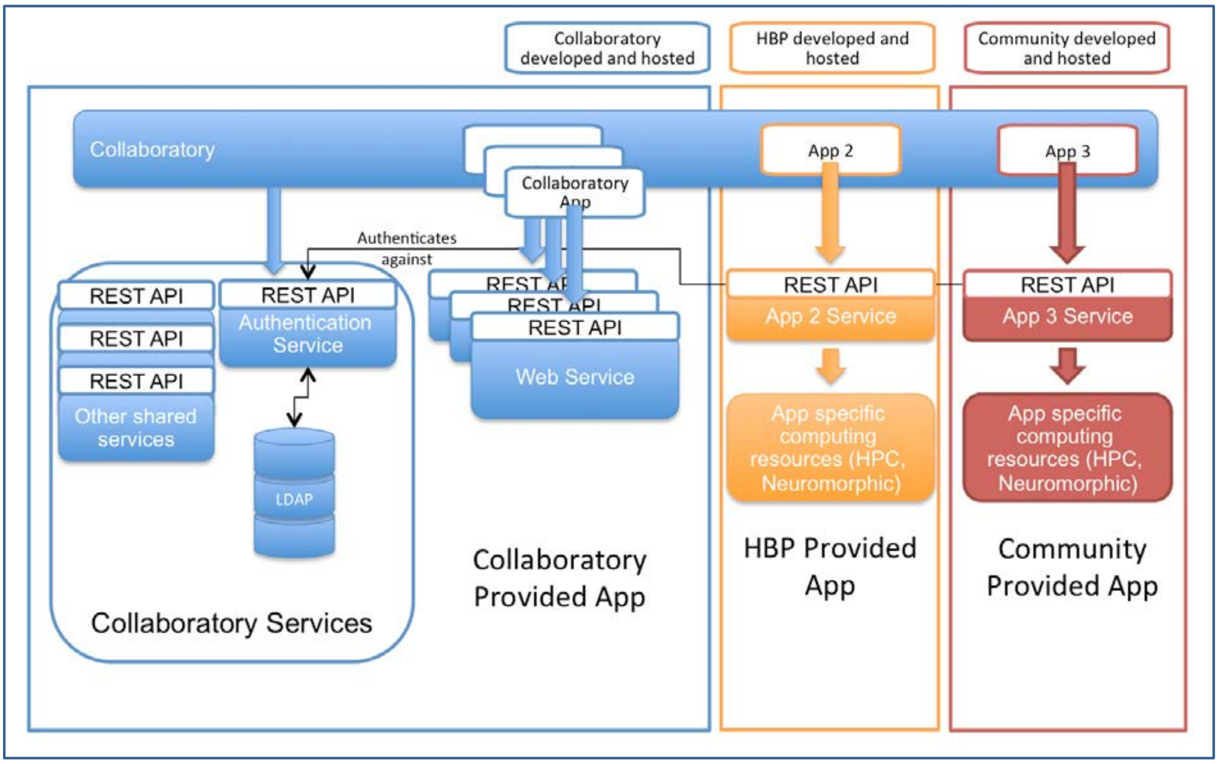


Figure 2: App Federation through the Collaboratory

2.3 FeDaPP/FENIX Architecture Overview

The Federated Data Pilot Project (FeDaPP), an initiative started by SP7 in Q1 of 2016, is planned to be implemented under FENIX. This initiative aims to deliver a federated data infrastructure. Figure 3 shows the high-level architecture design for FENIX which has been guided by the following considerations:

- Data is located close to the data processing
- Federating multiple data resources enables easy replication of data at multiple sites. This capability can be exploited to improve data resilience, data availability as well as data access performance.
- The federated Infrastructure as a Service (IaaS) is put in place and operated by multiple resource providers (initially mainly HPAC centres). The COLL will be used as a portal, but other interfaces may have to be supported as well.
- The Authentication and Authorization Infrastructure (AAI) implementation with OpenID Connect (OIDC) enables web-based applications to be seamlessly integrated with FENIX services.

- Security, flexibility, extensibility and scalability (in terms of users) are key consideration for the architectural design as user requirements and use cases are constantly evolving. In order to future-proof the design of infrastructure services, open standards for federated environments, such as OpenStack and similar solutions that are supported by key vendors, will be leveraged.
- Availability of the federated data infrastructure is a key feature that federation helps to achieve. The involved HPAC data centres individually provide level 1-availability, but by federating, e.g., information services, much higher availability levels can be achieved.
- Besides federation, the FENIX IaaS (see Fig. 3 below) provided by the HPAC data centres offers the possibility to bring archival and active storage close to scalable high-end compute capabilities.

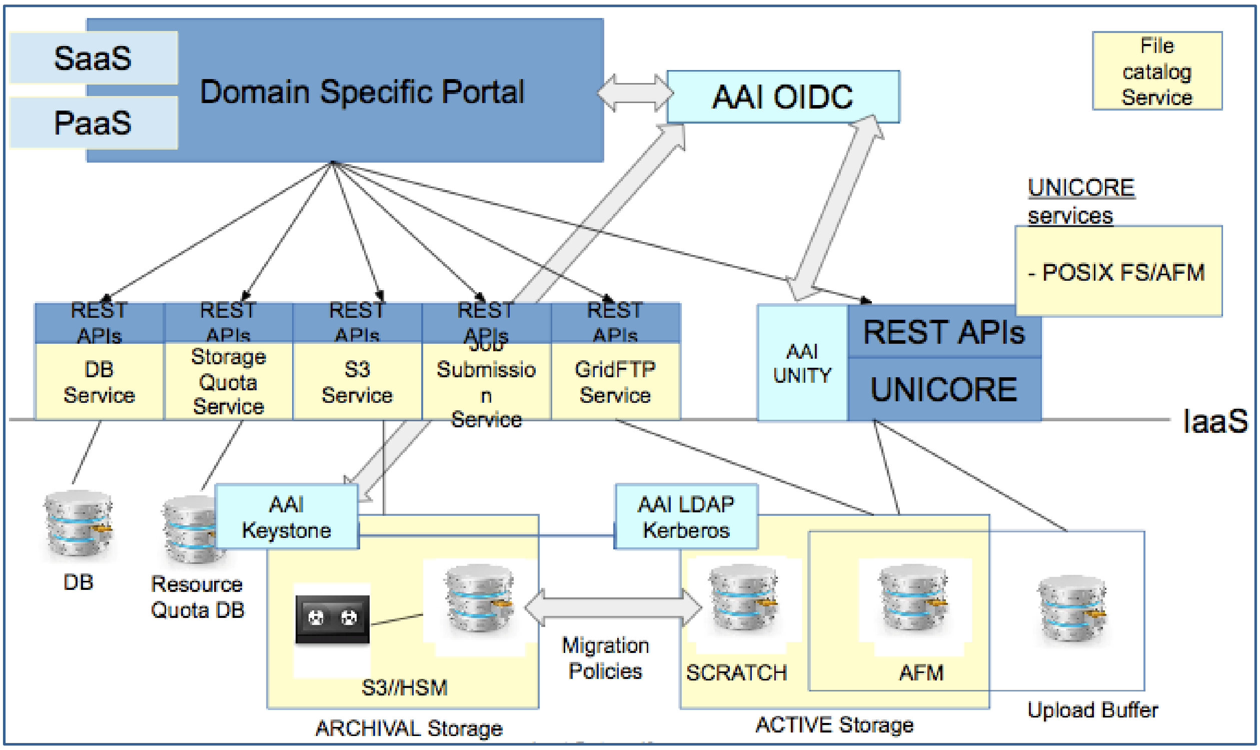


Figure 3: Schematic of the FENIX IaaS offering.

The FENIX IaaS offering provides a set of basic services (DB, Storage Quota, S3, Job Submission, GridFTP, and UNICORE) through REST APIs, as well as AAI. These services can be accesses by Domain Specific Portals, such as the COLL of the HBP. Efforts are under way to integrate newly installed supercomputing systems into the FENIX IaaS offering.

2.4 HBP Repodicible Data Processing and Interaction Architecture (HBP-RDPIA)

The HBP seeks to solve the following issues:

- The Data Sharing problem – find and share data easily.
- The Traceability problem – where did analysis X come from to help answer the question “Why does that plot look like that”?

- Evolving workflows – there is significant overlap between workflows for data-driven models and for predictive Neuroinformatics. These workflows are typically evolving rapidly throughout the model refinement process. Both computational and data transfer flexibility must be maintained in the relationship between the NIP and its underlying compute and storage.
- Efficient performance for IO heavy image processing workloads

2.4.1 Data processing

To address these issues the HBP-RDPIA architecture follows the model outlined in the three figures below. It is important to note that API support for large parts of the service functionality is assumed as it is required to enable complex scripted workflows.

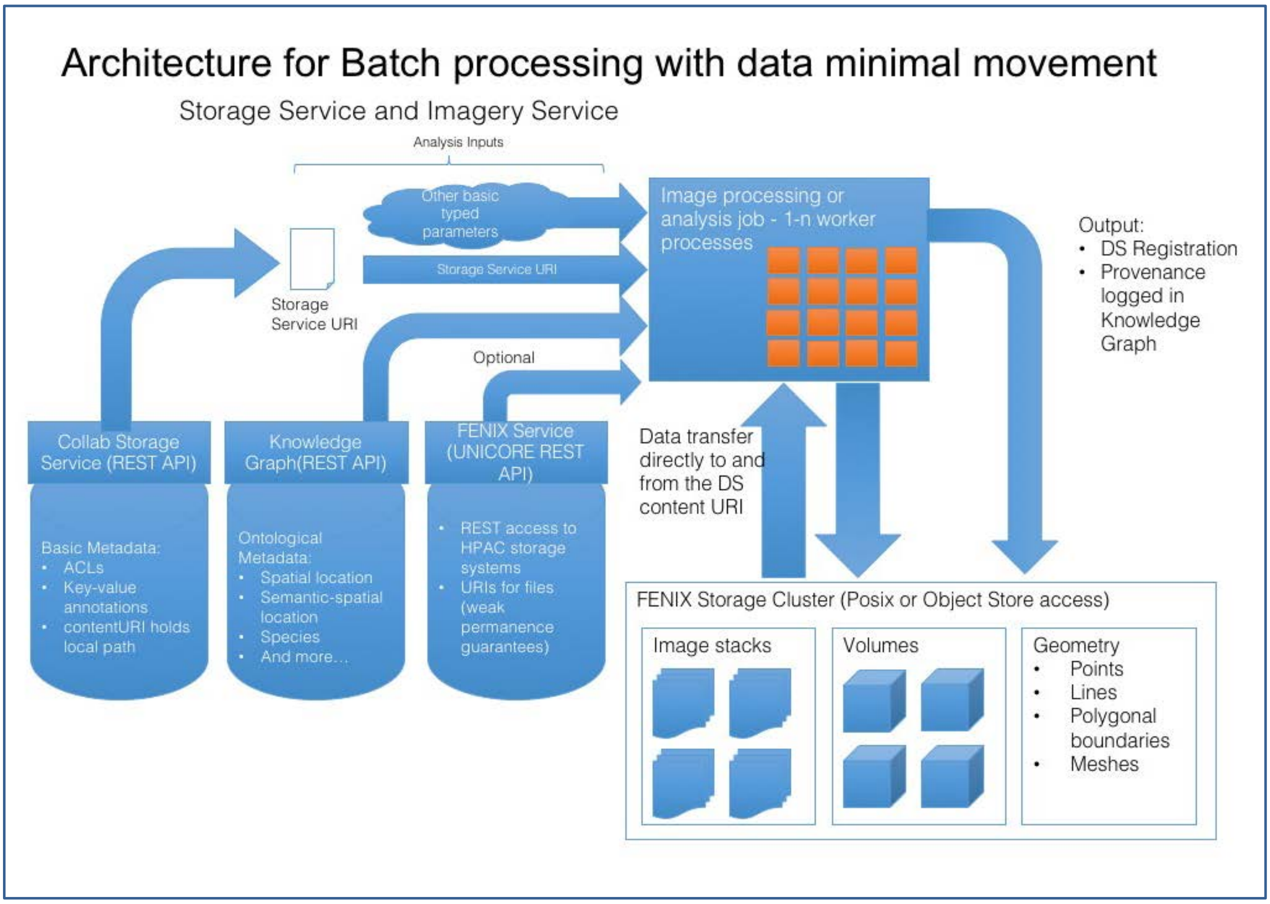


Figure 4: HBP-RDPIA for Batch processing of data with minimal movement

2.4.2 Visualising Data at Federated processing nodes

After data has been processed, users will want to visualise it. These services can be either through a COLL Jupyter notebook or through a web visualisation service such as the Atlas Viewer in the example below.

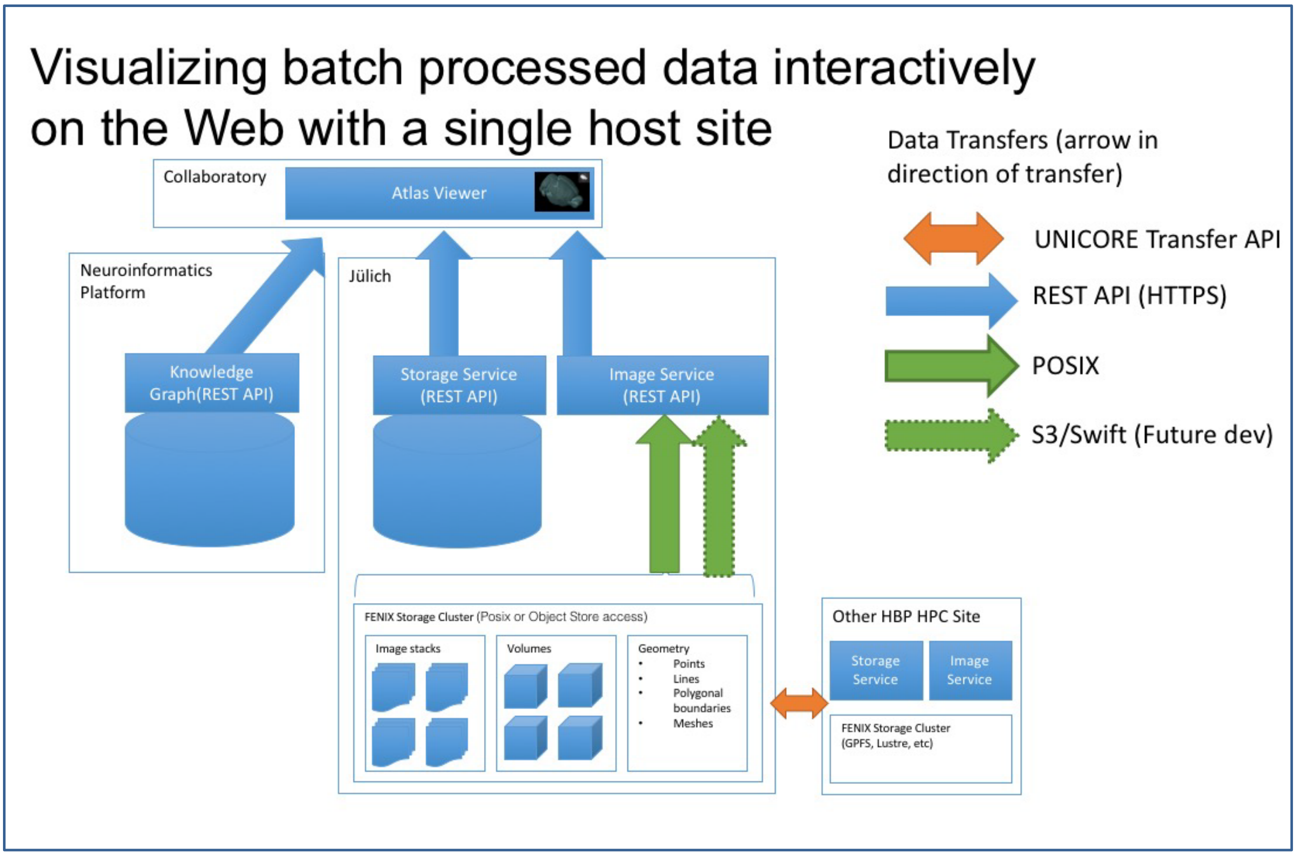


Figure 5: Example web-based data visualisation method

2.4.3 Interactive Data Processing

Equally important for HBP prioritised use cases is API-based usage of this architecture. This is described in the figure below.

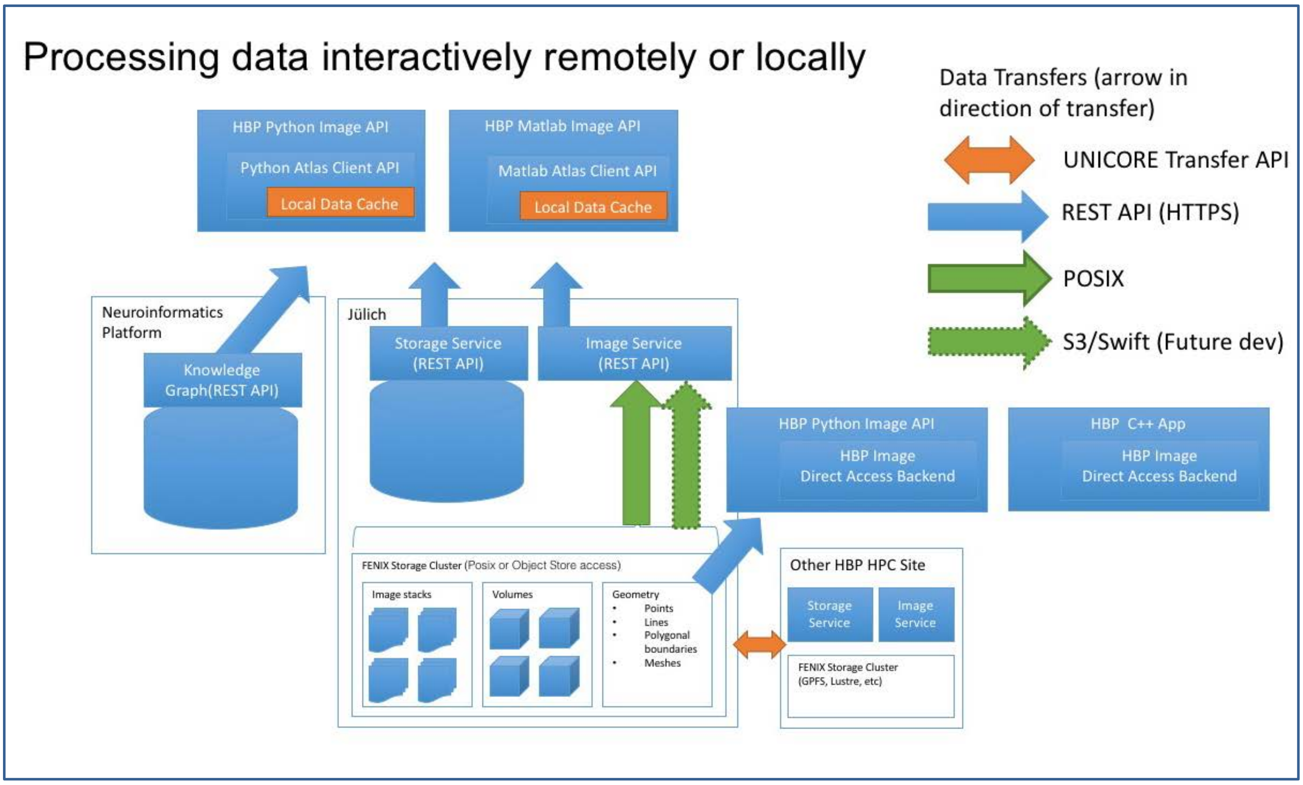


Figure 6: API interactions for remote and local data interactions

2.5 Serving the Science and the Platforms

The Co-Design Projects document how the Platforms will be architected and constructed to service specific scientific use cases. These Platforms are tied together by some common service needs, but the architecture described serves specific platforms in different ways.

1) SP5 will use:
   a) Data processing for atlas building and analytics.
   b) Web visualisation of atlases.
   c) Interactive data processing for rapid iteration of collaborative analytics.
2) SP6 will use:
   a) Data processing for model building and post-simulation analysis.
   b) Web visualisation of models and simulation results.
   c) Interactive data processing for rapid iteration of model building workflows.
3) SP8 may use:
   a) Web visualisation to view medical imaging overlaid on top of the atlases.
   b) Interactive data processing to explore publicly available research data sets or to make queries against federated MIP search services.

The investigation into ethical and privacy constraints of medical informatics use of these services will be done under the tasks of WP5.1, WP5.3 and WP5.5. The adaptation of these services will not be done under the current plan, but will require allocation of SP8 funding for this activity in future phases of the project.

4) SP9 will use:
   a) Data processing for post-simulation analysis.
   b) Interactive data processing for rapid iteration of collaborative analytics.
   c) COLL integration to give users prepared workspaces for to improve user adoption of Neuromorphic computing.
5) SP10 will use:
   a) COLL integration to give users easy access to rich virtual laboratories for the simulation of virtual robots and their environments.

2.6 Accompanying Documents

This document is part of a collection of documents, which form the basis of the HBP Architecture. The other two key documents to be considered alongside this document are D11.3.1-Collaboratory Architecture and FeDaPP_Architecture_Report_V01.docx. The Deliverable D11.3.1 is an updated architectural document which describes in more detail the integration strategies and architecture of the Collaboratory, along with key integration points into the NIP and FeDaPP/FENIX architectures. The FeDaPP Architecture report is a preliminary release of a document, a final version of which will be delivered at a later date. It is included in this architecture package to provide additional details of the FeDaPP architecture to those who are interested.

Finally, these 3 documents should be considered the vanguard of complete body of HBP IT architecture documents, which will also include the SP6, SP8, SP9 and SP10 specifications. Since the Platform release, these have either become living documents or are scheduled for resubmission. The exception is the Neuroinformatics Platform, which, due to its central role in the HBP Platform architecture, is described in the sections which follow.

3. The Neuroinformatics Platform (HBP-NIP)

This section presents the Neuroinformatics Platform (HBP-NIP or NIP), which complements the capabilities of the COLL (presented in D11.3.1) to form the central element orchestrating the dance of the IT architecture. After reviewing the overall goals of the Platform, it explains the architecture that addresses those goals as well as the Use Cases and their resulting requirements and the relationship to other Platforms. Where necessary, it describes in technical detail the architecture of planned integrations between the NIP, COLL and HPAC Platforms.

3.1 HBP-NIP: Overall goals

3.1.1 Principles

1) The HBP NIP is a central piece of the HBP Platform ecosystem that links data consumers with data producers. SP5 must therefore assume a proactive role in establishing and maintaining collaborations with other (consumer and producer) SPs.
2) It is unfeasible to have HBP Project staff “mass curate” community data for all contingencies. The HBP NIP must be driven by the needs of data-producers and data-consumers inside and outside the HBP. It needs to take the scientists by the hand.
3) Data quality depends as much on the data use as on the data production. Data quality must therefore be determined by interaction between target consumers and producers

during the Project Lifecycle and by community interactions with NIP registered data and metadata.

4) Platform development around priority use-cases requires strong support for developers (in SP5 and other SPs) and Platform users. SP5 must become the most collaborative SP in the Project by establishing and supporting interactions between all SPs.

3.1.2 Agile Data Integration

The HBP NIP is a collection of tools, services and specifications that focus on serving the needs of data-producers and data-consumers. In SGA1, the HBP has introduced a planning mechanism called *Co-Design Projects* (CDP), which ensures that tools and services are developed around strategic scientific use cases and in close interaction with scientists (platform users).

The **Co-Design Project** model has been extended with an on-going detailed component mapping activity with the intent of building a Product Breakdown Structure for the HBP. This is part of the Project Lifecycle Framework approach taken by DPIT (see D5.6.1) and the Science and Tech Coordinators in SP11. This analysis provides crucial insight on which data flows are occurring in the project and which Components are critical to allow the Platforms to support the project's scientific use cases. Data collected in initial steps will be refined throughout SGA1 to ensure that the project is coordinated with more clarity as time goes on. Synchronisation between producer and consumer is critical to ensure efficient use of resources.

The exploratory, or set-up, phase is where researchers determine candidate protocols to generate a certain type of data. They will follow the Project Lifecycle Framework to ensure that there is alignment between the budget of the producer and the quality needs of the consumer. Early parts of the Project Lifecycle will establish the quality criteria for a given data project, along with data analyses to ensure that outputs achieve unambiguous quality metrics.

Quality metrics such as these will never be total. The data-driven modelling approaches will provide a critical element of cross validation. This will reinforce the principles of integrating data into models as early as possible, and data generation activity should be paired with an immediate consumer (HBP or otherwise) wherever possible.

Once the producer and the consumer have agreed a data standard, the data producer will then apply this to all data produced during the later (engineering) phases of the project lifecycle. In these phases, the protocol applied to generate the data is generally not altered, to guarantee consistent data generation.

The Project Lifecycle for this type of situation encourages researchers to collaborate with each other using multiple formal phases and checkpoints. In brief:

1) **Project Implementation Proposal** – agree on what will be achieved qualitatively. Identify early proof of concept work required to mitigate technical risk.

2) **Testing sample** (to resolve format and other technical issues) – Based on this sample, first automated tests are written and integrated into a test suite. The main focus of the tests is format and format semantics.

3) **Representative sample** (test qualitative and quantitative issues). - This step will require multiple iterations, additional tests are expected to be added to the test suite, focusing on quantitative validation and format semantics.

4) **Production sample** (test qualitative and quantitative issues for production pipeline) - This step should result in additional tests being added to the automated test suite,

focusing on quantitative validation, and data variability induced changes in format semantics.

5) **Production pipeline operation** - Additional target Use cCses should be developed with HBP external community members. Additional quality criteria for these use cases must be defined in consultation with the wider community as it should result in additional tests being added to the automated test suite, focusing on quantitative validation and format semantics.

Agile projects using the Project Lifecycle and which involve two teams, need a specific point for quality checkpoints. The following diagram illustrates the relationship between two example teams involved in such a process.

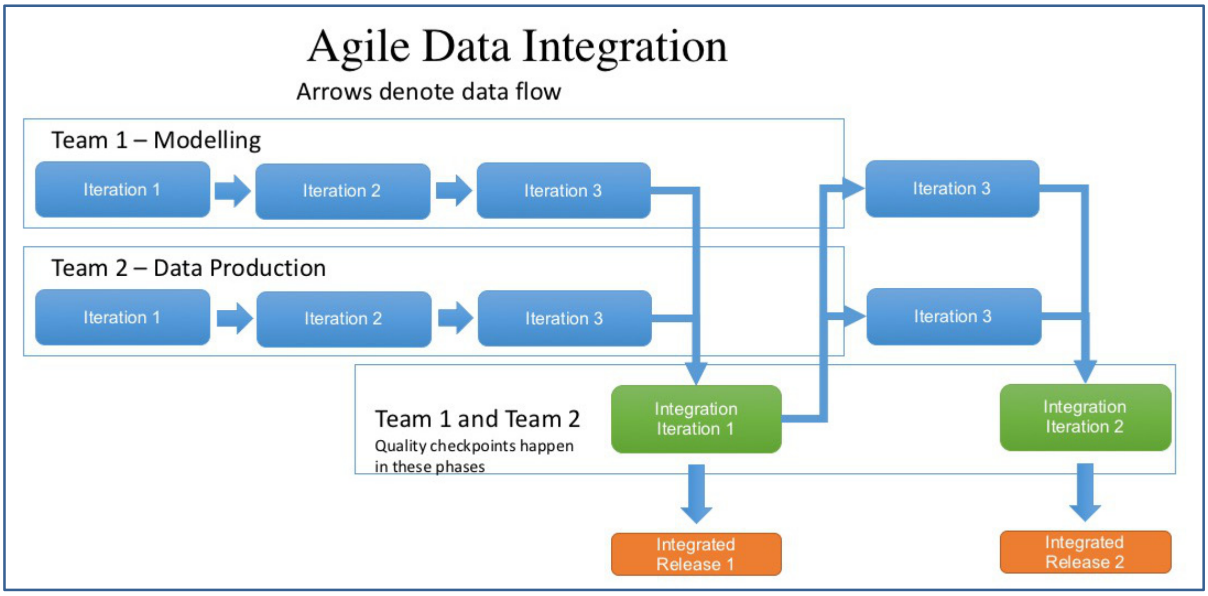


Figure 7: Agile Data Integration and Quality Checkpoints

3.1.3 *Project-wide Data Accessibility, Integration and Quality Model*

Ensuring that all data produced by HBP contributors is accessible to the community is one of the basic pledges the project has made to the community. To ensure we are living up to this promise, we have adopted a simple of criteria to evaluate the state of the data and models. All data and model producers are encouraged to reach the highest level of Data Accessibility Criteria (DACs), but for some of the intended uses of data and models, as well as ground reality in the community, may prevent this from being adopted. This approach of course also applies to Partnering Project and Strategic Research Partnerships.

Table 1: Data Accessibility Criteria

Data Accessibility Criteria	Stages
Provenance	1) Data Generation Provenance is accessible in text format. 2) Data Generation Provenance uses ontological terms for protocol, methods, experimenter roles, and entity classifications. 3) Data Generation Provenance is stored in a well-defined data provenance model (e.g. W3C PROV-DM, MINDS)
Spatial Anchoring	1) Location of Data, or a source sample, is available.

	2) Location of Data, or a source Sample, can be linked to spatial ontology. 3) Data, or it's source Sample, can be aligned to a standard reference space. (e.g. ABA, MNI, Collins)
Accessibility	1) Data is saved in a proprietary format, which is accessible via a software vendor. 2) Data is saved in an open format, which can be accessed via a publicly available file reader/service. 3) Data is available via a service that allows it's export in multiple formats.

The quality of data depends as much on the use of the data as on the production of data. The data quality metrics should therefore be defined in a dialogue between data producer and data consumer. In some cases, a single person or small group holds both roles. While this will also be true for some parts of SGA1 modelling work (e.g. Tasks in WP1.2 and WP6.2) HBP is committed to ensure that data is produced according to quality metrics that allows data reuse to addressing a wide range of scientific questions. Practically this means that researchers of different backgrounds (experimental neuroscience, medicine, computational modelling and ICT) need to define common data quality standards for data based on its intended use. Given the interconnected nature of work in SGA1 (see data flow and task dependency work) the minimal number of Task leaders needed to produce such a quality standard is two.

Beyond the initial production process, the NIP can support additional quality metrics based on usage. These will be based on common methodologies used in other communities. The software development community has dealt with this by developing tools and pragmatic procedures to deal with this in practical terms.

Arguably Github is the single largest code repository in the world. While a very large percentage of projects on Github are uninteresting to many developers, developers will agree search on Github for code before starting a project from scratch. This is largely due to the quality of the service and the community quality metrics it supports. Key quality metrics from Github:

1) Forked – number of users who have created their own copy to makes changes on.
2) Starred – number of users who like the project
3) Watchers – number of people who want to be informed in changes to the project.
4) Most Active – actively developed projects (recently used)

In a first approach Data/Model quality will be assessed along similar lines. Namely:

1) Starred – number of users who like the dataset
2) Cited – number of users who cited the dataset (for public datasets)
3) Used – number of users who have used the dataset in an HBP Platform tool.
4) Most active – the data/model is part of a Collab that has regular activity.

Items 1, 2, and 4 will be implemented as minor additions to the Knowledge Graph. However, item 3 requires a system capable of capturing large amounts of workflow provenance. The D5.6.1 work plan update includes a strategy to address this issue while minimising risk. See Task T5.5.1 for more details.

3.1.4 User Interface (UI) Support for pragmatic data integration

The previous section on Data Integration and Quality Management describes a strategy based on successful models for communities around software development. This strategy will require a very easy to use system for uploading data and annotating it with metadata so that it can be found.

There is a data registration interface in development at the BBP that needs to be prioritised for public release. Without public access to data registration, the NIP relies on limited SP5 resources to build its library of accessible data sets.

Striking the right balance between necessary detail and ease-of-use will be a very challenging User Experience (UX) problem.

In addition, community engagement tools will need to be built which support the proposed community quality metrics. Furthermore, the community quality metrics will likely need to be adapted to inform HBP Platform users of interesting datasets.

An additional effort will be needed to ensure that API support encourages computational users to easily leverage NIP registered results and to register their own results.

3.2 HBP-NIP: Architecture

3.2.1 Data Architecture

A large part of the functionality of the NIP is value-added services (SaaS – Software as a Service and PaaS – Platform as a Service) on top of existing data repositories (acting as IaaS – Infrastructure as a Service). This allows the NIP to leverage the Data Management practices of existing service provides, but it also requires that the NIP validate their vendors for compliance with best practices. This process will be made simpler in SGA1 by focusing on repository providers who are HBP partners or on commercial partners. Existing interactions with various data repository service providers are outlined below.

NIP teams will work closely with SP7 HPAC/FENIX teams to ensure that NIP requirements are clearly communicated and appropriately tested in all stages of development. The objective will be to deploy the following architecture:

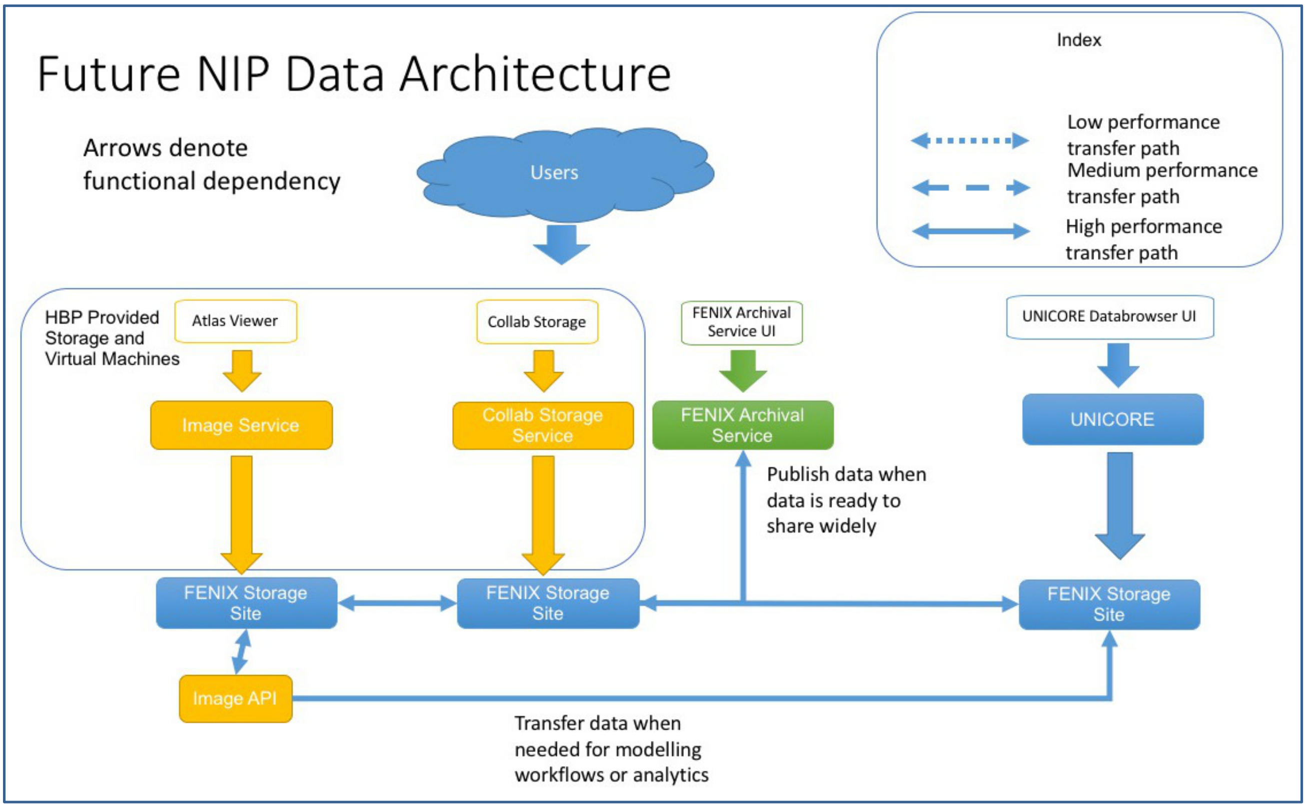


Figure 8: Future NIP Data Architecture

3.2.2 NIP Curation dataflow

The architecture above allows the NIP to provide tools and services to allow researchers to curate their datasets. In addition, some of the same tools and services are crucial to delivering data to end-users. Below is a high-level dataflow diagram which shows how data moves from data producers, through the various services into the hands of data consumers.

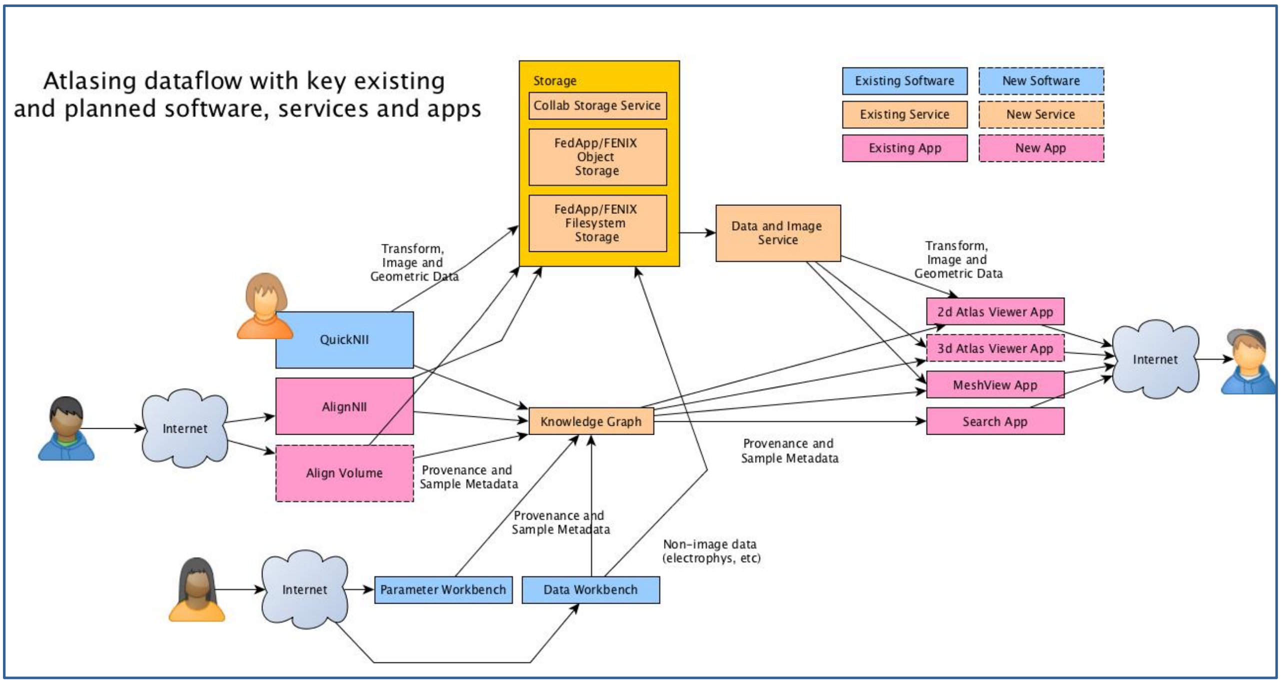


Figure 9: High-level curation dataflow

3.2.3 NIP-Collaboratory Integration

The NIP is developed as collection of web accessible services, web accessible applications, downloadable software packages and downloadable datasets. The web accessible services and applications follow the COLL Platform model described in D11.3.1. The downloadable software packages can be found in the Collaboratory Software Catalogue (https://collab.humanbrainproject.eu/#/collab/19/nav/2108). Datasets are found through the Neuroinformatics Search App. Figure 10 summarises the COLL-NIP interaction.

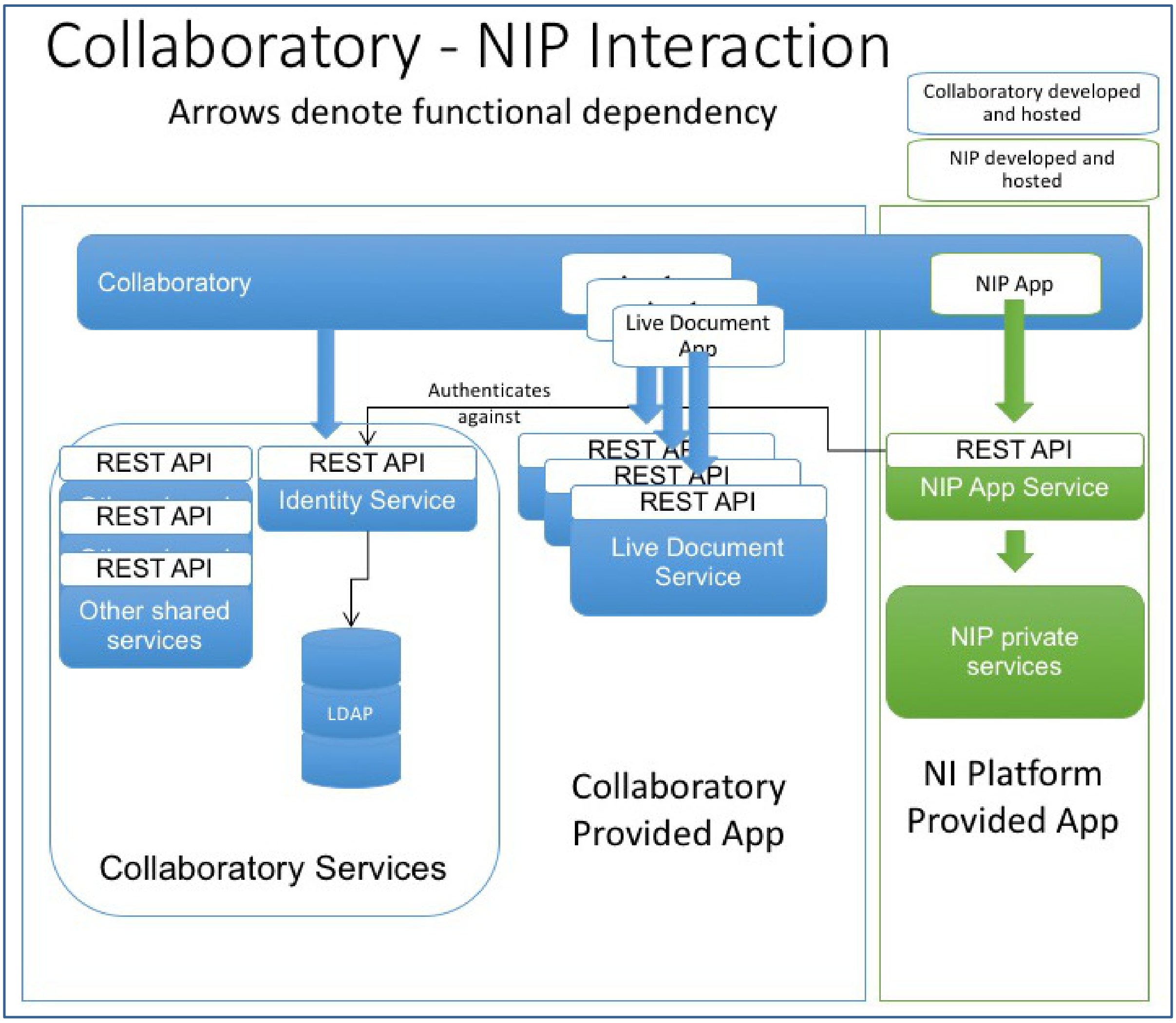


Figure 10: Collaboratory-NIP Interaction

3.2.4 Standard NIP App architecture

The web based App which the NIP has deployed through the COLL follow the common model. They provide a UI which interacts with a microservices architecture for application persistence, as described in Figure 3.

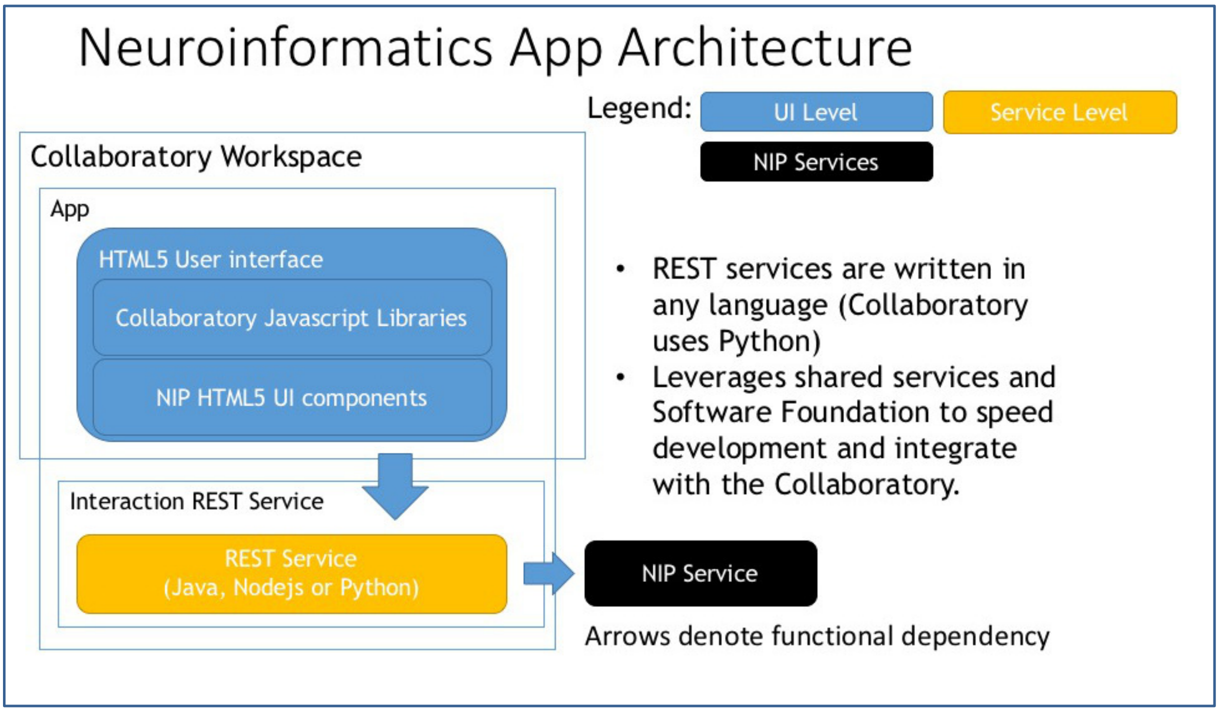


Figure 11: Neuroinformatics Web Application Architecture

3.2.5 Standard NIP API architecture

Most web applications developed for the NIP follow a common model. Wherever possible the implementation leverages proven open source technologies for all parts of the technology stack (see Figure 12).

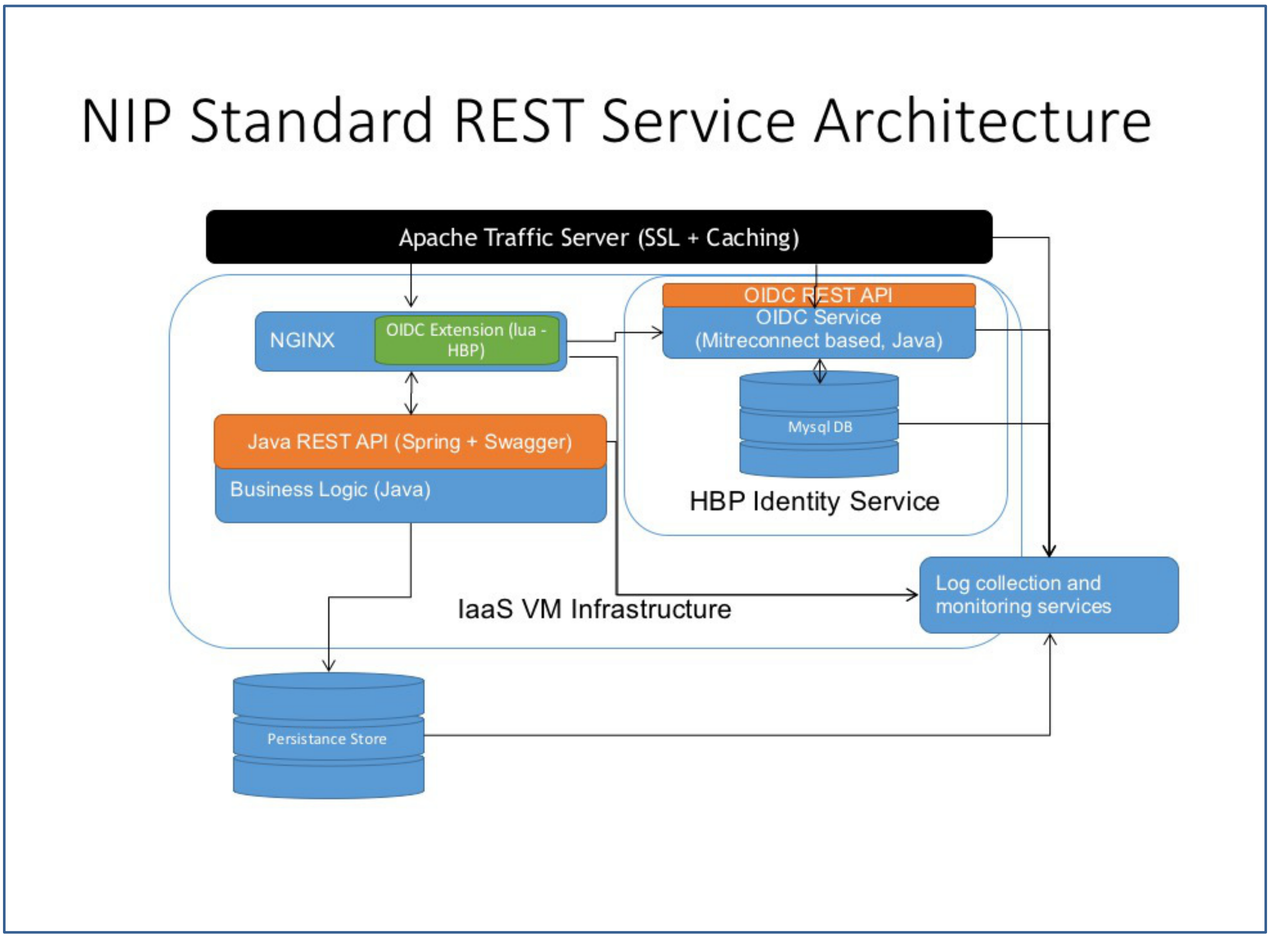


Figure 12: Standards-based NIP REST service architecture

This standard stack reuses system components developed by the Collaboratory team for NGINX reverse proxying and for the NGINX-hosted, Lua-based OIDC authentication module. It also makes use of the Swagger API description system (http://swagger.io/). There are specific variations on this architecture which are noted in the Component Status section.

3.2.6 Physical Architecture and Infrastructure Dependencies

The deployment and configuration management system described in Section 4.3.3, is the model used to support the NIP capabilities deployed for the Ramp-Up release. The physical Architecture of the Ramp-Up release is described below.

This will be replaced by an HBP Managed Physical Infrastructure running on commercial Cloud Providers and HPAC Platform sites. There will be an evaluation process for each service to determine whether it should be deployed to a commercial Cloud Provider or to an HPAC site. This will take into account the inherent advantages of the respective solutions, with cost on one hand and data locality and massive compute capabilities on the other hand.

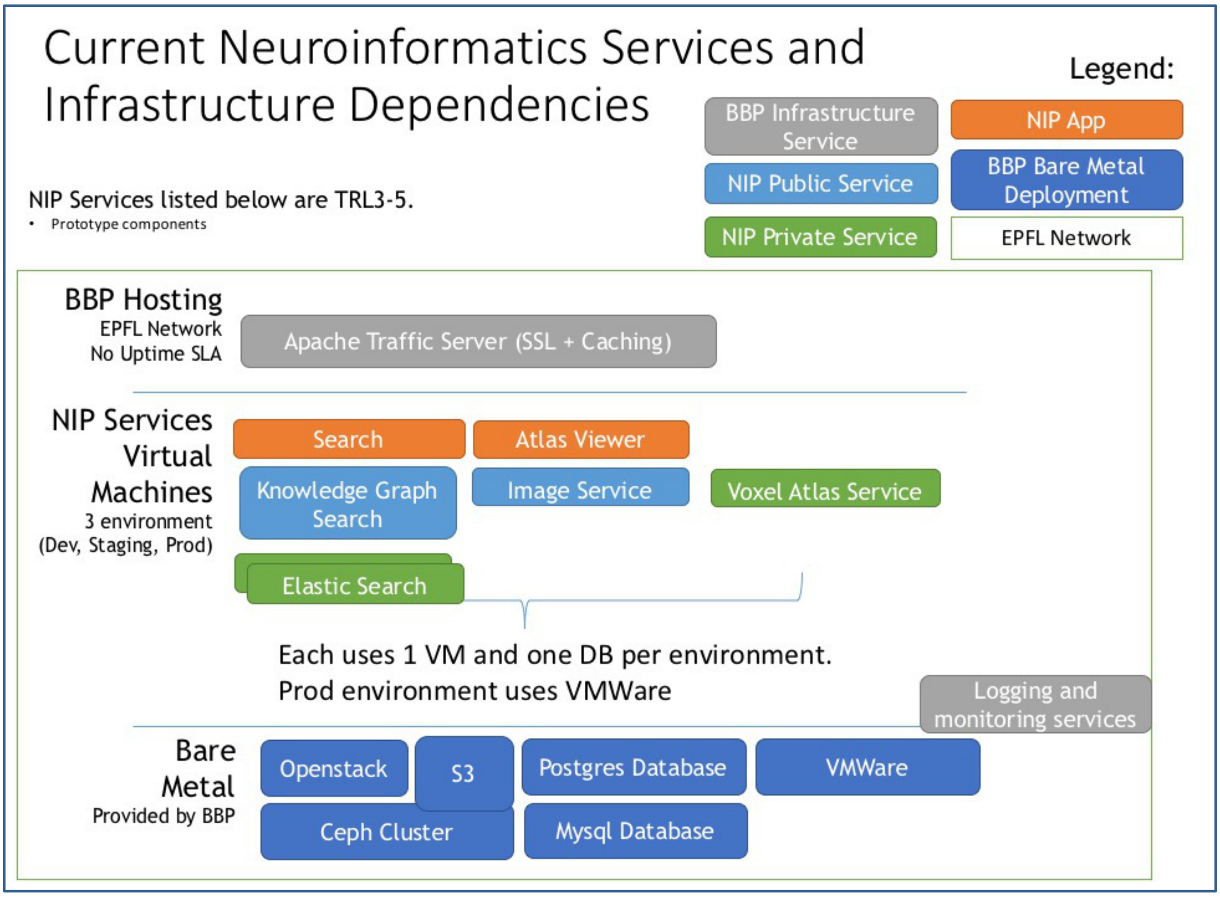


Figure 13: Legacy Neuroinformatics Services and Infrastructure

The HBP-COLL is architected to be a best-of-breed Software-as-a-Service (SaaS) offering. This has proven to be a robust model for organising large, loosely coupled software platforms in industrial settings in recent years, and has proven to be viable model for software delivery through extensive historical exploration.

As a result of its SaaS ambitions, the COLL is made up of components that are accessed through a network-accessible service API. The decision in favour of service orientation has implications for the architecture in terms of development, testing, deployment and monitoring discipline. Failure scenarios must be handled gracefully, and regular service upgrades must proceed with minimal User impact. User authentication must be handled in the most secure manner possible while still maintaining ease-of-use for non-technical Users. With all of this in consideration, the architecture presented below will provide a foundation that will enable the success scenarios for Users in a reliable, secure and user-friendly fashion.

3.3 Components

The following section describes web apps, services and software which form the backbone of the NIP. Other web apps, services and software are developed under SP5, but are not included here because they are not primary activities of a Task or are not essential to the atlas building and data sharing ecosystem.

It is expected that this list of will grow in SGA1 and beyond.

In all cases below, Apps Components have web services behind them doing the heavy lifting of the application functionality. However, in cases where the services are not offered as an independent API, only the App Component is listed.

3.3.1 Data Components

This section describes key data components which form the backbone some of the atlasing and data curation efforts in the NIP.

Waxholm Space rat brain atlas, v.2.0

Data released April 2014 (v. 1.0), November 2014 (v. 1.1), and February 2015 (v. 2.0). Public access through INCF Software Center.

http://software.incf.org/software/waxholm-space-atlas-of-the-sprague-dawley-rat-brain

Open access volumetric atlas of the Sprague Dawley rat brain based on ex vivo high resolution MRI and DTI, anchored to both Waxholm Space and stereotaxic space, in formats prepared for viewing with open access tools.

The atlas template consists of T2*-weighted anatomical MRI at 39 µm and diffusion tensor imaging (DTI) volumes at 78 µm resolution from an 80 day old male Sprague Dawley rat. Resampled DTI volumes matching the anatomical MRI (1024x512x512) are also provided.

Anatomical delineations in the atlas are based on image contrast, described in detail in the accompanying publications (Papp et al. 2014[1], Papp et al. 2015[2], Kjonigsen et al. 2015[3]). The latest version of the atlas contains 79 structures and includes a detailed parcellation of the hippocampal region. The atlas is anchored to both Waxholm Space and the stereotaxic space.

The atlas and accompanying MRI/DTI template volumes are provided as standard gzipped NIfTI files (.nii.gz). Labels are made available for ITK-SNAP (.label) and MBAT (.ilf). Bundled zip downloads are provided for the latest version of the most commonly used atlas and template files.

---

[1] Papp EA, Leergaard TB, Calabrese E, Johnson GA, Bjaalie G (2014) Waxholm Space atlas of the Sprague Dawley rat brain. Neuroimage 97:374-86. doi: 10.1016/j.neuroimage.2014.04.001 [Atlas version 1.0]

[2] Papp EA, Leergaard TB, Calabrese E, Johnson GA, Bjaalie JG (2015) Addendum to "Waxholm Space atlas of the Sprague Dawley rat brain" [NeuroImage 97 (2014) 374-386]. Neuroimage 105:561-2. [Atlas version 1.1]

[3] Kjonigsen LJ, Lillehaug S, Bjaalie JG, Witter MP, Leergaard TB (2015) Waxholm Space atlas of the rat brain hippocampal region: Three-dimensional delineations based on magnetic resonance and diffusion tensor imaging. Neuroimage. 108:441-9. doi: 10.1016/j.neuroimage.2014.12.080 [Atlas version 2.0]

Provided under Creative Commons Attribute Non-Commercial Share-Alike license (http://creativecommons.org/licenses/by-nc-sa/4.0/).

3.3.1.1 Standardized anatomical landmarks for registration of whole brain imaging datasets to mouse and rat brain Waxholm Space

Data released January 2015. Public access through INCF Scalable Brain Atlas.

http://scalablebrainatlas.incf.org

Standardised set of anatomical landmarks for registration of whole brain experimental image data from the mouse and rat brain to Waxholm Space(Sergejeva et al. 2015[4]). The use of standardised landmarks in combination with a common coordinate system for the rodent brain increases speed and accuracy of data integration.

Internal landmarks of the C57BL/6J mouse brain and the Sprague Dawley rat brain (Waxholm Space templates) have been identified across different MRI contrasts (T1, T2, T2*) and other modalities in different specimens and at different slice acquisition angles and different image resolutions. The use of these landmarks increases the accuracy of registration of experimental volumetric data sets to atlas templates and thereby facilitates comparisons of data from different experiments and data analysis requiring use of standardized atlas space.

3.3.1.2 Big Brain ultra-high-resolution 3D model of a human brain

Data released in 2013 (release with the original Science publication[5]). Second release with improved image registration quality and g/w matter classifications in 2015.

The Big Brain is an ultrahigh-resolution three-dimensional model of a human brain at nearly cellular resolution of 20 micrometers, based on the reconstruction of 7404 histological sections. It provides considerable neuroanatomical insight into the human brain, and allows the extraction of microscopic data for modelling and simulation. BigBrain enables testing of hypotheses on optimal path lengths between interconnected cortical regions or on spatial organisation of genetic patterning, redefining the traditional neuroanatomy maps such as those of Brodmann and von Economo.

Together with the reconstructed histological data, a gray and white matter classification as well as 3D surface reconstructions of the pial and white matter surfaces are provided.

The original data and reference material is publicly available at http://bigbrain.loris.ca

FTP access to the data releases via

ftp://bigbrain.loris.ca/BigBrainRelease.2013

ftp://bigbrain.loris.ca/BigBrainRelease.2015

The 3D volume can be interactively browsed as a part of the HBP human brain atlas at:

https://nip.humanbrainproject.eu/atlas/#/

The data is released under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0

[4] Sergejeva M, Papp EA, Bakker R, Gaudnek MA, Okamura-Oho Y, Boline J, Bjaalie JG, Hess A (2015) Anatomical landmarks for registration of experimental image data to volumetric rodent brain atlasing templates. J Neurosci Methods. 240:161-9. doi: 10.1016/j.jneumeth.2014.11.005.

[5] Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M., Bludau, S., Bazin, P., Lewis, L., Oros-Peusquens, A., Shah, N., Lippert, T., Zilles, K., Evans, A., 2013. BigBrain - an ultra-high resolution 3D human brain model. Science Vol. 340 no. 6139 pp. 1472-1475

International Public License.

3.3.2 App Components

This section describes web apps which form the backbone of the NIP.

3.3.2.1 Atlas Viewer App

Description: An Openlayers based viewer for 3d voxel datasets

Current URL: https://nip.humanbrainproject.eu/atlas

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#atlas-viewer

Depends on Services: Image Service, HBP Identity Services

Depends on APIs: Openlayers

Current TRL: TRL5

SGA1 Target TRL: TRL7

Further Details:

This App is a 2d web-based atlas viewer for very large multi-species volumetric datasets. This system is deployed to display multiple datasets of more than 1 teravoxel overlayed in the same view.

3.3.2.2 Search App

Description: A faceted search interface for data registered in the Knowledge graph.

Current URL: https://nip.humanbrainproject.eu/search

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#search-client

Depends on Services: Knowledge Graph Search

Current TRL: TRL4

SGA1 Target TRL: TRL7

Further Details:

This App is a search interface to the Search service and serves to allow NIP users to discover data in HBP and 3rd party repositories which supports the HBP strategic use cases.

3.3.2.3 Data Workbench

Description: A workbench for data curators and data providers to manage the metadata needed to make their data discoverable. This app will continue and expand on work delivered in the Data Registration App.

Current URL: https://nip.humanbrainproject.eu/registration/

Development Standard: BBP Standard

Documentation (prev version): https://nip.humanbrainproject.eu/documentation/user-manual.html#registration-app

Depends on Services: Knowledge Graph Service, HBP Identity Service

Current TRL: TRL4

SGA1 Target TRL: TRL7

Further Details:

In the next phase of development, this App is intended to support a range of dataset curation workflows through the entry of selected metadata. The workflows it will focus on are the so-called "light integration" workflows, with the focus collection of basic metadata needed data discovery. More specialised tasks around Atlas building will left to the AlignNII app described below.

The Data workbench will include a registration UI similar to the current Dataset Registration UI shown in Figure 14 and Figure 15, below.

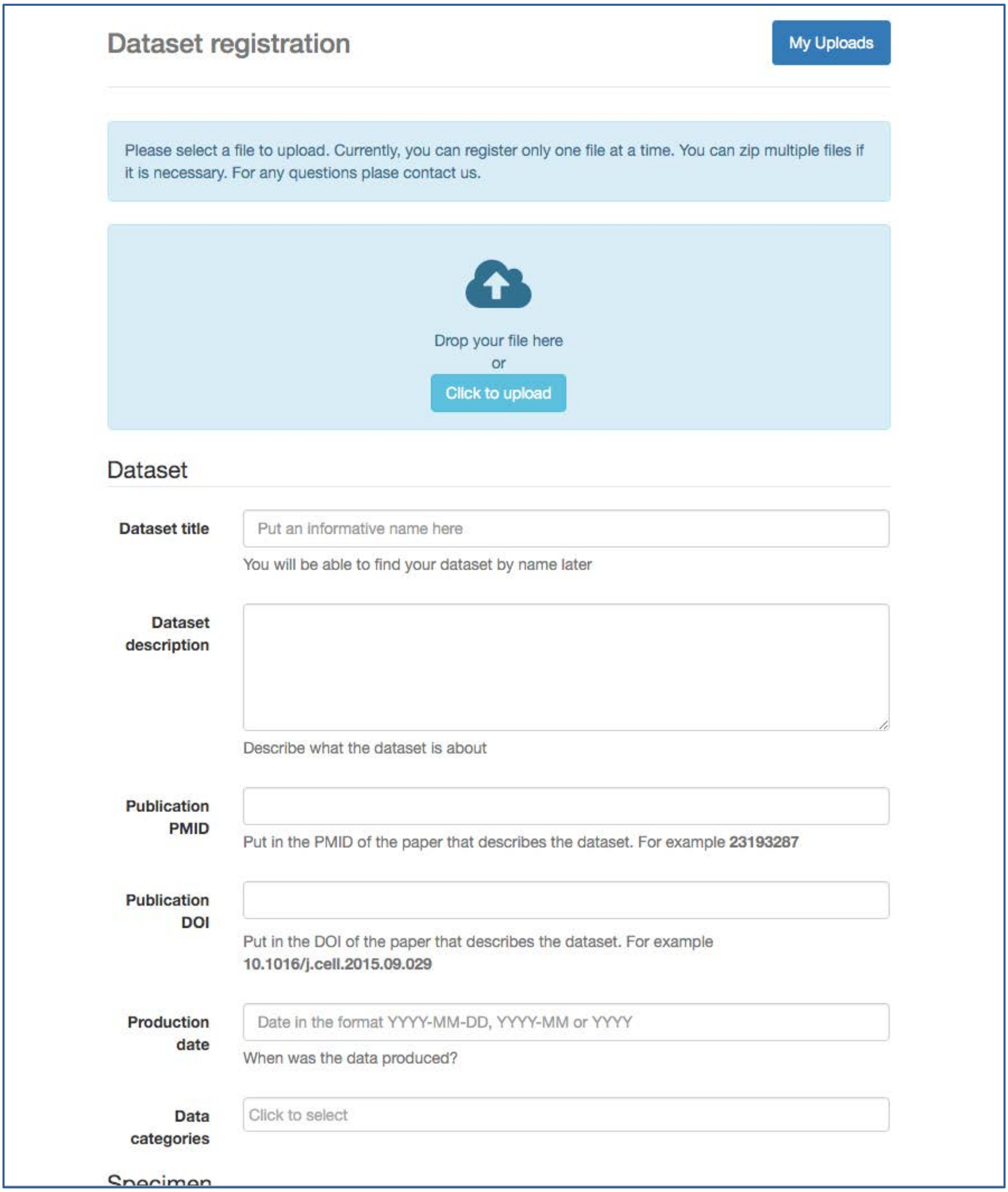


Figure 14: Data Registration Interface, page 1/2

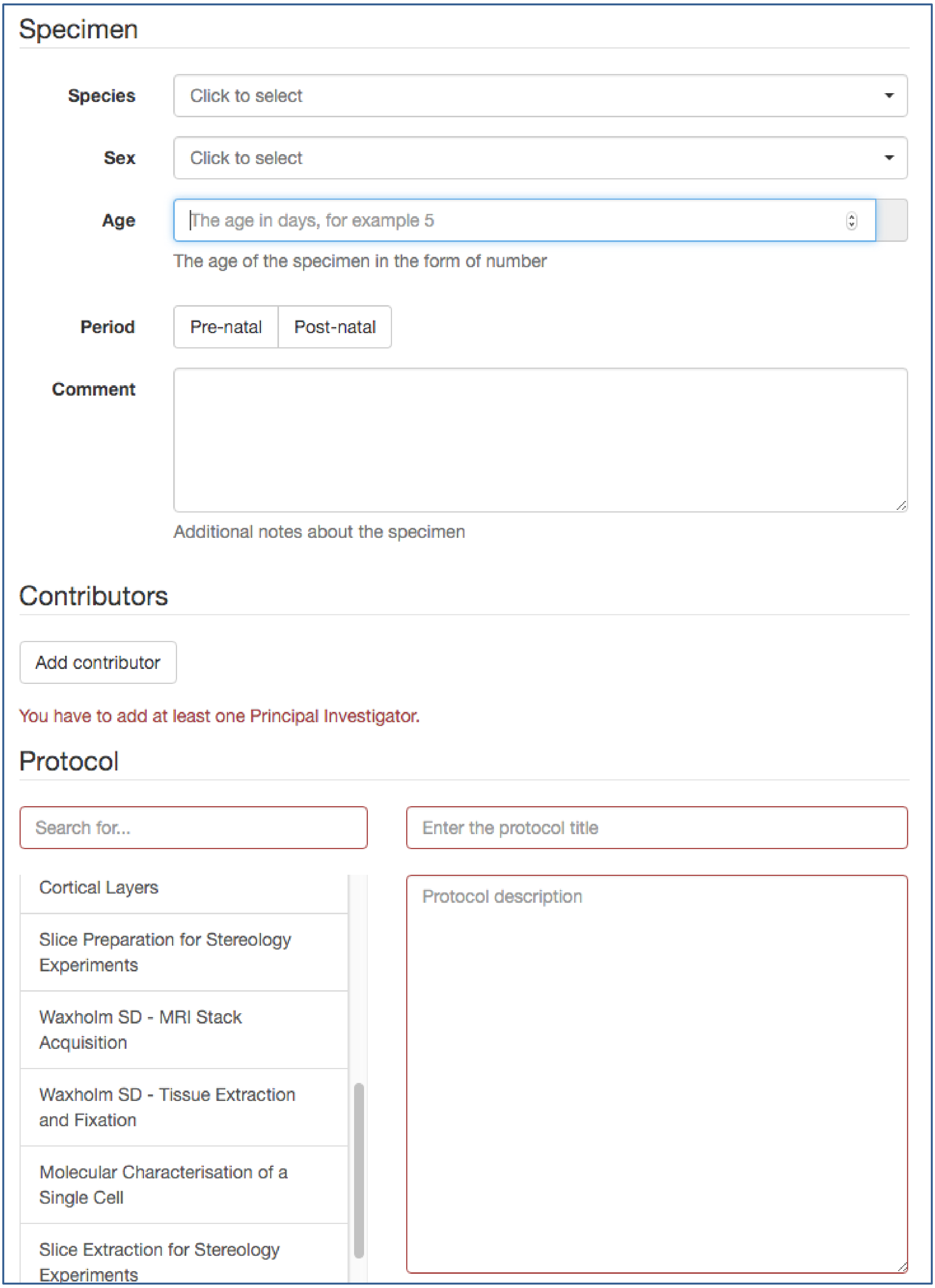


Figure 15: Data Registration Interface, page 2/2

3.3.2.4 Parameter Workbench

Description: A workbench for data curators and modellers to document the provenance of key parameters used in their models.

Current URL: NA

Development Standard: HBP Standard

Documentation: NA

Depends on Services: Knowledge Graph Service, HBP Identity Service

Current TRL: NA, new development

SGA1 Target TRL: TRL4

Further Details:

This App is intended to support a range of modelling workflows through the registration of parameters and parameter sets.

3.3.2.5 AlignNII App

Description: Online tool for anchoring of 2D experimental image data to 3D atlas templates

Current URL: http://www.nesys.uio.no/AligNII/

Development Standard: UIO Standard

Documentation: http://www.nesys.uio.no/AligNII/

Depends on Services: None

Current TRL: TRL7 (SLA not defined or monitored)

SGA1 Target TRL: TRL7 (SLA defined and monitored)

Further Details:

Tool integrated in the web-based Rodent Brain Navigator data system: internal HBP release October 2014, with most recent update February 2015.

Tutorial with video clips demonstrating the present version of the tool: http://www.nesys.uio.no/AligNII

Tool for user guided registration of 2D image data to volumetric atlas templates for the mouse or rat brain. With use of the tool, data are anchored to Waxholm Space (mouse and rat) or Allen Brain atlas space (mouse), facilitating data integration through standardised coordinate systems.

AligNII is a tool for user-guided registration of 2D experimental image data, typically high resolution microscopic images, to standardised volumetric brain atlas templates, such as the Waxholm Space atlases for the mouse and the rat, and the Allen Brain atlas for the mouse. A key feature in the tool is the capability to generate user defined cut planes through the atlas templates, matching the orientation of the cut plane of the 2D experimental image data, as a first step towards anchoring of images to the relevant atlas template. AligNII supports multi-modality 3D atlas packages and one-click switching between image modalites (e.g., MRI and DTI) while preserving the view location and UI settings in general, and in-depth numerical control over cut plane settings. The tool has been tested extensively and is improved based on user feedback. A series of research projects are currently using the tool for registration of project data to standardised atlas space.

AligNII uses NIfTI as standard container for volumetric and Java to handle the (NIfTI) data on the Navigator Web server. The current user interface technology is Adobe Flash. In order to get familiar with these main components, a 3-standard-planes viewer was developed first, ViiNII. This viewer already covered the most fundamental technical details, such as getting slices from the volume in near real-time, supporting various voxel types (adjustable intensity levels, direct RGB, palettised/labelled RGB with actual label names to display), coordinate transformations, and some user interactions for manipulating the views, which were presented via a windowed interface, inside the browser. Then this viewer was extended into AligNII adding a flexible cut with specialised controls and support for loading (histological) images and anchor them to the underlying volumetric data (typically an atlas).

Upcoming developments:

As plug-in technologies for browsers quickly lose popularity and support, AligNII is moving to HTML5. The “ViiNII phase” in HTML is nearly completed, thus a working (although not feature-complete) version of AligNII in HTML5 is expected to be ready early in the SGA1 phase.

Once this priority work has been completed, deeper integration with the KnowledgeGraph, Collaboratory and FENIX services are planned for the end of SGA1.

3.3.2.6 MeshView App

Description: Online viewer for Allen Brain mouse atlas and Waxholm Space atlases for mouse and rat

Current URL: NA

Development Standard: HBP Standard

Documentation: NA

Depends on Services: None

Current TRL: TRL7 (no SLA defined)

SGA1 Target TRL: TRL7 (SLA defined)

Further Details:

Web applications, internal HBP release March 2015, with updated versions during 2015 and full public access through The Rodent Brain WorkBench from June 2015.

http://rodentbrainworkbench.org http://rbwb.org [MeshView of mouse and rat brain atlases]

Waxholm Space Atlas of the Sprague Dawley rat brain

- meshes for 79 anatomical structures - 80 with inner ear) [WHSRat]
- http://www.nesys.uio.no/MeshGen/MeshView.html?bitlas=WHSRatV2.bitlas

Waxholm Space Atlas of the adult C57BL/6J mouse brain

- meshes for 26 anatomical structures - 27 with inner ear) [WHSMouse]
- http://www.nesys.uio.no/MeshGen/MeshView.html?bitlas=WHSMouseHier.bitlas

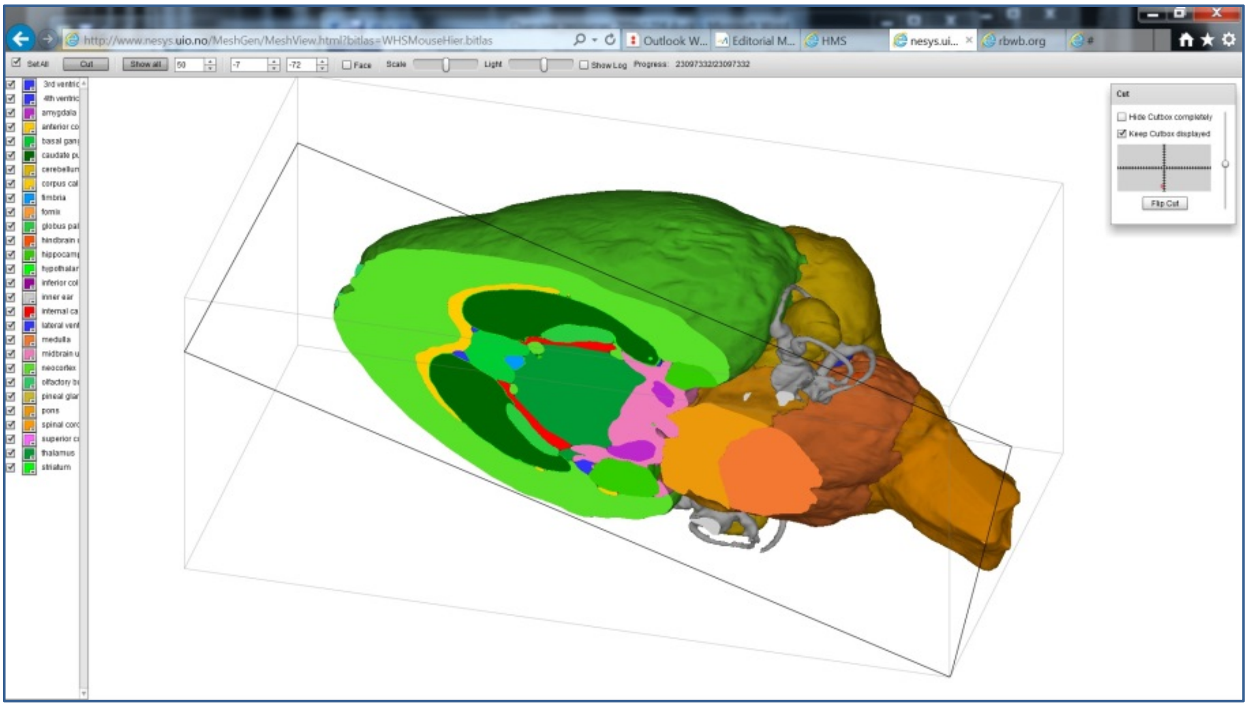
Allen Mouse Brain Atlas reference atlas

- (meshes for 106 anatomical structures including hierarchy) [ABAMouse]

- http://www.nesys.uio.no/MeshGen/MeshView.html?bitlas=ABAMouseHier.bitlas

Web application providing access to vector based versions of mouse and rat brain atlases in standardised 3D space for real-time viewing of shape, size and neighbouring relationships of brain structures, and viewing of boundaries of brain structures in any user defined cut plane.

The available mouse and rat brain volumetric atlases are presented through a browser-based viewer for real-time visualisation of mesh data. Since meshes representing adjacent structures typically intersect with each other, visual artefacts occur when rendering them. In order to ensure strict non-intersecting property of meshes, a custom mesh generator tool has been used to create the meshes based on the volumetric parcellations of structures. The tool generates a quad for each face of a voxel-cube which is facing a voxel and is not part of the structure, before applying a CESN-like (Constrained Elastic SurfaceNets) algorithm for smoothing the mesh. In the final release, it is the ai to support viewing of anchored 2D image data, and viewing of the locations of any data set.



3.3.3 Service Components

3.3.3.1.1 KSearch Service

Description: An essential service is intended to provide unified search indexes and APIs of Knowledge Graph and Collaboratory metadata across all services provided by the Collaboratory and Neuroinformatics Platforms.

API UI URL: https://nip.humanbrainproject.eu/api/ksearch/swagger/index.html

API URL: https://nip.humanbrainproject.eu/api/ksearch/

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#ksearch

Depends on Services: None

Current TRL: TRL5

SGA1 Target TRL: TRL8 (SLA defined and enforced by the API)

Further Details:

This service is intended to provide unified search indexes across all services provided by the Collaboratory and Neuroinformatics Platforms. This work will also define a standard API for populating the index. This service is currently in part on top of ElasticSearch. ElasticSearch is an open source horizontally scalable search index. See http://www.elasticsearch.org/ for more details.

See below for the architecture diagram which highlights planned developments.

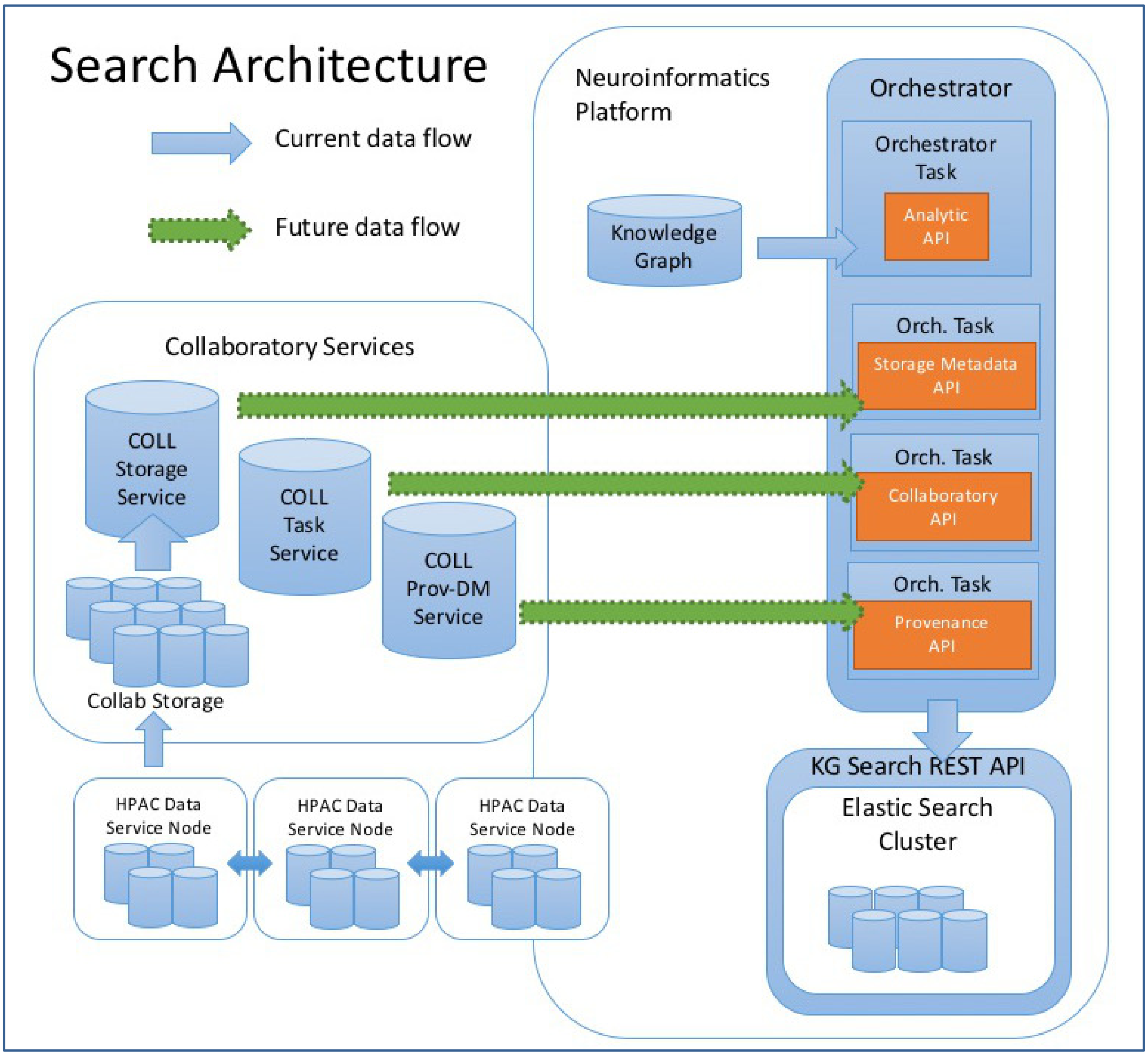


Figure 16: NIP Search Architecture

3.3.3.2 Knowledge Graph Service

Description: An essential service is intended to provide unified metadata service for experimental and modelling metadata necessary for data discoverability and provenance tracking. Currently access to this service is through the KSearch above.

API UI URL: https://nip.humanbrainproject.eu/api/ksearch/swagger/index.html

API URL: https://nip.humanbrainproject.eu/api/ksearch/

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#ksearch

Depends on Services: None

Current TRL: TRL5

SGA1 Target TRL: TRL8 (SLA defined and enforced by the API)

Further Details:

The Knowledge Graph developed in the HBP Ramp-up phase maintains experiment provenance data and is the primary source of truth for Neuroinformatics Platform search. As discussed in the Data Integration and Quality Management section one would like to be able to use the same search to find algorithmically produced datasets, models, simulations or post-simulation analyses. The experimental inputs form part of the context which would allow you to find the simulation and a key source of metadata on which to discover data and models in the NIP.

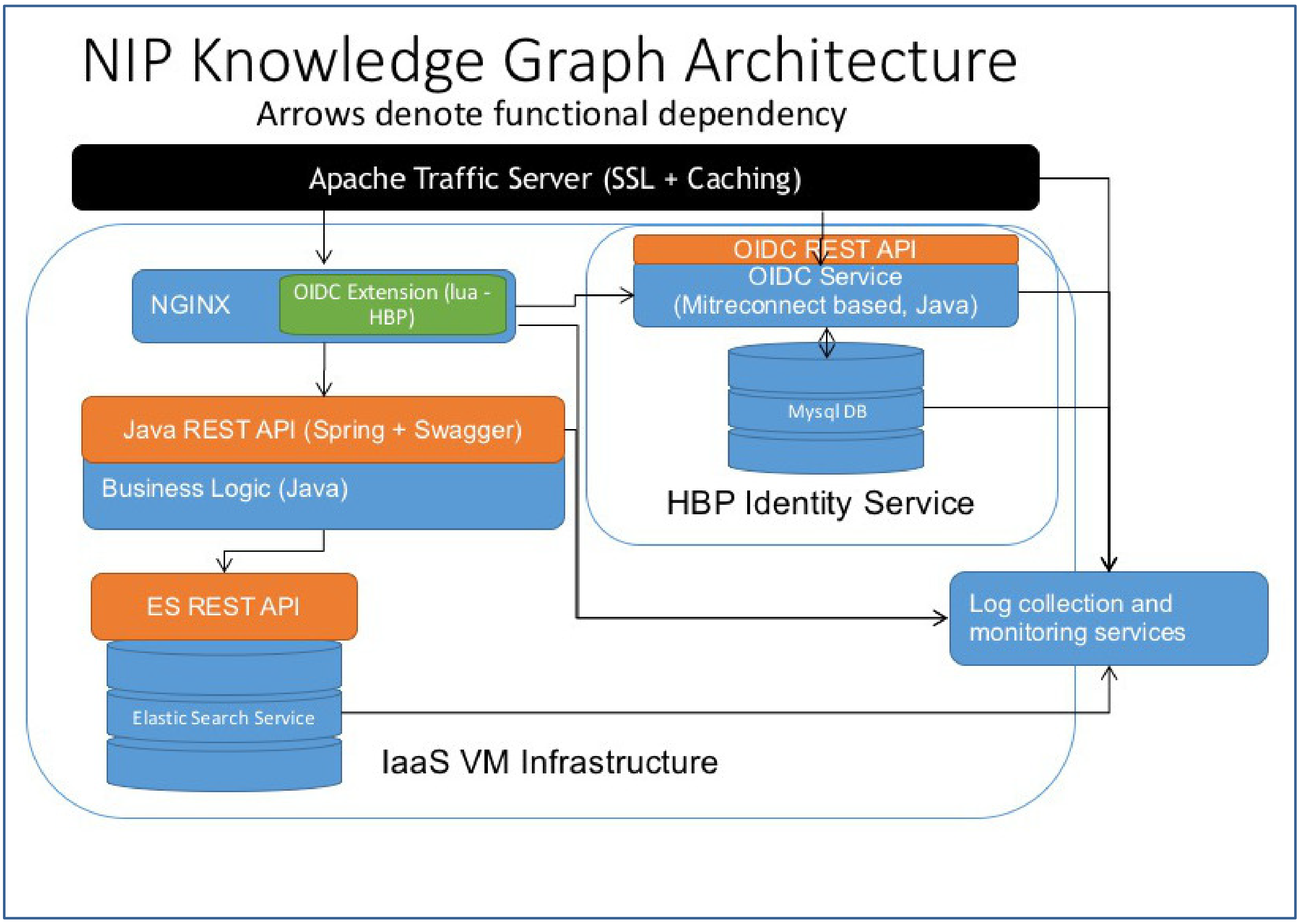


Figure 17: Current Knowledge Graph Service Architecture

Upcoming Developments:

A user might search for all models built with a particular class of mouse cells. This raises the problem of linking experimental data provenance to workflow provenance.

Under the current Knowledge Graph service, we would need to deploy one or more workflow provenance services. In doing so we will be creating an architectural debt which will be paid by each developer who wants to integrate with the provenance service.

This is the scenario described in Figure 18, below.

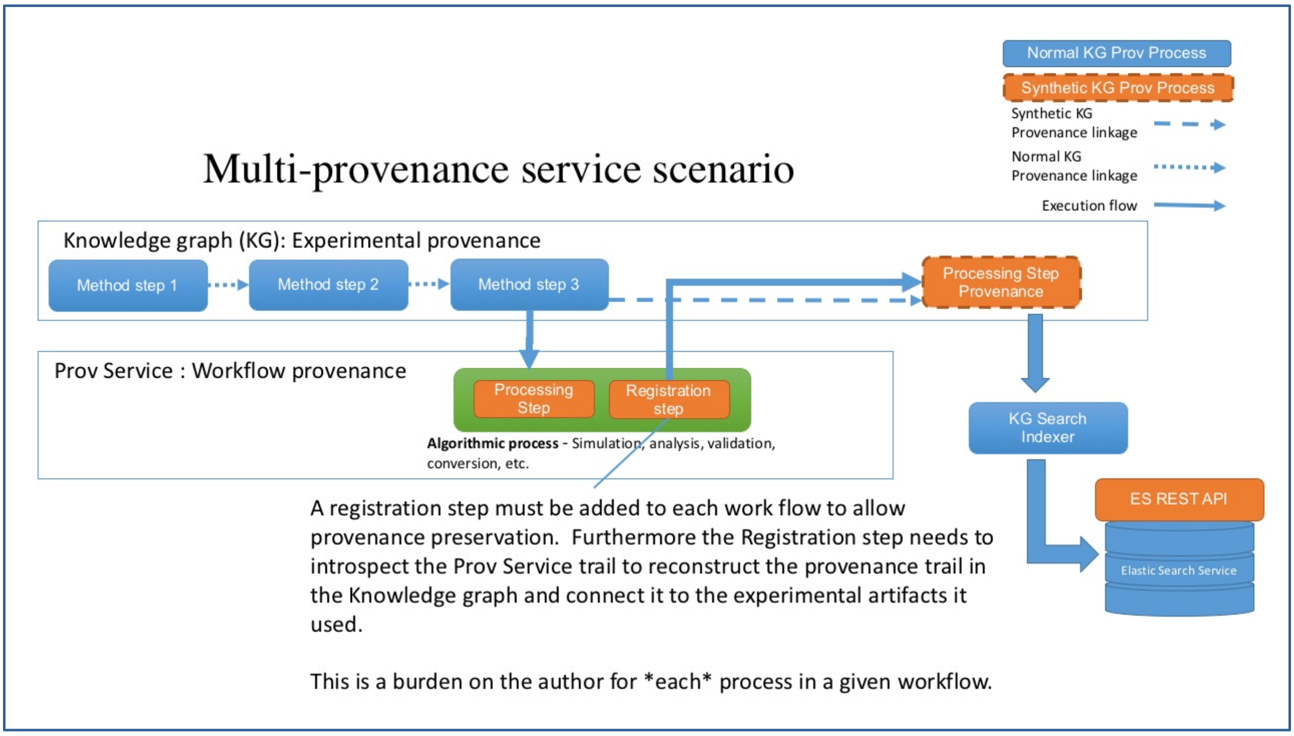


Figure 18: Multi-provenance service scenario workflow

To encourage adoption of provenance tracking as a standard practice for computational scientists we will instead work towards enhancing the Knowledge graph to enable it to function as a comprehensive and universal provenance system. The resulting scenario would be that shown in Figure 19, below.

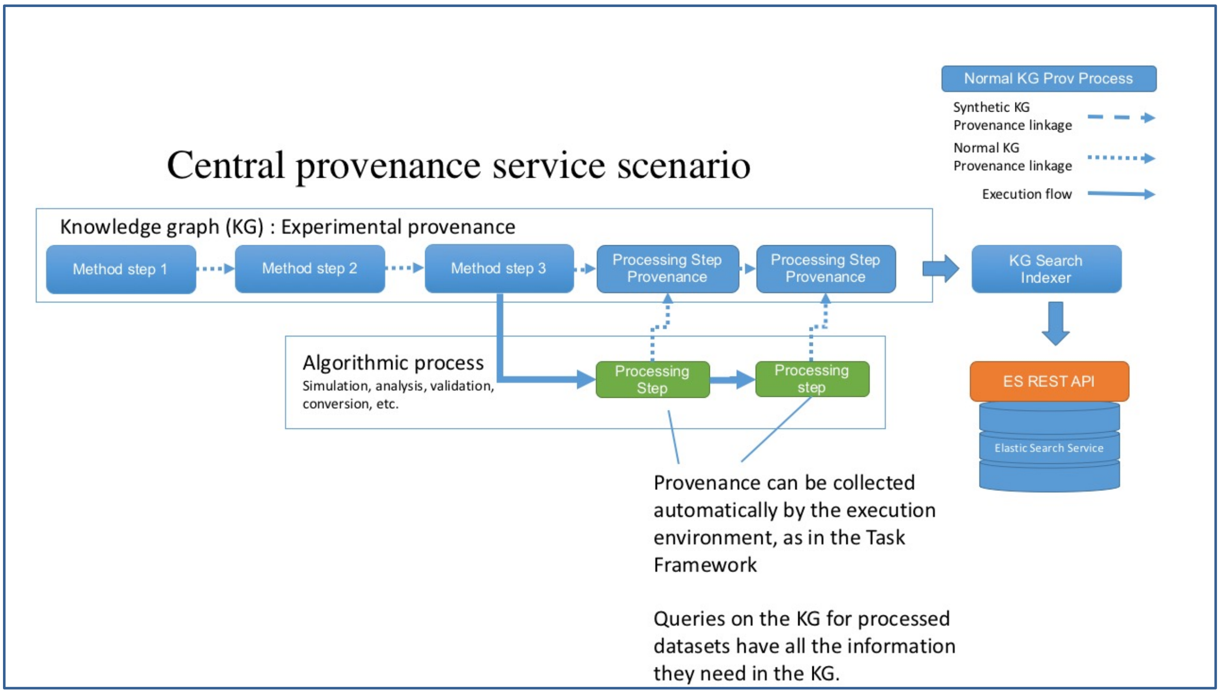


Figure 19: Enhancements to Knowledge graph to include provenance tracking

3.3.3.3 Image Service

Description: A

API URL: https://nip.humanbrainproject.eu/documentation/user-manual.html#image-service

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#image-service

Depends on Services: HBP Identity Service

Uses APIs: BBIC Python library

Current TRL: TRL5

SGA1 Target TRL: TRL8 (SLA defined and enforced by the API)

Further Details:

This Service provides data to the Atlas Viewer App for very large multi-species volumetric datasets. This is deployed to display multiple datasets of more than 1.25 teravoxels overlayed in the same view.

Atlases registered:

Template: Allen reference atlas v3, has aligned Allen ontological regions

Template: Waxholm rat brain atlas v2.0

- Waxholm rat brain atlas v2.0, 3-axis data, with mostly complete aligned Waxholm ontological regions
- Rat Golgi-Cox Stained, 3-axis data, not aligned to template (ie: Waxholm ontological regions are not aligned)

- Whole Brain Connectivity Atlas – R602, coronal data only, not aligned to template (ie: Waxholm ontological regions are not aligned)

Bigbrain, has white/grey matter parcellation

- Bigbrain, has white/grey matter parcellation
- Bigbrain with auditory cortex, no parcellation

MNI Colin 27, partial coverage of ontological regions

Infant atlas, 2 subjects, partial coverage of ontological regions

The current architecture is described below:

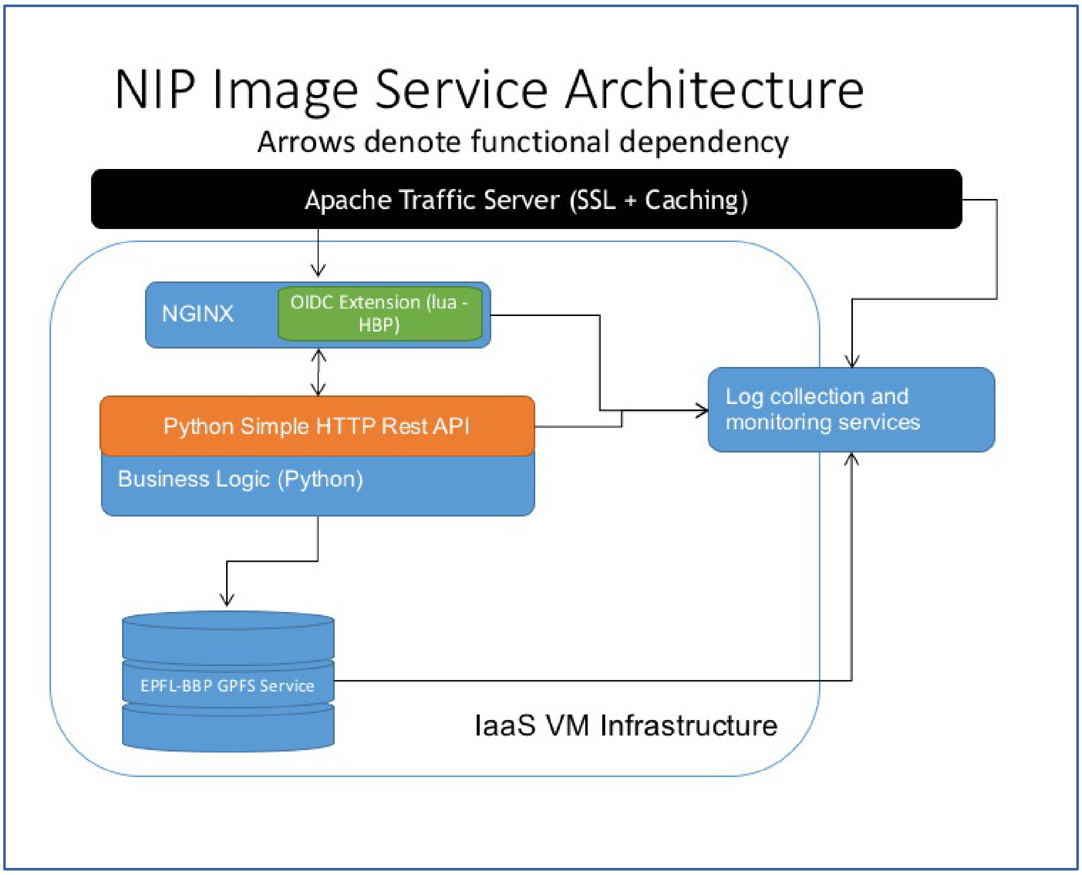


Figure 20: Image Service Architecture

Upcoming developments:

The current Image Service is an internal HBP product which, while very capable, is not based on standard image formats and does not leverage existing communities building similar tools. Since it was first developed in 2014, many community tools have advanced to the point where they are competitive.

As a result, the tasks involved in Atlasing apps and services will:

1) Evaluate the leading Image services in the community (eg: DVID, neurodata.io) for performance on target use cases, feature set and adoption by community. Select a single HBP Standard Image service based on this evaluation.
2) Work with SP7 to deploy the HBP Standard Image service with sufficient backing storage and attached compute capacity to SP7 data centres.

Formulate a work plan to integrate the atlas viewer system with the SGA1-planned ilastik web.

3.3.3.4 Ontology Service

Description: A read-only ontology resource managed by NIP developers for ontologies currently used throughout the NIP.

API URL: https://nip.humanbrainproject.eu/api/scigraph/docs/

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#scigraph-a-graph-based-ontology-service

Depends on Services: Github

Current TRL: TRL4

SGA1 Target TRL: TRL8 (SLA defined and enforced by the API)

Further Details:

See Documentation

3.3.3.5 VoxelBrain

API UI URL: https://nip.humanbrainproject.eu/api/analytics/swagger/index.html

API URL: https://nip.humanbrainproject.eu/api/analytics/

Development Standard: BBP Standard

Documentation: https://nip.humanbrainproject.eu/documentation/user-manual.html#voxel-brain

Depends on Services: None

Current TRL: TRL5

SGA1 Target TRL: TRL6

Further Details:

Further development plans for this component are still under consideration.

3.3.3.6 Elastic Search (Private)

Description: A scalable search index service. A critical architectural component which powers Knowledge Graph search.

Development Standard: BBP Standard

Documentation: https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html

Depends on Services: None

Current TRL: TRL5

SGA1 Target TRL: TRL8 (SLA defined and enforced by the API)

Further Details:

It is expected that Elastic Search will play a role in Knowledge Graph Search architecture for the foreseeable future. It is an industry standard opensource scalable search index developed in Java with client APIs in all major languages.

3.3.4 Software Components

3.3.4.1 HBP Knowledge Graph Python API

Description: A python API for Knowledge Graph and Knowledge Graph Search REST APIs

Development Standard: HBP Standard

Depends on Service: Knowledge Graph, Knowledge Graph Search

Current TRL: Doesn't exist

Target TRL: TRL8

Further Details:

Many of the add-on services like the registration UI will be built with Python. In addition, Python is well supported by existing Collaboratory Jupyter notebooks and other neuroscientific workflows. This work will aim to make integration of provenance registration into scientific workflows and web applications simple and straightforward to increase adoption of the HBP KnowledgeGraph inside and outside the HBP. This library will be open-sourced to facilitate community building.

3.3.4.2 QuickNII

Description: standalone tool for anchoring of 2D experimental image data to 3D atlas templates

Current URL: http://www.nesys.uio.no/AligNII/

Development Standard: UIO Standard

Documentation: http://www.nesys.uio.no/AligNII/

Depends on Services: None

Current TRL: TRL7 (SLA not defined or monitored)

SGA1 Target TRL: TRL7 (SLA defined and monitored)

Further Details:

Stand alone tool. Internal HBP release June 2015.

Tutorial with video clips demonstrating AligNII applies also for QuickNII (difference being near real-time operation with QuickNII as compared to the network dependent AligNII version): http://www.nesys.uio.no/AligNII

Standalone iteration of AligNII, developed to achieve near real-time operation of anchoring procedures, as compared to the network dependent AligNII version. QuickNII delivers user guided registration of 2D image data to volumetric atlas templates. With use of the tool, data are anchored to Waxholm Space (mouse and rat) or Allen Brain atlas space (mouse), facilitating data integration through standardised coordinate systems.

QuickNII is a standalone iteration of AligNII where the volumetric templates for the rat brain (MRI, DTI and structural parcellations for the Waxholm Space rat atlas) and mouse brain (volumetric histology and structural parcellations for the Allen Brain mouse atlas) are brought to the user's computer. This practice eliminates the network roundtrip, and thus results in near real-time operation with high refresh rate of the navigational and user-defined cut plane renderings. Other aspects of user experience are preserved (as outlined for AligNII), with the same controls, views and general look-and-feel. While QuickNII is a standalone tool, steps are taken in order to integrate it in the workflow provided by the Navigator data system. Current version of QuickNII is prepared to run in 64-bit Microsoft Windows environment, 3 Gigabytes RAM is required. The tool itself comes as a 620 Megabytes downloadable ZIP package, requiring 800 Megabytes of disk space after extraction (the package contains both the Waxholm Space rat brain and the Allen Brain mouse brain templates in a pre-processed format). The tool combines Adobe AIR and 64-bit Java technologies. Both runtimes are part of the package and no installation or any kind of privileged user rights are required for use. As both Adobe AIR and Java are platform-independent technologies, supporting other operating systems is mainly a question of re-packing.

3.4 HBP-NIP: Use Cases

3.4.1 Roles

- Computational Scientific User (CSU) – A User with scientific development skills and comfort in launching command line HPC jobs.
- Biological Scientific User (BSU) – A User with scientific expertise, but limited technical computing skills.
- Scientific User (SU) – A scientific User, either a CSU or a BSU.
- Scientific Developer (SCIDEV) – A User who is developing software to directly realise the scientific objectives. This User is usually working in close collaboration with scientists, both CSUs and BSUs.
- Developer (DEV) – A User who is developing software to realise engineering, operational and/or scientific objectives.
- Portal User (PU) – A User who accesses Platform functions through the Web GUI.
- Service User (ServU) – A User who accesses Platform functions through a programmatic Service Client API.
- Infrastructure Personnel (INFRA) - An infrastructure System Administrator or Developer, typically responsible for deploying and monitoring Platform services that are offered directly to customers.

3.4.2 Interactive Atlas Exploration (HBPNIP-UC-001)

Bill wants to explore the data that are available for the mouse. He intends to use this in later analyses, but to do so he must first know what is available. This is done through the 2D or 3D atlas viewer of the NIP integrated into the COLL.

Primary Actors: Scientific User, Bill.

Preconditions:

- One or more atlases for the mouse have been registered in the Knowledge Space.

Success Scenario:

1) Bill wants to explore available data for the mouse.

2) In the COLL, Bill would select the Mouse atlas, which would open the 3D viewer in the Atlas viewer mode with an initial context of Mouse.
3) From the initial context, Bill would be able to view a large number of data types in the spatial context of the default Mouse, Rat or Human reference space:
   a) 2D image slices
   b) 2D parcellations
   c) 3D volumes
   d) 3D parcellations
   e) 3D objects (geometry: meshes, lines, etc.)
   f) Registered data links
4) Data can be added to a COLL Project or will be added to the current COLL Project automatically if an analysis Task is run on the data.

3.4.3 Finding and Selecting Data through Direct Search (HBPNIP-UC-002)

This Use Case describes the interaction with the Search App for the pre-selection of data that are later used in the configuration of a Job prior to execution.

Primary Actor: Scientific User Abigail.

Success Scenario:

1) Abigail wants to extract improved dendrite synthesis parameters to feed to a prototype neuron synthesis program.
2) Abigail creates a new project, COLL Project 3, to hold her input data and the synthesis parameters she wants to generate.
3) Her current focus is Layer 5, so Abigail is wants to find reconstructed morphologies from Layer 5 of the mouse somatosensory cortex. She selects the Search App to execute a search against the NIP. She enters "mouse morphologies 'Layer 5'" and executes the search.
4) The search of the NIP returns a list of results that will be displayed in the Search App.
5) The search results will look similar to a Google search results list.
6) The search results can be exported to the 2D Atlas viewer to view the search results anchored to the relevant spatial or sematic-spatial location in the current slice of the 2D atlas viewer.
7) The search results can be exported to the 3D Atlas viewer to view the search results anchored to the relevant spatial or sematic-spatial location in the current slice of the 2D atlas viewer. Results clustering may be required if too many results share the same anchor location in the viewer.
8) Abigail selects the morphologies she thinks will provide a good source for synthesis parameters and links them to COLL Project 3.
9) Abigail then selects a morphology or collection of morphologies to run her preferred synthesis parameter extraction analysis on. She executes the analysis and the extracted synthesis parameters can then be added to the current COLL Project.

3.4.4 Uploading experimental data with metadata (HBPNIP-UC-003)

This Use Case describes the interaction with the Data Workbench Web App for the upload and annotation with metadata.

Primary Actor: Scientific User Abigail.

Success Scenario:

1) Abigail wants to upload data and annotate it with high level meta-data.
2) Abigail logs into the Data Workbench (DWB).
3) Abigail uploads data to one of the HBP Storage Services, possibly adding the data to a previously saved dataset collection.
4) Abigail selects the granularity she wants to start her annotation work with (single dataset or dataset collection).
5) Abigail ensures that an exemplar image/media clip for the dataset has been selected.
6) Abigail selects the protocol used to generate the dataset, or adds a new protocol via the DWB interface.
7) Abigail adds ontological metadata to the dataset adding new terms to the ontology, via the DBW, if necessary. While she does so she checks the "meta data completeness score" and "search return ranking score" of the dataset.

3.4.5 Identify and tag a parameter with metadata (HBPNIP-UC-004)

This Use Case describes the interaction with the Parameter Workbench Web App for the discovering and annotating a parameter.

Primary Actor: Scientific User Abigail.

Success Scenario:

1) While reading a publication Abigail finds a parameter she thinks will be useful for later modelling/analytics work.
2) Abigail logs into the Parameter Workbench (PWB).
3) Abigail uses the search function in PWB to see if somebody feed the parameter already into the system.
4) Abigail selects a search result and gets redirected to a parameter page. She checks the information in the system and determines that extra meta data can be added to increase discoverability of the parameter.
5) Abigail adds extra meta data to the parameter and leaves a comment on the discussion section of the parameter to ensure other users understand why the extra meta data was added.

3.4.6 Create a brain atlas (HBPNIP-UC-005)

Primary Actor: One Biological Scientific User Daniel

Success Scenario:

- Daniel would like to create a brain atlas from a stack of high resolution 2D images located on his computer, and share the brain atlas online.
- Daniel creates a project – Project1. He is now the owner of Project1.

- Daniel uploads 2D images to Project1.
- Daniel is requested to enter the metadata (MINDS).
- Daniel starts the converter via Web GUI. The converter creates pyramids, previews and thumbnails from the 2D images. This might be a time consuming task, so Daniel might log out and log back in later to check the progress.
- The conversion is finished. Now the following extra tools and features are automatically available via unique URLs for Daniel only:
  - Original 2D images, pyramids, thumbnails and previews.
  - 2D image viewers for individual 2D images in full resolution (zoom and pan).
  - 2D filmstrip viewers for a set of 2D images in full resolution (select image, zoom and pan).
  - Annotation editor for individual 2D images in full resolution (draw an annotation, zoom and pan).
- Via Web GUI Daniel selects which 3D template package he wants to anchor to and the 2D image to anchor.
- Daniel uses AligNII, allowing a step-wise process to anchor the 2D images to the 3D template package. The steps include:
  - Identification of the orientation of the 2D image stack in relation to the 3D template package.
  - Global scaling and positioning of the individual images (3D best fit) to the 3D template package.
  - Local positioning (region–by–region) of regions of interest.
- Daniel saves the anchoring results.
- The 2D images are now searchable using metadata and/or 3D spatial requests, and can be viewed individually, in filmstrips or in 3D viewer.
- Daniel can add other users to Project1 and give them read or read-write access OR make Project1 publicly (anonymously) read-only available. The tools inherit the access rights.
- Optional: the tools, original 2D images, pyramids, thumbnails and previews can be embedded into web projects.

3.4.7 Search by 3D spatial location or metadata (HBPNIP-UC-006)

Primary Actor: One Biological Scientific User John

Success scenario:

- John would like to find all available brain atlases and/or 2D images for a particular 3D location and/or a particular set of metadata.
- John logs in to COLL, defines the criteria and runs the search.
- The results can be displayed individually, in filmstrips, or in 3D viewer.

3.4.8 Annotate brain atlas (HBPNIP-UC-007)

Primary Actors: Two Biological Scientific Users: Bill and Daniel

Success Scenario:

- Bill would like to highlight certain areas on 2D images of the brain atlas created by Daniel (owner) as Project1.
- Daniel gives read-write access to Bill for Project1.
- Bill selects a 2D image and opens Annotation editor for it.
- Bill draws an annotation, saves it and repeats the process for every 2D image he wants to annotate.
- The 2D images with the annotations can be displayed individually or in filmstrips. The annotations can also be displayed in 3D viewer as 3D meshes. The annotations are searchable.

3.5 HBP-NIP: Functional Requirements

3.5.1 Authentication and Authorisation (HBPNIP-FR-001)

1) Services and apps with privacy considerations must be authenticated with the HBP Identity Service
2) Users must have access control based on COLL Project-specific groups of Users.
3) The Portal and its applications will be accessible only through the SSL protected https:// or wss:// protocols.
4) Single sign-on will allow a User to log in once for most applications.
5) Certain applications may require re-authentication to perform some privileged operations.
6) Access to data in COLL Projects, Artefacts, Parameters and their metadata will be controlled by the Collab authorisation system.

3.5.2 Search (HBPNIP-FR-002)

1) Artefacts uploaded to the NIP or COLL will be searchable by diverse metadata.
2) Artefacts produced to the NIP or COLL will be searchable by diverse metadata.
3) Artefacts in the NIP will be searchable across the COLL and NIP.
4) Tasks and workflows will be searchable by diverse metadata.
5) Parameter contents will be searchable.
6) COLL Project wiki descriptions will be full-text searchable.
7) The search service must provide a common search API for searching NIP and COLL metadata indexes.
8) The search service must return the first page of search results in less than five seconds. Since it will take a series of incrementally refined searches to find a User’s desired data, searches taking longer than five seconds to return the first page of 10-25 results will greatly reduce search utility.
9) Search queries for COLL data in collabs will only return results if the user performing the query has read access to the collab, to respect the intention behind the authorisation model for COLL collabs.

3.5.3 Common Service Interfaces (HBPNIP-FR-003)

1) All services must use the HBP Identity service as their source of truth for providing User metadata and for authenticating Users. This authentication can be exposed through one of the authentication mechanisms described in Sections 2.3.2 and 2.3.3.
2) All services — web or otherwise — must log their system logs through the system-wide syslog Logging Service.
3) All Services must provide a method to determine service health for the system-wide Monitoring Service.

3.5.4 Web Services (HBPNIP-FR-004)

1) Web Services will offer a REST binding using JSON.
2) REST services must conform to Platform standards.
3) REST services must provide web accessible documentation in a standard location.
4) REST services must have Platform-provided client libraries for Python.
5) Access to REST services must be authenticated via OpenID-Connect, which uses HBP LDAP as its source of authentication authority.
6) Web Service Interfaces must log API accesses. Ideally this will provide per-User accounting.

3.5.5 Non-Web Services (HBPNIP-FR-005)

1) Public facing non-Web Service interfaces must be authenticated with one of OpenSSH keys, Kerberos or X509 client certificates.

3.5.6 Atlas Viewer (HBPNIP-FR-006)

1) The Atlas Viewer must be usable as a viewer of an arbitrary collection of Artefacts. Sources of Artefact collections include COLL Projects and collections of COLL and NIP search results.
2) Each Artefact in a given collection will be displayable in a 3D reference brain viewer, along with its spatial location, if this is available. If the spatial information is not available, the Artefact will be displayable anchored to its most informative semantic spatial location. If neither is available, it will be displayable without a brain viewer anchor.
3) Each Artefact in a given collection will be displayable in a 2D reference brain viewer, along with its spatial information, if available. If the spatial information is not available, the Artefact will be displayable anchored to its most informative semantic spatial location. If neither is available, it will be displayable without a brain viewer anchor.
4) The Atlas Viewer must be integrated with the Neuroinformatics search such that search results are treated as a collection of Artefacts.

3.5.7 Atlas Viewer Data Management (HBPNIP-FR-007)

Image and geometry data download from the server to the browser. Compared to other Platform components, the amount of data downloaded to the browser will be large.

1) The download of various data elements from the atlases must be done in an asynchronous manner to ensure that the UI remains responsive.

2) The download process will be visible to the User. It must be possible for the User to cancel the download of a data set.

3.5.8 Atlas Viewer: View Modes (HBPNIP-FR-008)

The Atlas Viewer will have the following modes, each of them automatically selected based on application context:

1) List View
2) Tile View
3) 2D View
4) 3D View
5) Neuron selection interface.

3.5.9 Image Service (HBPNIP-FR-009)

1) The imagery service must provide bandwidth-efficient query and image download capabilities to allow use of the imagery service by a browser with a broadband internet class connection (10Mbit/s).
2) Contents of imagery services need to be provenance-friendly. This strongly implies identifying URL specifiers for image queries and immutable underlying data.
3) The COLL will need a way to register imagery services.
4) The COLL will need to perform health checks on externally registered imagery services. This could be done at least in part by an interface to the HPAC Platform.

3.5.10 Reference Space Transformations (HBPNIP-FR-010)

It is expected that the various atlases will not share a common coordinate space. Consequently, the Atlas Viewer will need a mechanism to view data from multiple coordinate spaces.

1) The Atlas Viewer and its associated Data and Image Services are expected to support limited collection of HBP Spatial Reference Systems. Data providers will be required to provide their data registered in a supported HBP SRS coordinate space.
2) A Reference implementation of a Spatial Reference transformation system exists, either in service or library form.

3.5.11 Performance: Image Processing and Atlas Building (HBPNIP-FR-011)

One of the most performance-sensitive areas in the NIP is in the building, analysis and viewing of atlases as required by CDP1, CDP2 and CDP3. There are a large number of high-quality high-performance image processing libraries which will be leveraged to allow SP5 to meet these needs. The supporting data and compute architecture must be able to handle a number of SMALL-, MEDIUM- and LARGE-scale data use cases where LARGE would be defined as follows:

- SMALL: O (10MB) processing and/or transfer
- MEDIUM: O (~1 TB) processing and/or transfer
  - Process the entire BigBrain dataset
  - Eg: ~2.5 hrs to transfer on a 1GBps link

- LARGE: O (10PB) 1 um^3 x 16bit voxel whole brain => ~8 PB
  - Eg:~83 days to transfer on a 10GBps link

We can the envision the following transformation, processing or analysis scenarios:

1) Slice images => slice image(s) (e.g.: artefact removal)
2) Slice sets => voxel volume(s) (e.g.: inter-slice alignment)
3) Slice sets => discrete extracted features (e.g.: ontologically linked neuron or glia geometry)
4) Voxel volume => slice image(s) (e.g.: plane visualisation, cut-plane alignment or format conversion)
   a) Scale: MEDIUM – many of the use cases for this transformation are interactive explorations and will need to scale for 30+ concurrent users on a single service instance (based on CDP3 requirements).
5) Voxel volume => voxel volume(s) (e.g.: atlas alignment)
   a) Scale: LARGE - Inter-template warpings are not likely to be calculated on full resolution volumes, but they will be applied on full resolution volumes. This will be a challenge to solve in collaboration with SP7.
6) Voxel volume => discrete feature(s) (e.g.: ontologically linked neuron or glia geometry)
   a) Scale: SMALL-MEDIUM – large image volumes will need to be traversed to extract particular geometries, such as long range connectivities. Similar problems have been addressed by existing software packages (https://github.com/neurodata/ndstore publication: http://arxiv.org/abs/1306.3543)
7) Voxel time series alignment
   a) Scale: LARGE - Time-series voxel data sets will be much smaller in size than a 1um^3 human brain atlas, but correlational analysis across time scales will be extremely compute and IO intensive (e.g.: Thunder https://github.com/thunder-project/thunder)

In addition, there is a need to allow scalable visualisation systems to visualise the above source and destination datasets in near real-time. This will also allow analyses above to operate on scaled versions of high-resolution datasets if deemed acceptable for the analysis in question.

- High-performance, low latency pyramided multimodal voxel static and time series data read.

The wide range of image operations and imaging modalities present in the HBP give rise to the following requirements:

- In many cases there will be a benefit in performing these operations on a lower detail image or volume than the full dataset. Support for image pyramids will be critical to high-performance data processing operations.
- The underlying storage method should be based on mature formats and be file- or object store-based, rather than database-based. This approach is taken due to:
  - Simplicity - adapting image-processing tools is significantly simpler. Deployment of systems consuming atlasing data directly will be simpler.
  - This will have a simpler path to scaling for parallel cluster or map-reduce style analyses (see Thunder https://github.com/thunder-project/thunder)

To achieve economies of scale the work of WP5.2 and WP5.3 under CDP1-C4 must establish a collection of image format standards. There should be a collection of standards rather than a single standard because different use cases benefit from different organisations of the underlying image data. The same task will need to work to establish robust conversion workflows between the collected standards. At the very least there are standards needed for:

- initial data collection – likely a collection of well supported standards: NIFTI, DICOM, Tiff stacks and MINC format.
- pyramided multimodal voxel static and time series data. A number of candidates exist here, including current NIP BBIC format and MINC2. These should be evaluated in collaboration with visualisation teams in SP7.

4. Common Architecture

4.1 Software Development

In Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) tiers of the service hierarchy, software is a crucial component. All research activities in the HBP are heavy consumers and producers of complex software packages. A robust approach to software quality will be a critical factor in the success of the HBP.

Drawing on the experience of other large-scale software projects, HBP identified the need to invest in software maturity early on.

These investments have continued and expanded to include the following best practices:

- Unit testing, typically with minimum test coverages (eg: 90% or higher on all Collaboratory Team developments)
- Automated Integration testing suites (automated post-deployment system testing)
- Continuous integration
- Ticket management for project feature planning
- Scrum training for Agile software development (either with an agile coach or and HBP practitioner)
- Coding standards
- Utilisation of Code review by Developer teams.

Bringing all these pieces together results in discipline called Continuous Delivery. All teams in HBP are using a variant of this approach with slightly different mixes of tools. These will be described in more detail in the upcoming D11.2.2 – HBP Software Engineering and Quality Assurance Approach delivered at SGA1- Month 10.

4.2 Operations Standards

4.2.1 Overview

For SaaS, PaaS and IaaS tiers the reliable and repeatable deployment software is a crucial component of the overall operations strategy. Monitoring of services will be based on best practices from industry along with crucial expertise from HPAC providers in the HBP and Federated infrastructure providers throughout the EU.

HPC has well-defined deployment and monitoring practices which are informed by a long history of service provision. For the other Platform teams, no such history exists so best practices have been adopted from Agile software development. Platform teams in the HBP have been adapting these techniques and tools according to their team needs.

To this end a DevOps model is in the process of being pragmatically adopted across HBP:

- Deployment lifecycle w/ dev, staging (optional) and production environments for all services.
- Continuous integration – unit testing, integration testing, repeatable software builds and package releases
- VM configuration development – source control and code review of configuration changes. Requires a programmable configuration system.
- VM configuration management – associate service configurations with specific VM resources.
- VM configuration deployment – deploy approved changes through an automated system.
- Object Storage – highly available, redundant storage for VMs and service data.
- Internet Gateway w/ Caching Proxy Server – services typically are not available directly via the internet. Best practice places a tuned caching proxy server between application servers and the open internet for reliability, flexibility and performance reasons.

4.2.2 HBP Standard DevOps Stack

For Platforms moving towards serving HBP-wide needs, infrastructure must be soundly governed by an organisation with the mandate to provide services for and on behalf of HBP. The planned HBP Standard DevOps Stack serves as the basis for development, deployment, configuration and monitoring of the following Platforms:

- Collaboratory
- SP5 – Neuroinformatics

This DevOps Stack will run on an HBP Managed tier physical Infrastructure running on commercial Cloud Providers and HPAC Platform sites. There will be an evaluation process for each service to determine whether it should be deployed to a commercial Cloud Provider or to an HPAC site. This will take into account the inherent advantages of the respective solutions, with cost on one hand and data locality and massive compute capabilities on the other hand.

Table 2: HBP Standard DevOps stack components

Service Category	Service	Provided by	Notes
Dev hosts	Cloud vendor – TBD	EPFL-HBPPCO procurement	
Staging hosts	Cloud vendor – TBD	EPFL-HBPPCO procurement	Optional environment
Production hosts	Cloud vendor and HPAC Platform	EPFL-HBPPCO procurement	
Continuous integration	Travis-CI	EPFL-HBPPCO procurement	
VM configuration development	Github and/or Phabricator	EPFL-HBPPCO	
VM configuration management	TBD	EPFL-HBPPCO	
VM configuration deployment	Puppet or Ansible - TBD	EPFL-HBPPCO	
Object Storage	Cloud vendor and FENIX	EPFL-HBPPCO	
Internet Gateway w/ Caching Proxy server	NGINX	EPFL-HBPPCO	
Monitoring	TBD	EPFL-HBPPCO	

4.2.2.1 HBP Standard Service deployment

The COLL and NIP will follow the HBP Standard development and deployment model, as described in the HBP System Engineering documentation and represented in Figure 21.

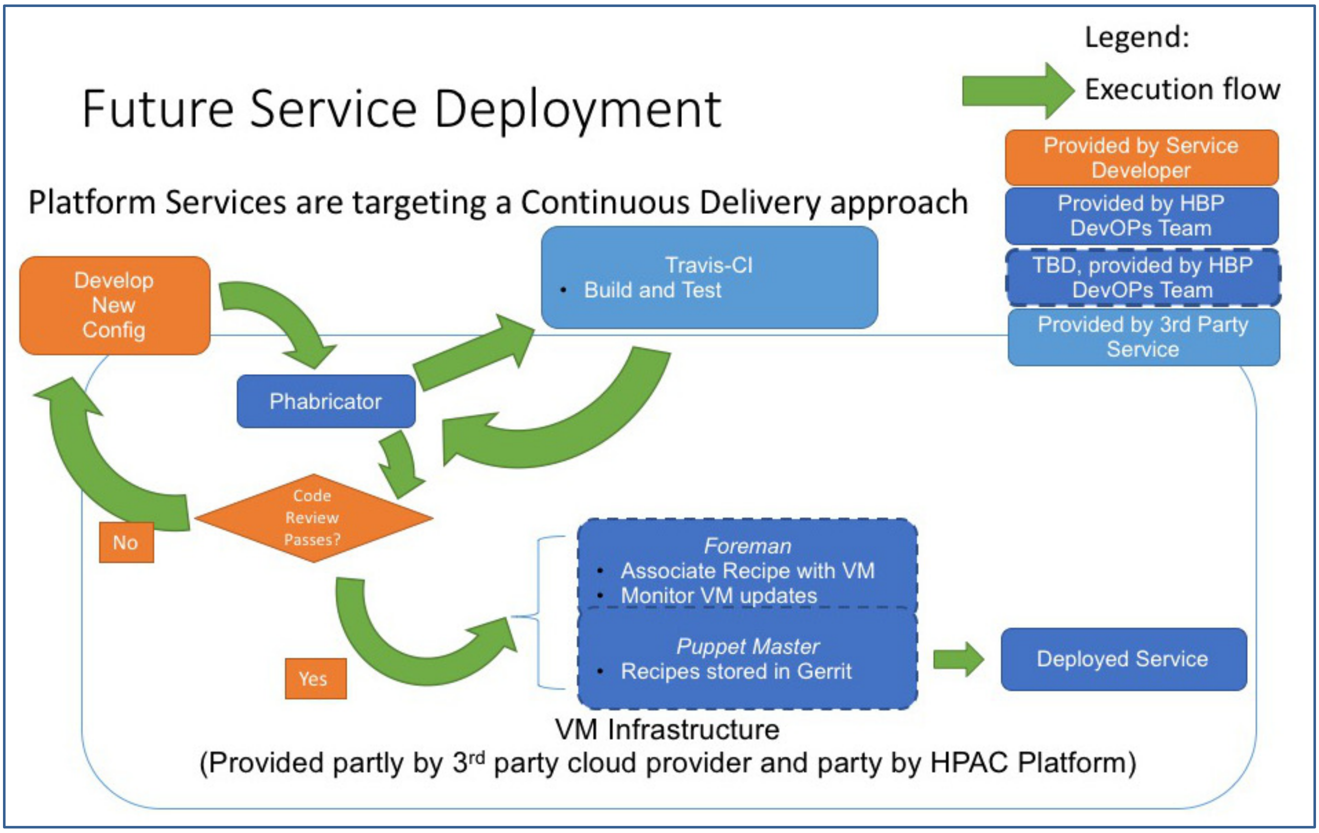


Figure 21: Planned HBP Standard Continuous development and deployment

This process makes extensive use of the following open source tools:

Git – a leading distributed Version Control system

Travis-CI – a leading Java-based Continous Integration (CI) system: https://travis-ci.org/

Phabricator – a leading code forge site with extensive project management feature set and code-review system change review system: https://www.gerritcodereview.com/

Puppet – a leading *devops* system configuration system: https://puppet.com/

Ansible – a serverless *devops* system automation framework: https://www.ansible.com/

4.2.3 Legacy Standard - BBP Standard DevOps Stack

For completeness, it important to talk about where many of these ideas were successfully piloted. In the RUP, the most widely used of the DevOps infrastructures, BBP Standard, was developed by various teams at the BBP and serves as the basis for development, deployment, configuration and monitoring of the Ramp-Up releases of the following Platforms:

- Collaboratory
- SP5 – Neuroinformatics
- SP6 – Brain Simulation Platform
- SP10 – Neurorobotics Platform

Table 3: BBP Standard DevOps stack components

Service Category	Service	Provided by	Notes
Dev hosts	Openstack	BBP Core Services	
Staging hosts	Openstack	BBP Core Services	Optional environment
Production hosts	VMWare	BBP Core Services	
Continuous integration	Jenkins	BBP Core Services	Jenkins uses Openstack VMs for jobs
VM configuration development	Git + Gerrit	BBP Core Services	
VM configuration management	Foreman	BBP Core Services	Integrates with VMware and Openstack
VM configuration deployment	Puppet	BBP Core Services	
Object Storage	Ceph	BBP Core Services	Used by both Openstack VMs and Collaboratory Storage Service
Internet Gateway w/ Caching Proxy server	Apache Traffic Server	BBP Core Services	
Monitoring	Icinga, Grafana, Kibana, syslogd, collectd	BBP Core Services	

4.2.3.1 Legacy BBP Standard Service deployment

The COLL and NIP currently follow the BBP Standard development and deployment model, as described in the HBP System Engineering documentation and represented in the figure below.

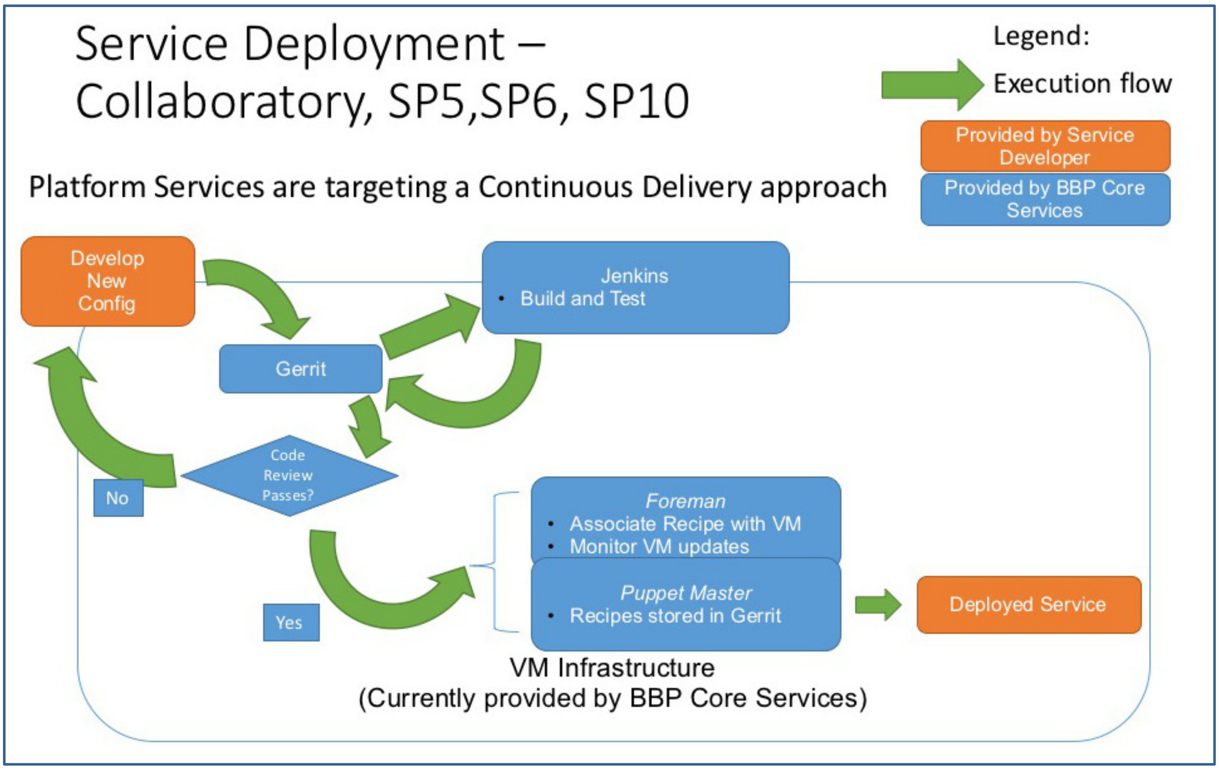


Figure 22: BBP Standard Continuous development and deployment

This process makes extensive use of the following open source tools:

Git – a leading distributed Version Control system

Jenkins – a leading Java-based Continous Integration (CI) system: https://jenkins.io/

Gerrit – a leading Git and Java-based source code change review system: https://www.gerritcodereview.com/

Puppet – a leading *devops* system configuration system: https://puppet.com/

Foreman – associates Puppet recipes with particular hosts: http://theforeman.org/

5. Extensibility and Platform Integration

5.1 Extensibility

As described in the HBP-COLL: Architecture section above, the HBP Collaboratory is designed from the ground-up to be extensible at all levels of the architecture. This has significant advantages. First, the HBP Collaboratory can be seen as self-extend to add each new feature addition. Secondly, it allows a tested extension model which can allow HBP Platform developers and third-party developers can use to add functionality to the Collaboratory. Finally, it requires HBP Collaboratory services and modules to be swapped out with alternative implementations readily as requirements change.

SP7 plays a key role in powering the Collaboratory and the HBP infrastructure architecture and should be considered a horizontal service provide both directly and indirectly through various HBP-COLL services. The details of this architecture are described in D5.6.2 “IT Architecture of the HBP Integrated System of Platforms”. Further details on the specifics of HBP-COLL extension can be found in the HBP-COLL developers guide, an attachment to this document.

For the remaining platforms, SP6, SP8, SP9 and SP10 have made heavy use of the Collaboratory to develop their solution verticals. The details of these interactions are described in the sections below.

5.2 Brain Simulation Platform

The Brain Simulation Platform (BSP) makes heavy use of python based Jupyter notebooks to document usage of their tools. These notebooks are offered through the HBP-COLL to allow users of the Platform to easily reproduce and customise model building, simulation and analysis task in an interactive fashion.

For less technically adept users, the BSP team is also building a collection of purpose-built web-based applications to increase the ease of use of certain workflows. These applications are largely build on the same foundations as the python notebooks above, but have been wrapped in web-based GUIs to ease understanding, configuration and execution of the workflows.

The two figures below describe the integration points between these two application paradigms and the HBP infrastructure. Case 1 describes the Jupyter notebook integration pattern and Case 2 describes the Collab Application integration pattern. It should be further noted that these two cases share similarities with the integration patterns of other HBP Platforms.

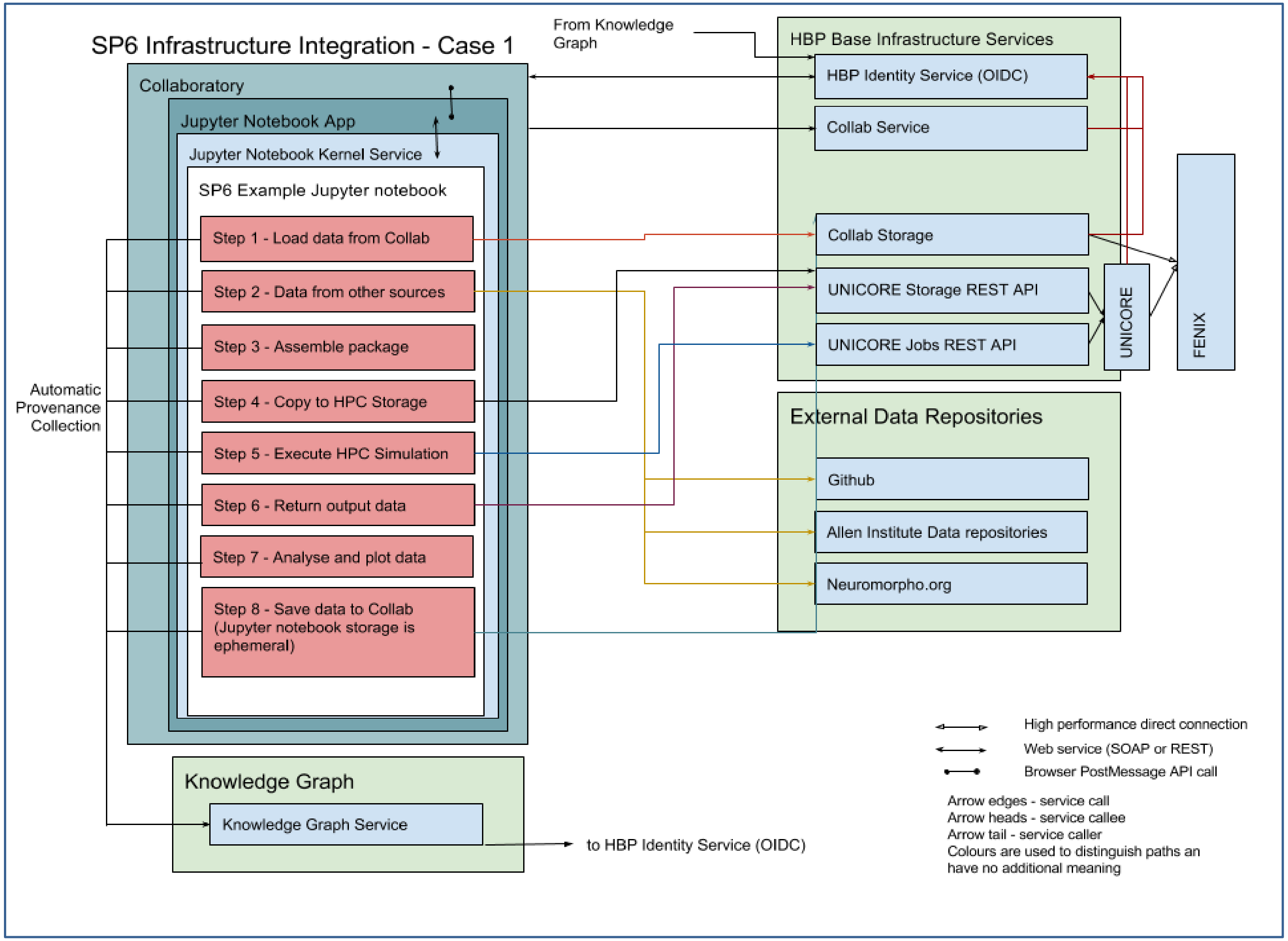


Figure 23: SP6 Infrastructure Integration - Case 1

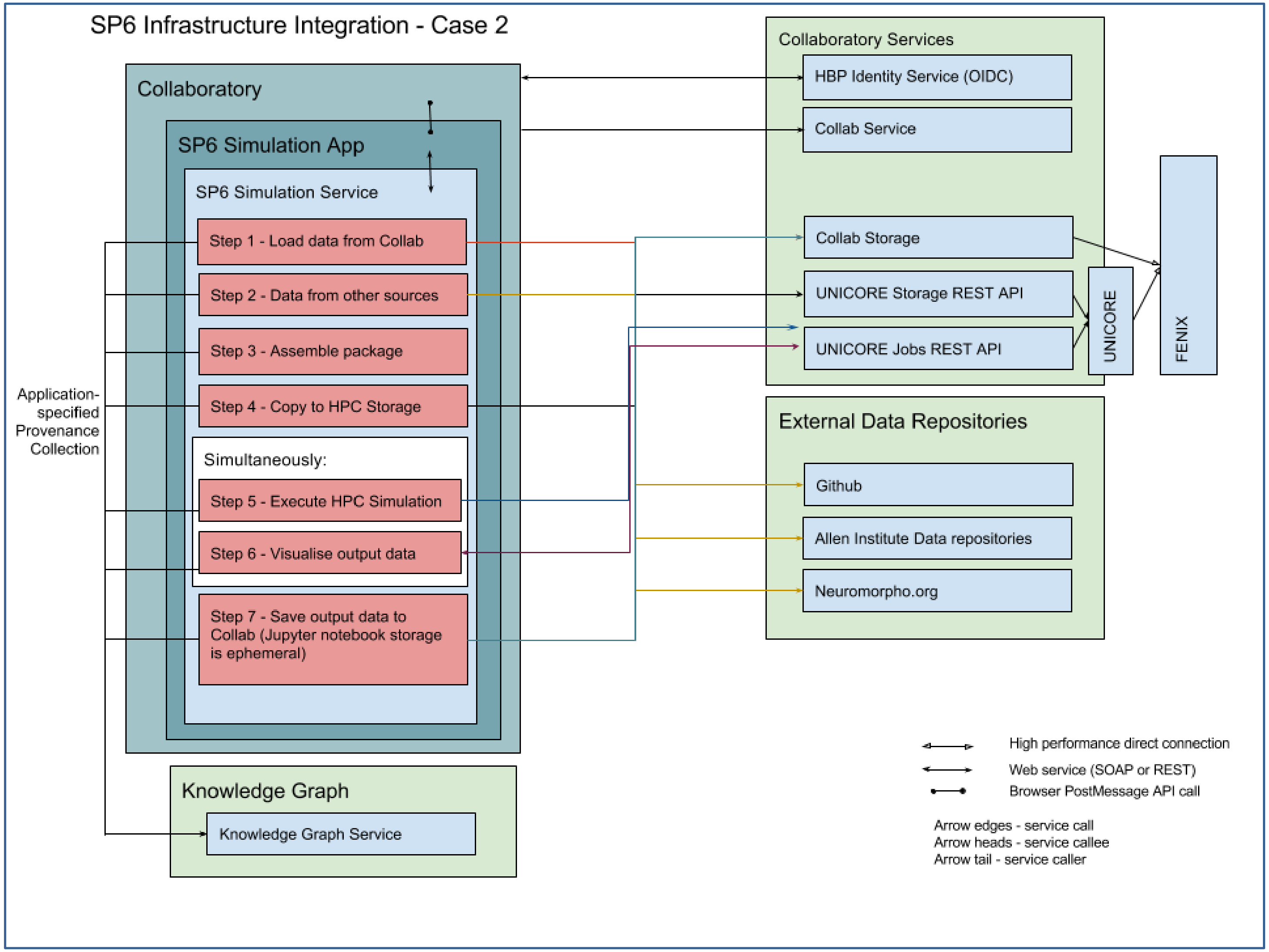


Figure 24: SP6 Infrastructure Integration - Case 2

5.3 Medical Informatics Platform

The Medical Informatics Platform (MIP) has different constraints on its authentication and authorisation practices due to privacy constraints which come when dealing with medical data. As a result, the MIP relies on external platform services much less than the other Platforms. It is expected that Medical Informatics could integrate more services from other SPs as the constraints on human data are better reflected in HBP Data Policy and later implemented in various services.

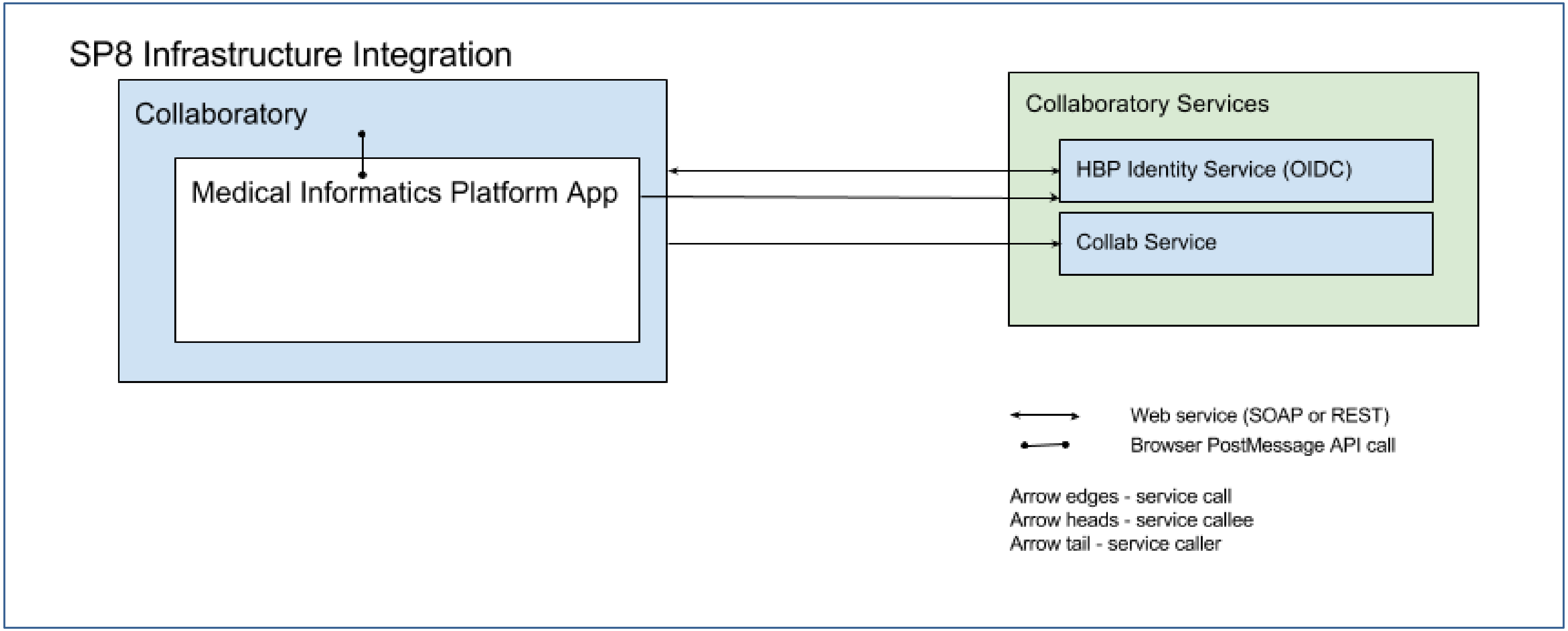


Figure 25: SP8 Infrastructure Integration

5.4 Neuromorphic Computing Platform

The Neuromorphic Computing Platform Collab App largely follow the model of the SP6 Case 2 Collaboratory App, but add an additional dimension integrating the Neuromorphic job services. While not diagrammed as below, it is worth noting that the Neuromorphic execution services can also be used from the Collaboratory Jupyter notebooks, thanks to the OIDC token integration in the Jupyter notebooks.

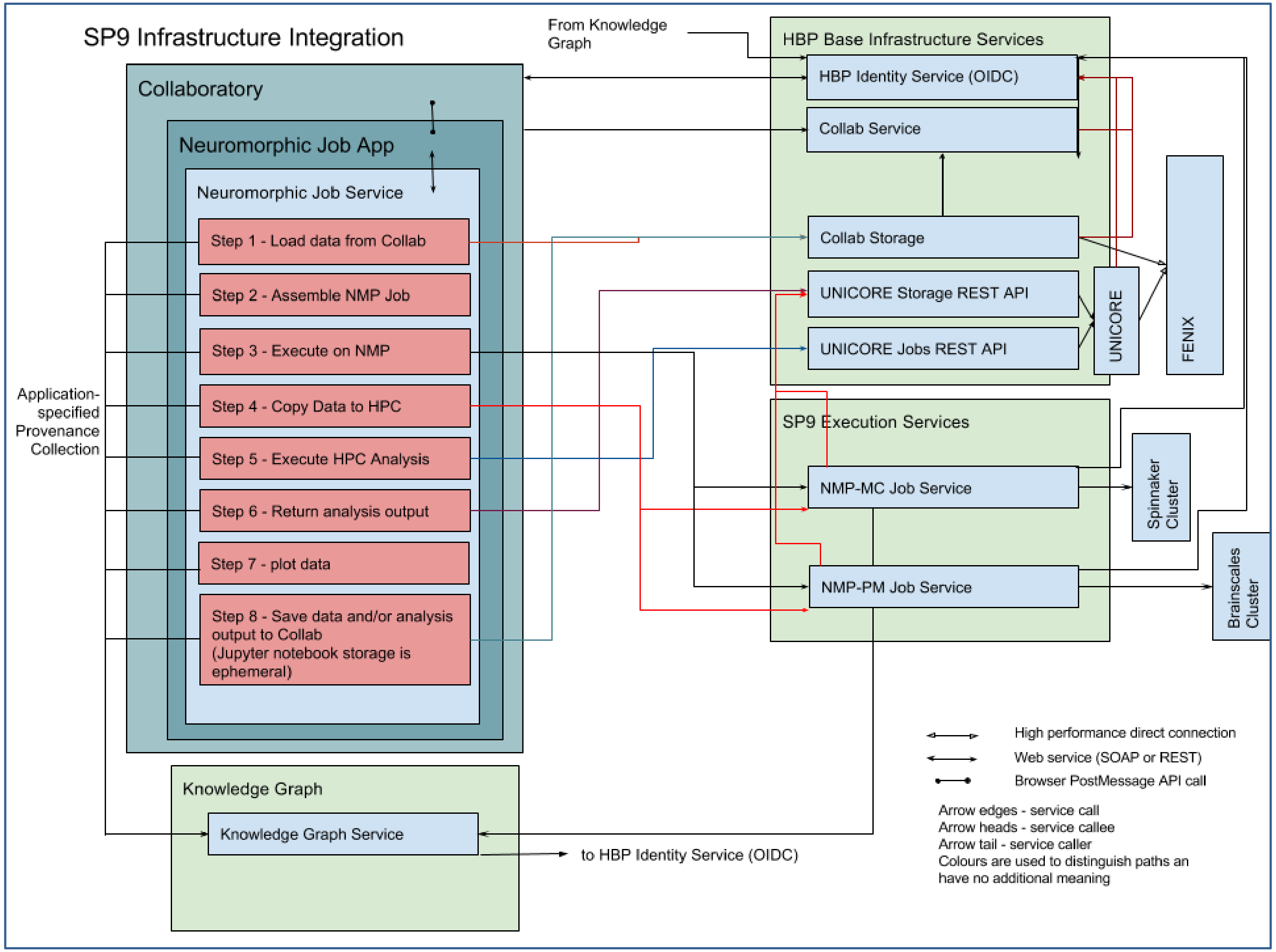


Figure 26: SP9 Infrastructure Integration

5.5 Neurorobotics Platform

The Neurorobotics Platform (NRP) also follows the SP6 Case 2 Collaboratory App integration model. This is expected to serve the HBP well in later versions of the NRP because the similarities in the integration patterns between the Neuromorphic Computing Platform (NCP) and the NRP will facilitate easy integration of NCP job execution services into the NRP simulation environment. This is a potential integration scenario which is not currently on the roadmap, but would be well supported and will only require having the NRP simulation backend collocated with the NCP for latency reasons.

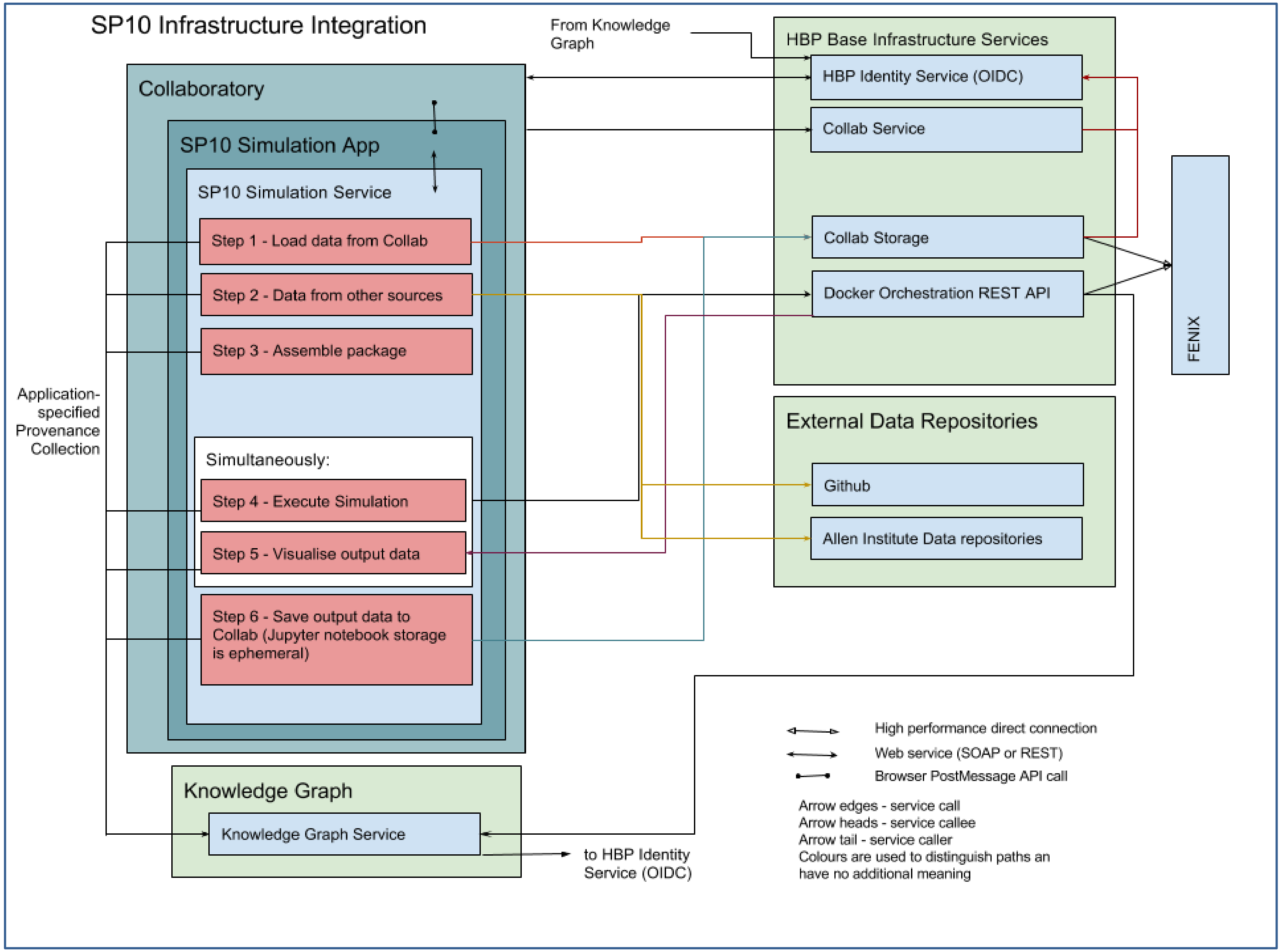


Figure 27: SP10 Infrastructure Integration

Annex A: Glossary

Term	Description
0-9	
2D Atlas	A 2D reference space, a collection of 2D parcellations, 2D images or a collection of registered data.
2D Parcellation	A collection of closed polygonal or spline boundaries at specific cut planes in 3D space. Each boundary is linked to one or more ontological elements.
2D Reference space	A collection of 2D subspaces each aligned with a specific cut plane in 3D space. Each subspace has a single coordinate origin and affine transformation.
3D Atlas	A 3D reference space, a collection of 3D parcellations, 3D voxel volumes and a collection of registered data.
3D Parcellation	A collection of meshes that define 3D boundaries. Each boundary is linked to one or more ontological elements.
3D Reference space	A set of 3D basis vectors with a single coordinate origin and affine transformation.
A	
Anchor	A spatial location with orientation and scale or a semantic-spatial with optional orientation and scale.
Artefact	A high data-density discrete data element, primarily meant to denote a file larger than 10kB which is not human readable or editable.
Atlas	A 2D or 3D atlas
B	
Biophysically realistic	Mathematical description of physical phenomena relevant to the biological processes of cellular behaviour. In particular, but not limited to, a model of the neuronal tree with the cable equation and ion channels by the phenomenological Hodgkin-Huxley formulation.
Brain System	A specific set of interacting brain regions.
C	
Capability supercomputing	Tightly integrated parallel supercomputer providing a high-speed and low-latency network between the computing nodes (in contrast to embarrassingly parallel computers).
Cockpit	Desktop, display wall or cave visualisation resource with a mechanism for good data locality.
Cognitive Architecture	A specific set of brain regions and interactions that are proposed to underlie specific cognitive capabilities.
Compute resource	A computer or collection of computers where a job can be executed
Configurability	The ability of the simulator to provide extensibility for additional mathematical formulations of novel physical phenomena and integrating this mathematics in the compute-critical inner integration loop.

Term	Description
Continuous time spike interaction	Representation of the time point of an action potential as a floating-point number with double precision on a continuous time axis. The action potential is generated at the time point of threshold crossing of the sending neuron and is communicated with full precision to its target cells, where it affects the receiving neuron, acting at the point in time after the application synapse delay.
CRUD	A commonly used acronym for Create, Read, Update, Delete.
Curation	A manual or analytic process involving a human to make decisions about some property of the reconstruction or one of its components. Curation can be applied to everything from electrical channel models in the Hodgkin-Huxley model to ontology names for a particular neuron morphology class.
D	
Datatype	A datatype is a semantically enriched mimetype. For example, the mimetype of a particular data file might be XML, but the datatype would be CircuitML, implying that the data file can be interpreted in a richer way. This allows the selection of editing interfaces and input data much more User friendly in the Collaboratory.
Detailed model	The finest level of representation with full geometry
E	
e-type	A short-hand form of the ne-type abbreviation defined below
Entity	A COLL Project, a file or a folder. Theses exist as part of a hierarchy; they may have a parent and children. Each entity has a series of predefined key values associated to it (like name, creation date…) and can also have some custom Metadata associated to it.
Exact integration	A method applicable to the integration of sets of linear differential equations. The solution agrees to the mathematically exact solution. Often formulated in terms of a matrix exponential.
F	
File	Entities that are required to have a parent but cannot have any children. In addition to the standard attribute they also have a Content URL that defines how to access the content of the file.
Folder	Plain Entities that are required to have a parent.
Full scale	Representation of a network with the natural number of neurons and synapses per neuron as found in the biological system.
G	
Glial cell	Non-neuronal cells that function in homeostasis and energy usage, which provide support and protection for neurons. They can be divided into microglia and macroglia types.
H	
HBP Collaboratory (HBP-COLL)	The unified web interface through which the web-accessible components of the six HBP Platforms and all other HBP activities are made available.
MINDS	A minimum metadata specification. Similar in spirit to the Carmen MINI specification but tailored to the Use Cases of the HBP.

Term	Description
Hidden Entity	Entities can be hidden from the Document Service REST API and through the Web GUI. Hidden data are not visible by default in the COLL Project browser. Hidden data will still be optionally visible (thought marked as hidden) to anyone it has been shared with. Hidden data are a separate function from true deletion. Deletion of data is a highly privileged operation that must be done by System Administrators on User request. See the Data Hiding Use Case for more information.
Host	A single operating system instance, running on virtualised or real hardware
I	
Ionic conductance models	Models that represent ionic permeation through the plasma membrane. Both stochastic and deterministic approaches should be covered. Extension toward molecular models (WP6.4.1) is envisaged.
J	
Job	An instance of an execution of a Task on a compute resource. For some Tasks, the compute resource will be selected by the User in the COLL on job launch. For other Tasks the execution will be decided by the Task.
L	
Level of resolution	The choice of abstraction applied to the representation of the network. The level of resolution of MolSim corresponds to single neurons or synapses. The level of resolution of NetSim corresponds to single neurons and synapses. The level of resolution of CellSims equates to electrical compartments coupled by conductances.
M	
m-type	A short-hand form of the nm-type abbreviation below
me-type	A short-hand form of the nme-type abbreviation below
Macrocircuit	The definition of the whole brain as a set of brain regions connected through long-range fibre tracts - the whole brain.
Mesocircuit	The definition of the smallest collection of midrange interacting microcircuits through their intra-areal or regional arbours - a brain area or region.
Metabolism	The set of chemical transformations within the cells of living organisms that maintain life.
Microcircuit	The definition of the smallest collection of short-range interacting neurons through their local arbours.
Microcircuit models	Models that represent an entire microcircuit, including 3D geometrical architecture, synaptic connectivity and neuronal and synaptic models.
Molecular level models	Models that are structurally accurate at the subcellular level (organelles, intracellular and extracellular spaces) and that contain molecules that ultimately follow cell biological rules of production, transport, localisation and degradation as well as the environment-dependent thermodynamics and kinetics of their interactions. Both stochastic as well as deterministic versions will be covered.

Term	Description
Molecular Simulations	Numerical simulations at the atomistic or coarse-grained level used for predicting structures of molecular complexes and the estimation of kinetic and thermodynamic parameters for molecular interactions. Molecular Simulations are based on atomistic structures of proteins available either from the Protein Data Bank or from homology modelling.
Morphology	The geometric definition of the shape of a neuron.
Multi-constraint fitting	The process whereby one data parameter or property constrains other data parameters or properties.
N	
n-type	A class of brain cells or a particular instance, depending on context.
ne-type	Abbreviation for an electrophysiological type of cell. This abbreviation is used to refer to a class of cells or a particular instance, depending on context.
ng-type	Abbreviation for a genetic type of cell. This abbreviation is used to refer to a class of cells or a particular instance, depending on context.
nm-type	Abbreviation for a morphology type of cell. This abbreviation is used to refer to a class of cells or a particular instance, depending on context.
nme-type	Abbreviation for a morpho-electrophysiological combination type of cell. This abbreviation is used to refer to a class of cells or a particular instance, depending on context.
np-type	Abbreviation for a protein type of cell. This abbreviation is used to refer to a class of cells or a particular instance, depending on context.
Neuro-glia vasculature (NGV)	The three principal components in neural tissue, which function as a unit to regulate blood flow and metabolism.
NEURON	Open source simulator NEURON (http://www.neuron.yale.edu) developed by Michael L. Hines.
Neuron model	Implementation of neuron dynamics defined as a set of differential equations. The implementation solves the dynamics within a finite time span given the incoming spike events are supplied. Incoming synapses can be modelled as currents or conductances.
Neuropil	Any area in the nervous system composed of mostly unmyelinated axons, dendrites and glial cell processes that form a synaptically dense region containing a relatively low number of cell bodies.
P	
p-type	An abbreviation for projection type, a to-be-determined classification scheme for determining classes of projections between meso-scale brain regions.
Parameter	A low data-density discrete data element that is primarily meant to denote a value that one might enter into a single form element. It might also be used to refer to a richer configuration document containing a group of settings.
Parcellation	One or more spatial boundaries associated with a set of discrete semantic concepts. Usually developed by manual, semi-automated or automated image analysis of landmarks.
Platform	Software components: libraries, services, APIs and their documentation that are to be used to build portals or cockpits.
Predictive reconstruction	The process whereby multi-constraint solutions yield a hypothesis and hence a prediction of the data parameter space.

Term	Description
COLL Project	COLL Projects are Entities with no parent. In addition with the standard attribute and Metadata associated with Entities, COLL Projects have also ACL that define which Users can access their content.
R	
Reconstruction data	Data that is used to parameterise a model.
Reconstruction process	A workflow that uses a configuration of the data parameters and implements a set of fundamental biological principles to constrain and instantiate the model.
Reference space	In 2D, a collection of slices with an optional 2D *parcellation*. In 3D, a voxel volume with an optional 3D *parcellation*.
Registered data	A URL accessible data set with an *anchor*.
Resources	Parameters, Artefacts, services, or compute capacity
REST	An acronym for REpresentational State Transfer, for a definition see http://en.wikipedia.org/wiki/Representational_state_transfer
S	
s-type	Abbreviation for the type of synaptic connection. This abbreviation is used to refer to a class of cells or a particular instance depending on context. Abbreviations for specific dimensions of a synapse include;
sa-type	Abbreviation for the anatomical type of synaptic connection. This abbreviation is used to refer to a class of synaptic connection or a particular instance depending on context.
sp-type	Abbreviation for the physiological type of synaptic connection. This abbreviation is used to refer to a class of synaptic connection or a particular instance depending on context.
SAN	An acronym for Storage Area Network, http://en.wikipedia.org/wiki/Storage_area_network
Semantic-spatial location	Association of semantic concept (e.g.: cerebellum) with a spatial boundary.
Service	A software function performed by a third party for a User or other Service. In the language of the COLL, Services consume Parameters, Artefacts and compute capacity. Services produce Artefacts and parameters.
Single neuron models	Models that represent entire neurons, including 3D structure, electroresponsiveness, synaptic activation and intracellular biochemical cascades (developed in WP6.4.1).
Site	A collection of hosts collected together in a single location. The grouping is potentially arbitrary. QIJ might be considered one site, LNMC another or one might consider EPFL a site unto itself.
Spatial location	2D or 3D location
Synapse model	A model representing synaptic plasticity, such as spike timing dependent plasticity (STDP). The implementation solves the dynamic equation describing the evolution of the synaptic amplitude, typically formulated as a differential equation, given the spike times of the presynaptic and possibly the postsynaptic neuron are given.

Term	Description
Synaptic models	Models that represent processes of synaptic transmission, including neurotransmitter release and postsynaptic receptor activation. Both stochastic and deterministic approaches should be covered. Extension toward molecular models and molecular networks (WP6.4.1) of neuromodulation, synaptic plasticity and homeostasis is envisaged.
Systems Biology Markup Language (SBML)	A mark-up language for representing standardised reaction networks within compartments.
T	
Task	A logical software unit. A Task takes Artefacts and Parameters as input, and produces Artefacts and Parameters as output. It may or may not be visible as a Service. A Task identifies its dependencies and its default parameters. Concretely, it is a software component that combines: • A Python-based Task entry point • A git repository or Python package index URL for the Task • A repository revision or package content specified by sha1 • A requirements file specifying all required dependencies. Tasks can have dependencies in non-Python languages, but these dependencies must be packaged for reproducible deployment.
Task definition	The collection of data that defines an individual Task
Task repository	A database of Task definitions
V	
Validation data	Data that is used to validate a model.
Validation process	A workflow that compares results obtained in the model when experimental protocols used to obtain the validation data are applied to the model..
Vasoconstriction	Narrowing of blood vessels resulting from constricting of smooth muscle cells within the vessel walls
Vasodilation	Widening of blood vessels due to relaxation of smooth muscle cells within the vessel walls
Voxel	A 3D unit volume, the 3D analogue of an image pixel.
Voxel volume	A 3D volume made up of voxels. Typically, the voxels densely fill a rectangular prism spatial bounding volume.
W	
Workflow	A tree of decision structures and Tasks. A Workflow takes Artefacts and Parameters as input, and produces Artefacts and Parameters as output. It may or may not be visible as a Service.

Appendix 1 – Resource List

Resources

Name	Default role
Backend Dev 1 - Collaborator	Backend Developer
Fullstack Dev 2 - Collaboratory	Fullstack Developer
Fullstack Dev 3 - Collaboratory	Fullstack Developer
Dev Ops 1 - Collaboratory	DevOps
Fullstack Dev / Product Owner - Collaboratory	Fullstack Developer
Fullstack Dev 4 - Collaboratory	Fullstack Developer
Frontend Dev 5 - Collaboratory	Frontend Developer
Backend Developer 1 - KG	Backend Developer
Backend Developer 2 - KG	Backend Developer
Backend Developer 3 - KG	Backend Developer
Jeff Muller	project manager
Fullstack Dev - Data Workbench	Fullstack Developer
Atlasing Dev 1 - Juelich	Fullstack Developer
Atlasing Dev 1- UIO	Fullstack Developer
FENIX	undefined

Appendix 2 - Gantt Chart

Name	Begin date	End date	Resources
FENIX VM Service - single site	07/12/16	05/04/17	
FENIX VM prototype capacity available	07/12/16	07/12/16	FENIX
FENIX VM prototype capability test	07/12/16	12/01/17	Fullstack Dev 3 - Collaboratory
Single-site HA capacity implemented	26/01/17	10/02/17	FENIX
VM capacity upgrade	13/02/17	15/03/17	FENIX
Production VM environment with internet accessible netw...	16/03/17	05/04/17	FENIX
COLL Platform-as-a-Service	09/02/17	31/03/17	
Pre-production testing	09/02/17	31/03/17	Dev Ops 1 - Collaboratory
ansible base recipe v0 for FENIX site 0	09/02/17	22/02/17	Dev Ops 1 - Collaboratory
Logging services - dev and prod	23/02/17	08/03/17	Dev Ops 1 - Collaboratory
Monitoring Services - dev and prod	09/03/17	20/03/17	Dev Ops 1 - Collaboratory
ansible base recipe v1 for FENIX site 0	21/03/17	24/03/17	Dev Ops 1 - Collaboratory
Neuroglancer viewer deployment v0	09/03/17	20/04/17	
Neuroglancer conversion scripts v0.1	09/03/17	16/03/17	Atlasing Dev 1 - Juelich
Conversion scripts v0.2 - adaptation for NIFTII (Waxholm ...	17/03/17	22/03/17	Atlasing Dev 1 - Juelich
Conversion scripts v0.3 - enhancements for initial viewer r...	23/03/17	28/03/17	Atlasing Dev 1 - Juelich
Bigbrain reference atlas w/ NG viewer v0	18/04/17	20/04/17	Atlasing Dev 1 - Juelich
Waxholm reference atlas w/ NG viewer v0	18/04/17	20/04/17	Atlasing Dev 1- UIO
Software Catalog FENIX migration	31/03/17	06/04/17	
Automate backups of software catalog database	31/03/17	31/03/17	Dev Ops 1 - Collaboratory
Automate ssl certificate renewal	03/04/17	03/04/17	Dev Ops 1 - Collaboratory
docker + ansible config of software catalog in prod	04/04/17	04/04/17	Dev Ops 1 - Collaboratory
Externalize database from docker container	05/04/17	05/04/17	Dev Ops 1 - Collaboratory
database import from EPFL to CSCS	06/04/17	06/04/17	Dev Ops 1 - Collaboratory
Jupyter Notebook FENIX migration	27/04/17	12/05/17	
Jupyter Swarm docker deployment in CSCS	27/04/17	10/05/17	Dev Ops 1 - Collaboratory
Selenium parallel test runner setup	27/04/17	03/05/17	Backend Dev 1 - Collaborator,Ful...
Selenium load testing of Jupyter Swarm in CSCS	04/05/17	10/05/17	Backend Dev 1 - Collaborator,Ful...
Jupyter Swarm in CSCS manual compatability testing	11/05/17	11/05/17	Jeff Muller
Switch production Jupyter app to use CSCS Swarm	12/05/17	12/05/17	Fullstack Dev 2 - Collaboratory
Storage V1 upgrade	31/03/17	28/04/17	
Opensourceable bbp-client	31/03/17	11/04/17	Backend Dev 1 - Collaborator
JS library upgrades for V1	31/03/17	11/04/17	Fullstack Dev 2 - Collaboratory
Instrumentation for COLL PaaS metrics in Storage service	17/04/17	28/04/17	Backend Dev 1 - Collaborator,De...
Jupyter Notebook Monitoring improvements	15/05/17	31/05/17	Dev Ops 1 - Collaboratory
Test Datadog analytics service	15/05/17	23/05/17	Backend Dev 1 - Collaborator
Test New Relic analytics service	15/05/17	23/05/17	Dev Ops 1 - Collaboratory
Integration of datadog in Storage service	30/05/17	30/05/17	
Integration of datadog in OIDC service	31/05/17	31/05/17	
FENIX S3 Storage Service - single site	02/02/17	08/06/17	
S3 Storage cluster 100TB	02/02/17	27/04/17	FENIX
S3 acceptance testing	28/04/17	08/06/17	Dev Ops 1 - Collaboratory
Atlasing infrastructure v2	09/06/17	27/06/17	
Testing of Neuroglancer deployment to FENIX S3	09/06/17	09/06/17	Dev Ops 1 - Collaboratory,Atlasin...
Rollout of Neuroglancer production to FENIX S3	27/06/17	27/06/17	Dev Ops 1 - Collaboratory,Atlasin...
Software Catalog FENIX migration	15/02/17	27/02/17	Dev Ops 1 - Collaboratory
Data Workbench MVP	03/04/17	25/08/17	
DW REST API v0 (self contained Postgres DB)	03/04/17	23/05/17	Fullstack Dev - Data Workbench
DW Web UI v0	03/04/17	23/05/17	Fullstack Dev - Data Workbench
Iterate on UI and backend services -> v0.1	24/05/17	23/06/17	Fullstack Dev - Data Workbench
Iterate on UI and backend services -> v0.2	26/06/17	25/07/17	Fullstack Dev - Data Workbench
Iterate on UI and backend services -> v0.3	26/07/17	25/08/17	Fullstack Dev - Data Workbench
Knowledge Graph MVP - restricted write access	29/05/17	08/09/17	
ACL specification	29/05/17	02/06/17	Backend Developer 1 - KG
Janusgraph deployment recipes, dev and prod	29/05/17	20/07/17	Backend Developer 3 - KG
Type Service REST API v0	05/06/17	07/08/17	Backend Developer 1 - KG
WAL DB spec + impl	05/06/17	12/06/17	Backend Developer 2 - KG
Ingest REST API v0	13/06/17	08/08/17	Backend Developer 2 - KG
Janusgraph backup and restore procedures	21/07/17	27/07/17	Backend Developer 3 - KG
WAL - Janusgraph Ingest Daemon	28/07/17	05/09/17	Backend Developer 3 - KG
Type Service deployment recipes, dev and prod	08/08/17	09/08/17	Backend Developer 1 - KG
Ingest REST API deployment recipes, dev and prod	09/08/17	10/08/17	Backend Developer 2 - KG
WAL - Janusgraph Ingest Daemon deployment recipes, de...	06/09/17	08/09/17	Backend Developer 3 - KG
Search MVP	15/06/17	20/07/17	
Elastic search ansible recipes with optional cluster, dev a...	15/06/17	20/06/17	Dev Ops 1 - Collaboratory
Develop Python rest gateway to Elastic Search	21/06/17	04/07/17	Fullstack Dev 2 - Collaboratory
Initial version of Python indexer daemon, indexing collab ...	21/06/17	04/07/17	Backend Dev 1 - Collaborator
Proof-of-concept implementation of ACL based filtering of ...	05/07/17	20/07/17	Fullstack Dev 2 - Collaboratory
Python rest gateway ansible recipes for dev and prod	05/07/17	06/07/17	Dev Ops 1 - Collaboratory
Indexer ansible recipes for dev and prod	05/07/17	06/07/17	Dev Ops 1 - Collaboratory
Initial version of Search UI replacing current Collab search...	05/07/17	20/07/17	Frontend Dev 5 - Collaboratory
Search v2	21/07/17	15/08/17	
Enabling indexer daemon for indexing of Data workbench	21/07/17	04/08/17	Backend Dev 1 - Collaborator
Enabling indexer daemon for indexing of Collab wiki docu...	02/08/17	15/08/17	Fullstack Dev 2 - Collaboratory
linkage with Data Workbench atlasing metadata, scope TBD	16/08/17	16/08/17	
Search usabilty refinement - fullstack	16/08/17	29/08/17	Fullstack Dev 2 - Collaboratory,F...
migration of trivial services from BBP infrastructure	16/08/17	05/09/17	Backend Dev 1 - Collaborator
decommissioning of UP provenance and task frameworks	30/08/17	30/08/17	Dev Ops 1 - Collaboratory
Infrastructure Migration - OIDC Service	31/08/17	11/10/17	
preparation of miigration guide for BBP customers	31/08/17	04/09/17	Jeff Muller
decoupling of OIDC service from BBP infrastructure	05/09/17	28/09/17	Backend Dev 1 - Collaborator,Ful...
migration of OIDC from BBP infrastructure	29/09/17	11/10/17	Fullstack Dev 2 - Collaboratory,D...
Knowledge Graph - NIP Integration	11/09/17	10/10/17	
Data Workbench, publish to Knowledge Graph	11/09/17	10/10/17	Fullstack Dev - Data Workbench
NIP Elastic Search population from Knowledge Graph	11/09/17	04/10/17	
Infrastructure Migration - Storage Service	12/10/17	27/11/17	
decoupling of Storage service from BBP Infrastructure	12/10/17	09/11/17	Backend Dev 1 - Collaborator,Ful...
migration of Storage service from BBP infrastructure	10/11/17	27/11/17	Backend Dev 1 - Collaborator,De...
OIDC and Storage Technical debt cleanup	28/11/17	11/12/17	
Read-write-admin separation	12/12/17	31/01/18	
Backend updates	12/12/17	10/01/18	Backend Dev 1 - Collaborator
UI updates	11/01/18	24/01/18	Frontend Dev 5 - Collaboratory
Collab app authorization review	11/01/18	31/01/18	Backend Dev 1 - Collaborator,Ful...
Group collab membership	01/02/18	07/03/18	
Backend updates	01/02/18	14/02/18	Backend Dev 1 - Collaborator
Team edit UI updates	15/02/18	23/02/18	Frontend Dev 5 - Collaboratory
Team view UI updates	26/02/18	07/03/18	Frontend Dev 5 - Collaboratory