| Grant Agreement Number: | 720270 | Project Title: | Human Brain Project |
|---|---|---|---|

| | |
|---|---|
| Document Title: | Second revised SP5 work plan for SGA1 |
| Document Filename[(1)]: | SGA1_D5.6.1_Resubmission_FINAL_20170116.docx |
| Deliverable Number: | SGA1 D5.6.1 Resubmission |
| Deliverable Type: | Report |
| Work Package(s): | WP 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8 |
| Dissemination Level: | PU = Public |
| Planned Delivery Date: | M 9 / 30 Dec 2016 |
| Actual Delivery Date: | M10 / 16 Jan 2017 |



| | |
|---|---|
| Authors: | SP5: Anastasia AILAMAKI, Jan BJAALIE (lead), Andrew DAVISON, Timo DICKSCHEID, Simon EICKHOFF, Pascal FUA, Sten GRILLNER, Sonja GRÜN, Fred HAMPRECHT, Anna KRESHUK, Pedro LARRAÑAGA, Trygve LEERGAARD, Daniel MALLMANN, Jean-François MANGIN, Jeff MULLER, Martin TELEFONT, Paul TIESINGA |
| Compiling Editors: | Jan BJAALIE, Jeff MULLER, Martin TELEFONT, Sofia ANDERSHOLM STRAND |
| Contributors: | Rembrandt BAKKER, Michael DENKER, Darius SIDLAUSKAS, Eduard TRULLS |
| Reviewers: | |



| | |
|---|---|
| Abstract: | |
| Keywords: | |

# Document Status

This revised work plan for SP5 is derived from Deliverable number D5.6.1 of October 10, 2016, developed by the HBP "Data Planning and Implementation Team", DPIT. The different document histories can be found in the respective DPIT working documents and DPIT notes as listed in the Appendix to D5.6.1 of October 10, 2016.

# Table of Contents



# Tables

# 1. Introduction

This second revised SP5 workplan for SGA1 has been developed in response to "result of the review of the "DPIT" results of your H2020 project 720270 HBP SGA1 (SGA1/DPIT) under FPA 650003 and of the SP5-related resubmissions of FP7 project 604102 HBP (RUP/SP5)" dated 28 November 2016. The revised workplan builds on the DPIT workplan of 10 October 2016, with changes explained in:

- "20161228 HBP answer letter to M6 DPIT review report" of December 28, 2016;
- "Detailed Response to Expert Review Report-20161228" of December 28, 2016;
- "Overview of changes in Second Revised SP5 workplan SGA1" of January 12, 2017.

## 2. Second Revised SP5 Workplan for SGA1

The following table presents the second revised work plan for SP5.

| Subproject no. | SP5 | Lead Participant | UIO | Starting: | M01 | Ending: | M24 |
|---|---|---|---|---|---|---|---|
| Subproject title | Neuroinformatics Platform | | | | | | |
| Activity Type | Research & Innovation | | | | | | |

| Participant number | 1 | 81 | 20 | 39 | 47 | 11 | 51 | 68 | 10 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Participant short name | EPFL | UIO | JUELICH | KI | UHEI | CEA | SKU | UPM | CNRS | |
| Person-months | 282.9 | 189.5 | 158 | 39 | 36 | 24 | 24 | 24 | 15 | 792.4 |
| Person-years | 23.6 | 15.8 | 13.2 | 3,3 | 3 | 2 | 2 | 2 | 1.25 | 66.1 |

**Objectives and Description**

| Work package No. | Work package title | Type of activity[1] | Lead Participant Short name | Lead Participant no.[2] | Person-months[3] funded+inkind | Start month[4] | End month[5] |
|---|---|---|---|---|---|---|---|
| WP 5.1 | Data Curation Support Lab | RIA | EPFL | 1 | 61 + work in SP7 | 1 | 24 |
| WP 5.2 | Multi-level Atlas of the Rodent Brain | RIA | UIO | 81 | 116 | 1 | 24 |
| WP 5.3 | Multi-level Atlas of the Human Brain | RIA | JUELICH | 20 | 117.5 | 1 | 24 |
| WP 5.4 | Data and Atlas Curation Tools | RIA | UIO | 81 | 160.9 | 1 | 24 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| WP 5.5 | Community-Driven Neuroinformatics Platform and Infrastructure Operations | RIA | EPFL | 1 | 147.1 +70 (in-kind) | 1 | 24 |
| WP 5.6 | Data Mining and Analysis Neuroinformatics Capabilities | RIA | UHEI | 47 | 60 | 1 | 24 |
| WP 5.7 | Tools and Curation for Integrated Parallelized Analysis of Activity Data | RIA | JUELICH | 20 | 40+9 from SP9 | 1 | 24 |
| WP 5.8 | Management and Coordination | RIA | UIO | 81 | 80.9 | 1 | 24 |

Table 1: Description for Work Package WP5.1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Work Package no. | **WP5.1** | Lead Participant | **Ecole Polytechnique Federale de Lausanne (P1 EPFL)** | Starting: | **M01** | Ending: | **M24** |
| Work Package title | **Data Curation Support Lab** | | | | | | |
| Activity Type | **Research & Innovation** | | | | | | |
| Participant number | **1** | **20** | Total | | | | |
| Participant short name | **EPFL** | **JUELICH** | | | | | |
| Person-Months | **61** | **0+work in SP7** | **61+work in SP7** | | | | |
| Person-Years | **5.1** | **0.0+work in SP7** | **5.1+work in SP7** | | | | |
| Objectives | | | | | | | |

Making HBP data discoverable and accessible are essential parts of the HBP pledge to the community.

This Work Package will support users in curating and sharing data/models with other researchers in the HBP, assist them in efforts to reach elevated level of data consistency, and support migration of their data/models into the open data domain. In the context of WP5.1 data curation refers to:

- enrichment of datasets/models with metadata considered useful for data discovery using a search

engine (species, age, protocol, provenance, etc.) [T5.1.1];
- ensuring that metadata terms are linked to entities in ontologies, to enable "smarter" queries (e.g. a search for "rodent" returns results tagged as "mouse" and "rat") [T5.1.2];
- transfer of data/models to HPAC storage solutions, to allow usage of data for FENIX supported workflows (visualization, alignment, simulation, etc.) [T5.1.3].

As a result of its activity, WP5.1 will:

- Make all Ramp-Up Phase (RUP) and SGA1 data/models discoverable, via metadata enrichment;
- Maintain and expand ontologies for data discoverability, link them where possible to community ontologies;
- Catalogue data standards used in the HBP;
- Engage the Community to annotate data and data resources with HBP Data Discoverability Tools;
- Enable researchers to make their data accessible using the HBP Collaboratory, and the federated data infrastructure FENIX.

WP5.1 will also ensure implementation of HBP ethical guidelines concerning 1) verification (by data providers) that data registered in HBP infrastructure have been generated in compliance with HBP accepted ethical guidelines, and 2) management of data security and privacy at all levels of the SP5 infrastructure.

**Description of work and role of participant**

Task 5.1.1 - Data and Parameter Workbench Support and Curation Team (EPFL (28), M01 - M24):

This Task is responsible for metadata enrichment of all data and models produced during RUP and SGA1. This will be done by providing support to HBP researchers in the use of tools developed in T5.4.1 and perform assisted curation with researchers in SP1-10. The Task is also responsible for engaging HBP external researchers to drive adoption of services developed in T5.4.1 and offer training in curation workflows and best practices (100% service & support).

- Ensure all RUP/SGA1 data is discoverable in services provided by T5.4.1.
- Assist HBP researchers in adopting tools, developed in T5.4.1.
- Assist HBP researchers in adopting enrichment workflows developed in T5.4.1.
- Organize, verify and improve metadata enrichments provided by SPs 1-10 and external researchers, to improve discoverability of data and parameters.

**Product 1-1 Metadata Enriched RUP/SGA1 data and models**

RUP and SGA1 Data and Models will be enriched so as to make them discoverable via the Neuroinformatics Platform (NIP).

SGA1 – UC 1-47, 49, 50, 52, 53, 57, 59 -79

**Product 2-2 Living report on adoption of T5.1.1 tools by HBP and external Researchers**

This report will provide an overview of how the outreach and training Component of this Task faired.

SGA1

Task 5.1.2 - Ontologies for the integration of Models in the KnowledgeGraph (EPFL (33), M01 - M24):

This Task will perform the knowledge engineering of the ontologies required to support the data-driven model building. This concerns all data and models of HBP. The ontologies will be developed to support the KnowledgeGraph queries. This is necessary to feed the modelling pipelines in the Brain Simulation Platform and other Platforms.

**Product 3-1 Identification of HBP users' Use Cases**

In order to build scientifically relevant ontologies, we will need to interview HBP users in order to understand which Use Cases our ontologies need to support. We will document these Use Cases, as well as the resulting ontology architecture.

SGA1 – UC 1-47, 49, 50, 52, 53, 57, 59 -79

**Product 4-2 Ontology engineering**

This Task will build necessary ontologies (create new or integrate/adapt existing where applicable) to support the process of curating all RUP/SGA1 data and models. These ontologies will enable HBP users to perform queries on the KnowledgeGraph that address their scientific Use Cases.

SGA1 – UC 1-47, 49, 50, 52, 53, 57, 59 -79

Task 5.1.3 - Large-Scale Federated Data Accessibility Services (JUELICH (0 / SP7)), M01 - M24):

This Task will work on leveraging HBP storage services in SP7 and the Collaboratory by supporting users in:

- Downloading a dataset stored within the federated data infrastructure that is provided in SP7;
- Initiating a data transfer from an archive repository in the federated data infrastructure to an active repository in order to use this dataset e.g. for an analysis;
- Searching for and using datasets stored in the federated data infrastructure of SP7 from the Collaboratory;
- Making datasets available to the Consortium or the public by uploading them into the federated data infrastructure.

This Task closely collaborates with T7.1.2 (which deploys and operates the data services of the federated data infrastructure) and with T7.5.7 (which provides general support for the HPAC Platform (SP7)).

**Product 6-1 Support for data upload and download**

Supporting users to upload data to a repository in the federated data infrastructure or to download it from such a repository.

SGA1 – UC all except 27, 28, 29, 44, 49, 51, 58

**Product 7-2 Support for data transfers**

Supporting users to transfer data between an archive and an active repository in the federated data infrastructure.

SGA1 – UC all except 27, 28, 29, 44, 49, 51, 58

**Product 8-3 Support for data searches**

Supporting users to search for datasets stored in the federated data infrastructure from the Collaboratory.

SGA1 – UC all except 27, 28, 29, 44, 49, 51, 58

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.1.1 | List of KPIs for internal and external user adoption of Products/services offered by WP5.1 | 1 | M10 | Service and Support WP |

Table 2: Description for Work Package WP5.2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.2 | **Lead Participant** | Universitetet I Oslo (P81 UIO) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Multi-level Atlas of the Rodent Brain | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 81 | | 51 | **Total** | | | |
| **Participant short name** | UIO | | SKU | | | | |
| **Person-Months** | 92 | | 24 | 116 | | | |
| **Person-Years** | 7.7 | | 2.0 | 9.7 | | | |

**Objectives**

In this WP, multilevel atlases of the rodent brain will be constructed through spatial anchoring of data produced by other SPs to common reference atlases. Through the anchoring process, location will be assigned to all data, allowing query-based retrieval and data mining. Additional efforts will be made to import relevant experimental data available from external sources.

The results will be spatially anchored data and metadata organized in the KnowledgeGraph, to be shared with the research community and used for strategic data mining and analytics.

Description of work and role of participant

Task 5.2.1 - Maintenance of Rodent Atlases (UIO (12), M01 - M24):

This Task will focus on updating of content and functionality of the rodent reference atlases. Atlas maintenance efforts will include:

- Updating the image templates used for the 3D reference atlas when newer and better templates become available. More sophisticated templates will improve spatial registration and allow more advanced analyses;
- Managing the image templates and delineation files constituting the reference atlases, validating and integrating new or adjusted delineations, and compiling downloadable files in different formats;
- Making use of image data of multiple categories (products from RUP listed below and new data registered in T5.2.2) to reveal salient anatomical features needed to identify anatomical and functional boundaries of key importance for data interpretations;
- Expanding current anatomical delineations by adding structures at different levels of granularity based on the existing image templates;
- Updating, sharing, and managing customized atlas versions with granularity tailored for different analytical purposes.

The product list below is divided into two parts: products from RUP and products to be delivered in SGA1. In the list, atlases consisting of data sets with a specific feature are named accordingly (e.g., Neurotransporter atlas for GLT1, Product 21-13).

**Completed RUP products that provide a starting point for the work to be conducted in SGA1:**

**Product 9-1 Waxholm Space**

A spatial reference system based on internal brain landmarks according to the Waxholm Space standard has been applied to the MRI template used to create the WHS rat brain reference atlas (See Products 2-4, 20, 22 below), together with the widely used stereotaxic coordinate system based on cranial sutures and related stereotaxic landmarks.

RUP 5.1.1 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76.

**Product 10-2 Waxholm Space rat brain reference atlas, v1.0**

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 of 76 major structures are based on image contrast as described in detail in the accompanying publication (Papp et al. Neuroimage 97:374-86. doi: 10.1016/j.neuroimage.2014.04.001, 2014).

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 11-3 Waxholm Space rat brain reference atlas, v1.0.1**

An updated version of the volumetric atlas template used for the Waxholm Space rat brain atlas, transformed to comply with the NIfTI-1 spatial orientation standard. Correspondingly transformed atlas delineations and new voxel coordinates describing the position of the WHS origin are provided.

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 12-4 Waxholm Space rat brain reference atlas, v2.0**

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 TI 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. In version 2.0, delineations of 79 brain structures are provided, including a detailed anatomical parcellation of the hippocampal region based on image contrast as described in the accompanying publication (Kjonigsen et al., Neuroimage. 108:441-9. doi: 10.1016/j.neuroimage.2014.12.080 2015).

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 13-5 Enriched Waxholm Space rat brain reference atlas: PLI / M2 data**

3D reconstructed serial histological images showing muscarinic (M2) and polarized imaging fibre orientation data non-linearly warped to the WHS rat brain reference atlas, providing a basis for further delineations of anatomical structures as well as analyses of fibre orientations and M2 receptor distributions across the whole brain.

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 14-6 Atlas of rat brain cyto-and myeloarchitecture**

Image repository holding microscopic images of normal rat brain sections (cut in coronal orientation) stained to show cytoarchitecture (Thionine) and myeloarchitecture (Woelche, J Psychol Neurol 51: 199-202, 1942). Data from this collection have been used in two publications (Leergaard et al., 2010; White et al., 2013). Image data are available to HBP collaborators and can be interactively inspected in a web-based viewer tool.

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 15-7 Whole brain connectivity atlas of the rat**

The Whole Brain Connectivity Atlas is an interactive resource providing access to experimental tract-tracing image data showing the brain-wide distribution of efferent axonal connections originating in the rat primary somatosensory cortex. The atlas contains the results from six experiments in which an anterograde tracer (biotinylated dextran amine, BDA or *Phaseolus vulgaris* leucoagglutinin, Pha-L), was injected in SI forelimb or whisker representations. Images were collected at 200-400 micrometer intervals across the entire brain. Images from four cases were 3-D reconstructed into histology volumes. The results of data analyses were published in Zakiewicz et al., 2011, 2014, 2015). Image data can be interactively inspected a web-based viewer tool, or downloaded for further analysis.

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Product 16-8 Volumetric Allen mouse brain reference atlas**

Volumetric version of the Allen mouse brain reference atlas version 2.0 converted to NIfTI format for use with different atlas viewer tools (QuickNII, MeshView, CutNII, see T.5.4.2).

RUP 5.1.1 – UC 1, 5, 7, 23, 32, 33, 62, 63, 69, 76

**Product 17-9 Update of Allen mouse brain reference atlas**

Incorporation of an updated Allen mouse brain reference atlas version 3.0 in NIfTI format for use with the image registration tool QuickNII (see T.5.4.2).

RUP 5.1.1. – UC 1, 3, 5, 8, 13, 15, 17, 18, 19, 23, 38, 39, 40, 41, 46, 62, 63, 76

**Product 18-10 Three-plane high resolution architectonic atlas of the rat hippocampal region**

Web application providing access to high resolution microscopic images through the hippocampal region with series of histological sections cut in standard planes (coronal, sagittal, and horizontal) and stained to visualize chemical marker substances (calbindin and parvalbumin), and neuronal nuclei, an interactive viewer tool for virtual microscopy of microscopic images, as well as encyclopaedic information about anatomical structures and boundaries in the hippocampal region. The coronal images are publicly available. Sagittal and horizontal series are currently only available to HBP collaborators. Detailed anatomical descriptions of structures and boundaries are published (Boccara et al., Hippocampus 23. doi: 10.1002/hipo.22407, 2014).

RUP 5.1.1 – UC 1 ,3, 5, 8, 13, 15, 17, 18, 19, 23, 38, 39, 40, 41, 46, 62, 63, 76

**Product 19-11 Standardized anatomical landmarks for registration of whole brain imaging datasets to mouse brain Waxholm Space**

Standardized set of anatomical landmarks for registration of whole brain experimental image data from the mouse brain to Waxholm Space. Internal landmarks of the C57BL/6J mouse brain (Waxholm Space template) 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.

RUP 5.1.1 – UC 1, 3, 5, 8, 13, 15, 17, 18, 19, 23, 38, 39, 40, 41, 46, 62, 63, 76

**Product 20-12 Tg-Arc-Swe atlas**

Transgenic mice carrying the Arctic (E693G) and Swedish (KM670/6701NL) amyloid-precursor protein (AβPP) develop amyloid-beta (Aß) deposits in the brain that resemble Alzheimer's disease neuropathology (Lord et al., Neurobiol. Aging 27:67-77, 2006). The tg-ArcSwe atlas provides access to Aßx-40 immunolabelled histological section images from representative 12-month old tg-ArcSwe mice that show the morphology and spatial distribution of Aßx-40 plaque deposits across the brain. Image data can be interactively inspected via a web-based viewer tool, or downloaded for further analysis.

RUP 5.1.1 – UC 13, 46

**Product 21-13 Neurotransporter atlas: GLT1**

The VGLUT3-atlas provides online access to high-resolution images showing the distribution of vesicular glutamate transporter 3 (VGLUT3) in the normal rat brain. Image data can be interactively inspected via a web-based viewer tool, or downloaded for further analysis. Re-use of data from these repositories is allowed, provided that reference is given to the original publication.

RUP 5.1.1 – UC 13

**Product 22-14 Neurotransporter atlas: VGLUT3**

The GLT1-atlas provides online access to high resolution images showing the distribution of the glutamate transporter GLT1 (slc1a2, excitatory amino acid transporter, EAAT2), C-terminal variants GLT1a and GLT1b) in the normal mouse and rat brain, visualized by immunohistochemistry using antibodies against GLT1a and GLT1b. Image data can be interactively inspected a web-based viewer tool, or downloaded for further analysis.

RUP 5.1.1 – UC 13

**Product 23-15 tTA mouse brain atlas: Nop**

Conditional transgenic mouse models based on the tetracycline-regulatable transactivator (tTA) system involve bigenic constructs in which a tTA-regulated promoter controls the expression of a gene of interest. The spatial location of conditional gene expression depends on promoter activity. The tTA atlas resource contains histological image data showing the brain-wide location of promoter expression in a bigenic promoter-reporter mouse line that encodes the microscopically detectible Product ß-galactosidase (Nop/LacZ). The web-application provides access to high-resolution microscopic images showing the spatial distribution of ß-galactosidase, relevant for selection of promoter lines for construction of conditional models, and interpretation of observations from models constructed using these promoter lines. Image data can be interactively inspected online in virtual microscopy viewers, or downloaded for further analysis.

RUP 5.1.1 – UC 77

**Product 24-16 tTA mouse brain atlas: PrP**

Conditional transgenic mouse models based on the tetracycline-regulatable transactivator (tTA) system involve bigenic constructs in which a tTA-regulated promoter controls the expression of a gene of interest. The spatial location of conditional gene expression depends on promoter activity. The tTA atlas resource contains histological image data showing the brain-wide location of promoter expression in a bigenic promoter-reporter mouse line that encodes the microscopically detectible Product ß-galactosidase (PrP/LacZ). The web-application provides access to high-resolution microscopic images showing the spatial distribution of ß-galactosidase, relevant for selection of promoter lines for construction of conditional models, and interpretation of observations from models constructed using these promoter lines. Image data can be interactively inspected online in virtual microscopy viewers, or downloaded for further analysis.

RUP 5.1.1 – UC 77

**Product 25-17 tTA mouse brain atlas: PrP-CamKII**

Conditional transgenic mouse models based on the tetracycline-regulatable transactivator (tTA) system involve bigenic constructs in which a tTA-regulated promoter controls the expression of a gene of interest. The spatial location of conditional gene expression depends on promoter activity. The tTA atlas resource contains histological image data showing the brain-wide location of promoter expression in a bigenic promoter-reporter mouse line that encodes the microscopically detectible Product ß-galactosidase (PrP/CamKII/LacZ). The web-application provides access to high-resolution microscopic images showing the spatial distribution of ß-galactosidase, relevant for selection of promoter lines for construction of conditional models, and interpretation of observations from models constructed using these promoter lines. Image data can be interactively inspected online in virtual microscopy viewers, or downloaded for further analysis.

RUP 5.1.1 – UC 77

**Product 26-18 Database on rat cerebro-cerebellar connectivity**

Database holding 3-D reconstructed data (point coordinates) representing the spatial distribution of cerebropontine axonal projections and pontocerebellar projection neurons in the rat pontine nuclei. The data are derived from 74 anterograde or retrograde axonal tracing experiments (112 tracer injections placed in cerebral or cerebellar cortex as single, dual or triple injections), where tracer labelled axons and cells in the pontine nuclei were coded by point coordinates, 3-D reconstructed, and spatially registered to a common coordinate system. The spatial distributions of labelled elements and topographical organization patterns have been described in several publications. Visualizations of the 3-D data are available as standardized diagrams, and animated sequences (**www.rbwb.org**; choose Cerebro-cerebellar I). Data are shared as xyz point coordinates in a standard coordinate system for the pontine nuclei, which is spatially defined in the Waxholm Space rat brain atlas.

RUP 5.1.1 – UC 4, 13, 17, 23, 38, 46

**Products to be delivered in SGA1:**

**Product 27-19 Customized versions of Allen mouse brain atlas tailored for different analyses**

The full version of the Allen mouse brain reference atlas contains a very large number of small structures. For many types of analyses this level of granularity is inappropriate. To facilitate semi-quantitative analysis of spatial distributions of labelled markers in images that have been anchored to the Allen mouse brain reference atlas, we will create several custom versions of the atlas with different granularity (i.e. number of structures). These custom versions will be bundled in the QuickNII tool.

SGA1 – UC 1, 3, 5, 8, 13, 15, 17, 18, 19, 23, 38, 39, 40, 41, 46, 62, 63, 76

**Product 28-20 Allen mouse brain reference atlas with white matter structures parcellated**

In the Allen mouse brain reference atlas, white matter regions are not segmented as individual regions. This is a limitation for several visualization and analytic purposes. We will therefore create an updated version of the atlas where the white matter is parcellated into separate smaller structures.

SGA1 – UC 1, 3, 5, 8, 13, 15, 17, 18, 19, 23, 38, 39, 40, 41, 46, 62, 63, 76

**Product 29-21 Waxholm Space rat brain reference atlas enriched with additional receptor data**

Enrichment of the Waxholm Space rat brain reference atlas with brain-wide data on the spatial distribution of excitatory, inhibitory, and modulatory receptors, expanding on the results from RUP efforts (Product 5, above). These data will be used as strategic data for modelling purposes and to drive efforts to further delineate the WHS reference atlas (Product 22, below).

SGA1 – UC 4, 13, 17, 23, 38, 46

**Product 30-22 Expanded and improved Waxholm Space rat brain reference atlas (with additional and corrected structure delineations) based on registered multimodal data**

The Waxholm Space rat brain reference atlas has been delineated in several steps (Products 1-4), above). Enrichment of the reference atlas by spatial registration of different types of experimental data (Products 5, 21, see also T5.2.2) provides new substrates for further and more refined delineations of additional brain structures and systems. Utilizing new data registered to the atlas space, we will update and expand this reference atlas resource. The WHS rat brain atlas is already widely used (> 41,000 downloads) and improved delineations will have great value for the community.

SGA1 – UC 4, 13, 17, 23, 38, 46

Task 5.2.2 - Curation of Rodent Atlas Data (UIO (32), M01 - M24):

This Task will curate the spatial anchoring of experimental data to the reference atlases using tools and standards developed in WP5.4. The anchoring will be performed by the researchers producing the data (SPs 1–3), with guidance from the curation team. The outcome will be spatial metadata that will be added to other metadata (T5.4.1) in the KnowledgeGraph.

Location will be assigned to data from SP1 through the anchoring of the image data to the reference atlases. Experimental data from SP1 includes a wide range of strategic data on mouse brain molecular, cellular and system organization, including genetics data, proteomics data, and system-level descriptions of neural connectivity and brain vasculature. Data will be generated using various experimental techniques, and will include 2-D or 3-D images covering entire brains or discrete regions (e.g. electron microscopy stacks, electrophysiological recordings, and molecular signatures), or tables and lists of genes and proteins of interest, together with different accompanying data descriptions derived by quantitative analysis.

For each category of experimental data, unique metadata are required in order to curate the spatial anchoring of data to reference atlases. These metadata requirements will be delivered to T5.4.1 to be embedded in the tools for metadata curation and will be used in the services provided by T5.1.1.

The outcome of this data curation will be an organized collection of large data sets with searchable spatial (atlas coordinate based) or semantic (atlas brain structure-based) metadata and with access to the original data through the KnowledgeGraph.

The curation team will develop and optimize procedures for anchoring of data, contribute to training, supervision and coordination, and validate the quality of the anchoring. It provides a Tier-2 curation service to HBP.

**Product 31-1 Tutorials, training and supervision in assignment of spatial metadata**

The QuickNII tool (see T.5.4.2) is quite mature and a workflow for spatially defining large numbers of serial image data to reference atlas space has been tested and optimized. We will produce a tutorial and course material for providing practical training in the use of the tool and associated procedure. The UiO team will train users in hands-on courses as well as remotely by email and teleconference contact. Data produced in RUP (see T5.2.1) will be used for training purposes. Support will be given to researchers using the tools and procedures at HBP meetings, by phone, email, and teleconference.

SGA1 – UC 1, 3, 4, 5, 6, 8, 9, 13, 17, 18, 19, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 32-2 Validation and approval of spatial metadata before final entry in KnowledgeGraph**

Spatial metadata assigned to experimental data will be reviewed by a team of expert curators who will evaluate spatial accuracy and completeness, and approve data for ingestion in the KnowledgeGraph.

SGA1 – UC 1, 3, 4, 5, 6, 8, 9, 13, 17, 18, 19, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 33-3 Optimized procedure for anchoring of 2D image data to reference atlas**

As optimized versions of the QuickNII tool for spatial registration are released for HBP and public use, data registration procedures and associated tutorials will be updated and optimized, also taking into account accumulated user experiences.

SGA1 – UC 1, 3, 4, 5, 6, 8, 9, 13, 17, 18, 19, 21, 22, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 34-4 Procedure for anchoring of 3D image data to reference atlas**

This procedure will build on the QuickNII tool and procedures for anchoring 2D data to reference atlas space, and be adapted for handling of volumetric image data as series of pre-aligned 2D images.

SGA1 – UC 3, 13, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73

**Product 35-5 Procedure for non-linear warping of 2D image data to reference atlas**

This procedure will describe how methodology for non-linear warping of 2D image data to reference atlas space, as a follow up after anchoring of images using QuickNII (see Product 3, above).

SGA1 – UC 1, 3, 5, 8, 9, 13, 17, 18, 19, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 36-6 Procedure for non-linear warping of 3D image data to reference atlas**

Methodologies for warping of image volumes to reference atlas space exist, but require specialised tools and expertise. We will first establish procedures for non-linear warping of different data types (T5.1.2), and secondly develop these for use by non-experts.

SGA1 – UC 1, 3, 5, 8, 9, 13, 17, 18, 19, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 37-7 Curation of semantic spatial metadata delivered in T5.4.1**

Here a team of experts in assignment of anatomical location will evaluate the granularity, accuracy and completeness of semantic spatial metadata assigned to experimental data delivered in T5.1.1, prior to ingestion in the KnowledgeGraph.

SGA1 – UC 1, 3, 5, 6, 8, 9, 13, 17, 18, 19, 21, 22, 23, 31, 32, 33, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

Task 5.2.3 - Harvesting and Curation of Strategic Metadata from Existing Third Party Data Repositories (UIO (24), M01 - M24):

The goal of this Task is to make a selection of publicly available datasets hosted in external neuroinformatics data repositories discoverable through NIP. Efforts will focus on identifying and evaluating data collections that have high relevance for predictive analytical and modelling Tasks (WP5.5, SP6). Data curators will employ tools developed in WP5.4 to register the anatomical spatial locations of third party data to the HBP reference atlases by importing or converting already provided semantic and/or spatial metadata from the external source, or by spatially anchoring image data to the rodent reference atlases.

In SGA1 the primary objective will be to define specific procedures for harvesting data from important repositories, and to validate these by harvesting and curating 5-10 complete data sets from each data repository. Following optimization of procedures, additional data will be added to the extent possible.

**Product 38-1 Metadata and links to Allen Institute mouse brain data repositories (https://www.alleninstitute.org/)**

Strategic mouse brain gene-expression and connectivity data shared in Allen Institute repositories will be selected for ingestion into the KnowledgeGraph in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KnowledgeGraph together with links pointing to the native data stored at the Allen Institute.

SGA1 – UC 1, 3, 5, 13, 17, 18, 19, 21, 22, 23, 29, 32, 38, 39, 40, 41, 46, 62, 63, 73

**Product 39-2 Metadata and links to data shared via the mouse connectome project (www.mouseconnectome.org)**

Strategic data shared by the Mouse Connectome Project will be selected for ingestion into the KnowledgeGraph in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KnowledgeGraph.

SGA1 – UC 1, 3, 5, 13, 17, 18, 19, 21, 22, 23, 29, 32, 38, 39, 40, 41, 46, 62, 63, 73

**Product 40-3 Registration of rat brain hippocampal connectivity descriptions from temporal lobe database (www.temporal-lobe.com)**

Strategic mouse data shared via the Temporal Lobe repository will be selected for ingestion into the KnowledgeGraph in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KnowledgeGraph.

SGA1 – UC 1, 3, 5, 13, 17, 18, 19, 21, 22, 23, 29, 32, 38, 39, 40, 41, 46, 62, 63, 73

**Product 41-4 Registration of rat brain connectivity descriptions from the BAMS database (https://bams1.org/)**

Strategic data about rat brain connectivity shared via the Brain Architecture Management System will be selected for ingestion into the KnowledgeGraph in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KnowledgeGraph.

SGA – UC 4, 13, 17, 18, 19, 21, 22, 23, 29, 32, 38, 39, 40, 41, 46

Task 5.2.4 - Strategic Mining of Data Anchored to Rodent Atlases (UIO (24), M01 - M24):

This Task has three key objectives:

- To test and validate complete and operative pipelines going from data ingested in the KnowledgeGraph to interpretations of data and use of data for modelling and simulation;
- To clarify needs for quantitative data analysis and determine how such analyses should be conducted;
- To provide Use Case examples yielding strategic data that can be fed into predictive analytical and modelling Tasks (WP5.6, SP6).

The starting point for the mining of data anchored to atlases will be the requirements and scientific questions relevant for the modelling Tasks (SP6). Using the search tools, including the new tool for spatial search, the KnowledgeGraph will be queried for relevant data that in turn will be submitted to analyses, aimed at producing robust, quantitative interpretations of data content. The data retrieved from the KnowledgeGraph will be very heterogeneous, and focus will be given to determine suitable analytical paths for different categories of image data or other non-quantitative data. Tools to be employed include, but are not limited to, tools available through HBP. The mining of data anchored to atlases will result in user feedback to tools developers and data curation teams. It will represent a final and key validation of the data curation. The feedback will be used to fine-tune tools and procedures in the curation process.

**Product 42-1 Region-wise key characteristics of neuronal morphologies**

Strategic experimental data describing neuronal morphologies will be harvested from the pool of data accumulated in the KnowledgeGraph. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region-of-interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

SGA1 – UC 2, 7, 29, 31, 32, 33, 34, 36, 37, 47, 77

**Product 43-2 Region-wise key characteristics of cellular elements**

Strategic experimental data describing key cellular properties (densities, spatial distributions within sub-regions, and other properties of different cell types) will be harvested from the pool of data accumulated in the KnowledgeGraph. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region-of-interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

SGA1 – UC 2, 7, 29, 31, 32, 33, 34, 36, 37, 47, 77

**Product 44-3 Region-wise key characteristics of subcellular elements**

Strategic experimental data describing key sub-cellular properties (densities, distributions of receptors, ion channels, synapses, *etc.*) will be harvested from the pool of data accumulated in the KnowledgeGraph. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region-of-interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and

SP6.

SGA – UC 2, 7, 29, 31, 32, 33, 34, 36, 37, 47, 77

**Product 45-4 Region-wise key characteristics of connectivity**

Strategic experimental data describing long-range neural connections will be harvested from the pool of data accumulated in the KnowledgeGraph. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region-of-interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

SGA1 – UC 2, 7, 29, 31, 32, 33, 34, 36, 37, 47, 77

Task 5.2.5 - Prediction of Cellular, Synaptic and Connectomic Composition, Distributions and Properties of the Rodent Brain (SKU (24), M01 - M24):

As the coverage of brain areas will be incomplete, the missing information will need to be filled in. The first step for this Task is to quantify the available data in each brain area in terms of a distribution with a limited number of parameters. These parameters can then subsequently be estimated, based on similarity between the area at hand and areas for which these parameters are known, using regression models and other machine learning methods.

Previous analyses have shown that the gene expression data in the Allen Brain Atlas can be used to define this similarity and leads to adequate prediction of the mesoconnectome.

This Task will include single cell transcriptomics data from T1A2 together with synapse densities from T1B2, projections from T1B7, and cell, synapse density and connections from T1B9, to make predictions for cellular, synaptic and connectomic composition, distributions and properties in terms of cell types adapting the approach in. These predictions will be delivered to Tasks T6.2.6 and WP6.3.

**Product 46-1 Synapse Generator Pipeline**

The synapse generator pipeline provides an implementation of the first draft for the brain addressing system, a Deliverable of RUP T5.4.1. It comprises two Components. The first Component is the morphology viewer, called as part of a Jupyter script, which allows for the extraction (and selection) from the NIP (and Neuromorpho.org) of the neurite morphologies relevant for building the connectivity of the long-range projection under study. The second Component is the SynapseGenerator itself, implemented as a toolbox in Matlab, which processes the selected morphologies into a Gaussian mixture model (GMM) and then generates synapse locations according to the GMM densities. The constructed model is validated through the construction of a connectivity matrix, the statistics of the generated dendritic and axonal morphologies and the distribution of synapses across branches, for which subroutines are provided in the SynapseGenerator. Both Products are available from the NIP, together with a tutorial illustrating their use. A draft technical report describing the algorithms is available.

RUP 5.4.1 – UC 76, 77, 38, 43, 34

**Product 47-2 The connectomic composition predictor**

The Task will develop an algorithm to predict cell-type specific long-range projections, exploring and evaluating the performance of machine learning techniques such as random forests, logistic regression and Bayesian techniques to integrate the underlying multimodal and multilevel data. Pilot implementations will be provided in Matlab and/or R for ease of exploration using the recipient Use Cases. More robust implementations of the most effective algorithms will be in Python (and in C for the most computation intensive Components) as part of follow-up work in SGA2. The input data derives from reconstructions of single cell axonal projections of thalamocortical neurons; the Allen mouse gene-expression atlas together with single cell data obtained with RNA-seq and the Allen mouse mesoconnectome. The feasibility of including diffusion tensor imaging (DTI) data could also be explored. Deliverables associated with this Product are a report on the algorithm and a review of the quality of the different underlying data sources as well as a pilot implementation that will be delivered to the NIP as a downloadable Matlab toolbox.

Note 1: the same techniques can also be used to predict cellular and synaptic composition, but given the PMs suggested associated with this Task, development, implementation, evaluation and tool delivery of these Products would not seem feasible, hence that part should be postponed until sufficient additional resources become available in SGA2.

Note 2: this Task relies on Tasks T5.2.2, T5.2.3 and T5.2.4 to provide functional access to third party data.

| SGA1 – UC 76, 77, 38, 43, 34, 5, 66, 67 | | | | |
|---|---|---|---|---|
| Schedule of relevant milestones | | | | |
| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
| MS5.2.1 | KPIs for rodent atlas user adoption | 81 | M10 | |
| MS5.2.2 | 2D and 3D image data from HBP Partners spatially anchored to atlas with metadata in KnowledgeGraph | 81 | M12 | |
| MS5.2.3 | SGA1 and SGA2 Roadmap for integration of HBP and 3rd party rodent data into the rodent atlases. Validated by the CDPs and SP6 (T5.2.2) | 81 | M12 | |
| MS5.2.4 | Prioritized roadmap of properties targeted for prediction development. Validated by the CDPs and SP6 T5.2.6 | 51 | M12 | |
| MS5.2.5 | Re-prioritized roadmap of properties targeted for prediction development based on progress to M15. Validated by the CDPs and SP6 (T5.2.6) | 51 | M18 | |

Table 3: Description for Work Package WP5.3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.3 | **Lead Participant** | Forschungszentrum Jülich Gmbh (P20 JUELICH) | | **Starting:** M01 | **Ending:** | M24 |
| **Work Package title** | Multi-level Atlas of the Human Brain | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 20 | 12 | 81 | 68 | **Total** | | |
| **Participant short name** | JUELICH | CEA | UIO | UPM | | | |
| **Person-Months** | 56 | 24 | 13.5 | 24 | 117.5 | | |
| **Person-Years** | 4.7 | 2.0 | 1.1 | 2.0 | 9.8 | | |

**Objectives**

The goal of WP5.3 is to provide:

- Tools specific to the Human Brain Atlas;
- Spatial alignment of datasets produced in SP2;
- Alignment and/or import of strategic data from external repositories;
- Mapping of external repositories of human data to enable access of datasets through the integrated atlas.

Description of work and role of participant

Task 5.3.1 - Curation of Human Atlas Data (JUELICH (38) M01 - M24):

In this Task, curators will:

- Spatially anchor datasets in the common template for human data, using tools co-designed and developed in WP5.4;
- Save the anchoring data in the KnowledgeGraph, making the dataset discoverable in the Atlas;
- Collect provenance metadata on key datasets;
- Provide access in well-specified standards;
- Pre-process datasets appropriately for the above activities.

**Product 48-1 Whole Human Brain Cytoarchitectonic and Maximum Probability Maps registered and curated**

The JuBrain cytoarchitectonic and maximum probability maps from the RUP will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1- UC 70, 71

**Product 49-2 Human V1 laminar profiles from RUP registered and curated**

Human V1 laminar profile of Figure Ground Segmentation (as produced by SP2 T2.1.8) will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is

today referred to as the "Knowledge Graph").

SGA1 – UC 71

**Product 50-3 Big Brain Release 2015 registered and curated**

The 2015 release of the Big Brain will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 70

**Product 51-4 Infant atlas and major tracts in infant brains registered and curated**

The RUP human dataset "Infant atlas and major tracts in infant brains" will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 71

**Product 52-5 Quantitative human receptor data in selected areas registered and curated**

Quantitative human receptor data, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph"). The same holds for the extended data to be delivered in M12 by MS MS2.2.1.

SGA1 – UC 71

**Product 53-6 Morphologies of selected human neurons registered and curated**

Morphologies of selected neurons, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 71

**Product 54-7 Whole brain connectivity atlas registered and curated**

Morphologies of selected neurons, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 71

**Product 55-8 Human Intracranial Database accessible in the NIP**

An atlas of long and short white matter structural connectivity, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 71

**Product 56-9 Wistar rat brain fibre orientation model registered and curated**

The Wistar rat brain fibre orientation model delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the "KnowledgeGraph").

SGA1 – UC 70

Task 5.3.2 - Alignment and Import of Strategic Data from External Repositories (CEA (24), M01 - M24):

The objective of the Task is to make publicly available datasets discoverable within the Atlas and seamlessly accessible to users through the atlas interfaces, as well as to HPC workflows that run on HBP infrastructure through NIP service APIs, as developed in WP5.5.

Using the tools developed in WP5.4, curators will spatially anchor strategic datasets available in external resources and save those within the KnowledgeGraph.

The Task will either create interfaces to important external repositories and atlases, or produce replicas on the federated HBP storage infrastructure. The initial focus will be to interface to the Allen Human Brain Atlas, using its well-defined API, and to replicate the Human Connectome Project data (connectivity data for ~1200 subjects). The reason for replicating the latter is that

Task 5.4.4 is going to implement HPC workflows for applying established neuroimaging tools to this data. This is not feasible if the data is accessed from a remote site. Importing these two repositories will provide a case study for future (SGA2 and beyond) import of external datasets.

**Product 57-1 Software for external datasets import/integration**

a) Integration via APIs. Data are not duplicated, but rather accessed directly on the Allen/HCP servers. Web-based APIs will be developed to on-the-fly transform metadata and data from original representation to a representation suitable for HBP.

b) Dataset replication. Data are replicated on HBP repositories (exploiting FENIX), and transformed to HBP standard formats for data/metadata. The conversion in this case is offline.

SGA1 – UC 71, 73, 75

The best solution will be evaluated.

**Product 58-2 Human Connectome Project data searchable in NeuroInformatics Platform**

Data from Human Connectome Project are fully accessible and searchable from NIP. Metadata comply with HBP standards. The user can navigate the data using software developed in Task 5.5.2, and can use these data for analysis and/or simulations. Here, the data are replicated on HBP repositories (exploiting FENIX), and transformed to HBP standard formats for data/metadata. The conversion in this case is offline.

SGA1 – UC 71, 73, 75

**Product 59-3 Allen Human Brain Atlas data searchable in the NeuroInformatics Platform**

Data from Allen Human Brain Atlas are fully accessible and searchable from the NIP. Metadata comply with HBP standards. The user can navigate the data using software developed in Task 5.5.2, and can use these data for analysis and/or simulations.

SGA1 – UC 71, 73, 75

Task 5.3.3 - Cross-scale Interactive Spatial Alignment Tools for Partial Volumes (JUELICH (18), UIO (13.5) M01 - M24):

The Task goal is to provide semi-automatic spatial registration of partial volumes into existing human brain templates across scales. Such volumes could be block-wise quantitative images natively acquired in 3D, or image stacks with proper spatial coherence such as region-wise reconstructions from histological sections. For example, it should be possible for a user to spatially align an ultra-high resolution ROI from light-microscopy with the BigBrain model to enrich the atlas with more volumetric details.

Due to larger datasets and higher inter-subject variability, human images require different approaches from rodent images, such as input of pairs of corresponding points for landmark-based registration, an iterative refinement of the registration using additional landmarks, and the possibility to align into reference volumes which do not fit into the client side working memory.

Hence, this Task will develop a web-based tool and associated back-end in close co-development between Jülich and UIO, in order to make a tool that matches the needs as well as possible, while sharing relevant code and functionality with other tools of T5.3.3. In particular, the front-end being developed in Products 60-1 and 60-2 will be shared with Product 87-13.

**Product 60-1 Web-based multi-resolution three-planar viewer for large image volumes**

This product is a web-based component that presents the user with three synchronized orthogonal planar views (coronal, axial, and sagittal) of an image volume, and a cross cursor to pinpoint a particular point within this volume. A multi-resolution data streaming strategy is used, in order to allow for remote visualization of reference volumes that do not fit into the client side working memory. This product will either build on the big data viewer from Task 5.4.3 (Product 89-1), or apply the same 2D web rendering engine as Product 85-11. It will use the HBP image service (Product 101-1) as a back-end for serving the multi-resolution data. It will be used by Product 61-1, as well as Product 87-13, as a component of their user interface for entering the landmarks.

SGA1 – UC 71, 72

**Product 61-1 Selection, management and navigation of many landmarks**

This product will implement a web-based HTML5 interface that allows interactive input and management of many 3D landmarks, i.e. explicit corresponding points in two synchronized views: a view of the incoming partial 3D volume that is being spatially anchored, and a view of the reference template volume. Either volume should be visualized through an image service which supports streaming of multi-resolution data, in order to allow for very large volumes. This product is based

on Product 60-1, which provides three orthogonal planar views into the incoming and template volumes. Development of the front-end will be shared with Product 87-13, which also needs an interface for entering of corresponding 3D landmarks in two volumes.

SGA1 – UC 71, 72

**Product 61-2 Affine transformation estimation from landmarks**

This product will implement a back-end that can derive an optimal 3D affine transformation from a set of explicit pointwise landmark correspondences, entered using the front-end developed in the previous Product 61-1. This is different from Product 87-31, which provides intensity-based registration for whole brain data from rodent. However, both tools will share code where possible, e.g. for actually applying the estimated transformations to the source image.

SGA1 – UC 71, 72

**Product 61-3 Iterative workflow loop for landmark adjustment**

This product will implement a controlled interactive loop, during which corresponding landmarks can be updated, added, or deleted in order to refine the registration. The effect on the registration made visible after each update, and the option is given to the user to re-slice the incoming partial volume in order to align its axes with the axes of the template volume.

SGA1 – UC 71, 72

**Product 61-4 Connection and interoperability NIP services**

This Task will link the alignment tool to the KnowledgeGraph and HBP data services, including a streaming image service (Product 101-1), so that reference templates from the HBP atlas can be naturally accessed.

SGA1 – UC 71, 72

Task 5.3.4 - Machine Learning and Statistical Methods for Modelling Cellular and Subcellular Morphologies (UPM (24), M01 - M24):

This Task will focus on cortex pyramidal human cells and will be done in collaboration with T6.2.1 and T1.2.1.

The first goal is to develop machine-learning models to identify morpho-electrical neuron types using unsupervised classification and multi-output regression. To refine these types, the hypothesis of homogeneity of these neuron types across the various cortical layers will be tested. Then pyramidal types per layer might be obtained and, consequently, supervised classification models that discriminate among layers can be derived.

The second goal is to model the geometry of human pyramidal cells. This involves: (a) Hierarchical basal arborisation models covering from its orientation to the detailed dendritic branching of each arbour. The models will be specially designed Bayesian networks to capture how these shapes grow radial- and depth-wise. Exact and approximate inference methods will allow the response to any query to be retrieved. Sampling from the Bayesian network models will generate virtual human pyramidal cells; (b) Repairing models of incomplete basal dendrites. These models will be "encapsulated" Bayesian networks equipped with (exact and/or approximate) inference techniques and imputation methods based on the Expectation-Maximization algorithm; (c) Analysis of dendritic wiring optimality. Heuristic optimization based on graph theory and degree- and role-constrained minimum spanning tree algorithms will be developed to test whether wiring economy is related to the way in which dendrites grow.

A third more fine-grained goal is to provide a morphological classification of dendritic spines from a probabilistic mixture-based clustering approach that will additionally be able to simulate virtual spines from each class.

This Task will collaborate with SP6 (T6.2.1 Models of human cells), with data generated in SP2, and with SP1 (T1.2.1 The pyramidal neuron of the mouse and human). In addition, SP9 can take advantage of this Task when developing the neuromorphic chip of a pyramidal cell.

**Product 62-1 3DSynapsesSA**

An R package to analyse and visualize the three-dimensional spatial distribution of a set of synapses has been developed. It includes an embedded web-based Graphical User Interface and technical documentation.

RUP 5.4.2 – UC 33

**Product 63-1 3DSomaMS**

An R package to characterize a soma according to morphological features taken from its image reconstruction using a mathematical model. It includes a web-based Graphical User Interface and

technical documentation.

RUP 5.4.2 –UC 36, 37

**Product 64-3 3DPyrStructure**

An R package to analyse and reconstruct the basal dendrite arrangement and its particular dendritic trees based on a hybrid Bayesian network model. It a web-based Graphical User Interface and technical documentation.

RUP 5.4.2 – UC 34, 35

**Product 65-5 Single cell arborisation model**

Hierarchical Bayesian model of the basal arborisation morphology built from three-dimensional pyramidal neuron reconstructions.

SGA1 – UC 33

**Product 66-5 Incomplete basal dendrites repairing model**

Data-driven repairing model that detects cut-points in the basal arborisation and then repairs them using a growth model built from complete three-dimensional neuron reconstructions.

SGA1 – UC 32

**Product 67-6 Spine morphology feature extractor**

Matlab module that, given a three-dimensional spine reconstruction, computes a set of characteristic morphological measures that unequivocally determine the spine shape.

SGA1 – UC 30

**Product 68-7 Spine morphology clustering**

Probabilistic clustering model that discovers groups of spines according to their morphological features.

SGA1 – UC 30

**Product 69-8 Spine morphology simulator**

R package that samples from a probabilistic model to create realistic 3D virtual spine representations that match morphologically to the real ones.

SGA1 – UC 30

**Product 70-9 Morpho-electrical neuron types Component**

Model that predicts unseen neuron electrophysiological variables from the set of morphological ones. The model has to be validated via regression performances. Apart from the prediction Task, we want to learn models that describe relationships among variables, giving thoughtful insights in neuron behaviour.

SGA1 – UC 31

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.3.1 | List of KPIs for internal and external user adoption, of Products/services offered by WP5.3 | 20 | M10 | |
| MS5.3.2 | Atlas registration of 3rd party datasets | 12 | M18 | |
| MS5.3.3 | Integration of HCP data | 20 | M24 | |
| MS5.3.4 | Tool for manual partial volume alignment | 20 | M24 | |
| MS5.3.5 | Integration of initial HBP human atlas | 20 | M24 | Produce living |

| | | | | |
|---|---|---|---|---|
| | templates, maps and datasets | | | milestone |
| MS5.3.6 | SGA1 and SGA2 roadmap for integration of HBP and 3rd party human data into the human atlases. Validated by the CDPs and SP6 (T5.3.1 and T5.3.2) | 20 | M12 | |

Table 4: Description for Work Package WP5.4

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.4 | **Lead Participant** | Universitetet I Oslo (P81 UIO) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Data and Atlas Curation Tools | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 1 | 20 | 81 | **Total** | | | |
| **Participant short name** | EPFL | JUELICH | UIO | | | | |
| **Person-Months** | 44.9 | 68 | 48 | 160.9 | | | |
| **Person-Years** | 3.7 | 5.7 | 4.0 | 13.4 | | | |

**Objectives**

This WP will deliver and maintain a suite of tools to enable the development of multi-level atlases. In addition, this WP delivers and maintains tools for general SP5 user support in WP5.7. This includes tools and procedures to assign anatomical location (spatial metadata) to experimental data produced by other SPs, tools to register and curate metadata in general, and 2D and 3D viewers to query and explore the atlases using common APIs. To ensure maximum modularity and reusability tools will be co-designed to fulfil both rodent atlas and human atlas functionalities when possible. Co-design applies in particular to metadata curation tools and spatial anchoring tools. Methods to harvest metadata from third party repositories will be developed to make more data discoverable through the atlases. Tools of specific relevance for the Human atlas are specified in WP5.3.

Description of work and role of participant

Task 5.4.1 - Tools for Metadata Curation (EPFL (41.9), M01 - M24):

This Task will develop, deploy and maintain software essential to data discoverability and accessibility across the HBP Platforms. It will use open source code where appropriate, and the software will be released as open source from the first code commit. User Support based on these tools is provided in WP5.7. Input from WPs managing various categories of data (WP5.2, WP5.3, and WP5.7) will be used to develop unique metadata schemas for each data category. The following functionalities will be provided:

- A WebApp, and APIs, for Meta-Data Annotation of Data to be saved in KnowledgeGraph;
- A WebApp, and APIs, for Curation of Parameters;
- A webclipper, for manual parameter extraction from web resources;
- A WebApp, and APIs, for Ontology Management and expansion;
- A WebApp, and APIs, for a Data-Standards Catalogue;
- A WebApp, to search and navigate HBP, and community, Data and Parameters. Users will be able to put a “Shopping Basket”, to build data and model collections, which can then be shared, or exported for processing/display in the HBP Platforms, and standalone apps;
- A WebApp, and APIs, to easily deposit and access data and models contributed to the Neural Activity Resource (NAR) (developed with T5.7.2) as well as the underlying database software infrastructure.

**Product 71-1 Registration in KnowledgeGraph**

Datasets are annotated with high-level metadata and stored in a database.

RUP 5.6.1 – UC not specified

**Product 72-2 Data Workbench (API, WebApp, MetaData DB)**

Data Workbench allows users to enrich their data artefacts with metadata, to make them discoverable via a search interface (API and web). While annotating the data, users will contribute to the development of ontologies where concepts are missing, and feed a data-standard database.

SGA1 – UC 1-26, 28-47, 50, 52, 53, 57, 59-79

**Product 73-3 Parameter Workbench (API, WebApp, MetaData DB)**

Parameter Workbench allows users to capture parameters annotate them with meta, using the same ontology as the data Workbench, to make them discoverable via a search interface (API and web). The Parameter Workbench will also allow users to discuss with other users the likelihood that parameters reported in the literature and online databases can be taken at face value, or how they might need to be adjusted to reflect the current state of knowledge.

SGA1 – UC 2, 3, 6, 34, 35, 39, 65, 66, 79

**Product 74-4 Neural Activity Resource Development (API, WebApp, MetaData DB)**

Products produced in T5.1.1 and operationally supported by the work in T5.7.1. The development in T5.8.2 is the in-depth adaption of the resource together with the users.

SGA1 – UC 8 – 28, 62, 63, 79

Task 5.4.2 - Integrating 2D Atlas Viewers and Manual Spatial Registration Tools (UIO (48), M01 - M24):

This Task will include further development of a key tool from the RUP, QuickNII, for anchoring of 2D experimental image data to 3D reference atlases. New functionality will increase the speed of the anchoring process through optimized user interfaces and propagation of anchoring information through large series of images.

A new product for 2D to 2D non-linear warping will be added and considered for inclusion in QuickNII. This will allow the user to add a second step to the anchoring process, achieving a higher level of precision when required.

As a first step towards making use of the anchoring information for interpretation of data and further analysis, the viewer LocaliZoom will be developed. This viewer will make full use of the advantages of 3D atlasing by allowing simultaneous viewing of individual section images with overlay of the relevant atlas cut plane. The tool will also include functionality for reading spatial coordinates from the reference atlas and adding annotations (also in reference atlas coordinate space) to the images.

An updated version of the mesh viewing tool from the RUP, MeshView, will be prepared to ingest output from LocaliZoom and provide functionality for basic viewing of annotations from LocaliZoom in reference atlas space and with simultaneous viewing of 3D mesh representations of brain structures available in the reference atlases. This will allow users to select information from multiple experiments.

Finally, a back-end tool for 3D to 3D non-linear warping will be developed and will thus allow registration of whole brain 3D volumes to the reference atlases. The user will access this tool through the front-end software developed under Task 5.3.3.

Tools developed under this Task will be part of a pipeline starting with the curation of metadata required to perform the anchoring of images to reference atlases and ending with the retrieval of data for viewing, annotation and analysis. All tools will be adapted to the different data categories produced in SP1 and the data standards developed by the Ontology Definition Team (ODT) to be realized in Task 5.1.2 and supported in Task 5.1.1. All tools will use standard APIs to view any image stack registered in the HBP KnowledgeGraph.

**Product 75-1 AligNII, from RUP: online tool for anchoring of 2D experimental image data to 3D reference atlases**

Tool for user guided anchoring / affine registration of 2D image data to volumetric atlas templates in the form of isotropic image volumes in NIfTI format. Using this tool, data are anchored to the Waxholm Space (mouse and rat) or Allen mouse brain reference atlases, facilitating data integration through standardized coordinate systems. The tool is integrated in the web-based Rodent Brain Navigator data system.

RUP 5.1.6 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 76 [No further developments planned for SGA1.]

**Product 76-2 QuickNII, from RUP: standalone tool for anchoring of 2D experimental image**

**data to 3D reference atlases**

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 anchoring of 2D image data to the Waxholm Space (mouse and rat) or Allen mouse brain reference atlases, facilitating data integration through standardized coordinate systems.

RUP 5.1.6 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76 [Further developments under Product 83-9 below.]

**Product 77-3 CutNII, from RUP: custom-angle slicer for 3D reference atlases**

Simplified version of QuickNII, allowing custom-angle slicing for isotropic image volumes in NIfTI format. The software is delivered bundled with pre-processed data from the Waxholm Space (mouse and rat) or Allen mouse brain reference atlases. Grayscale template volume and label volume for brain structures are provided. The application requires 1 gigabyte of memory to run. The interactive viewer allows inspection of anatomical features from non-conventional angles and is suitable for generating customized atlas plates matching the orientation of experimental serial images independent of the angle of sectioning.

RUP 5.1.6 – UC 4, 13, 17, 23, 38, 46 [No further developments planned for SGA1.]

**Product 78-4 MeshGen, from RUP: Tool for generating smooth surface meshes from volumetric segmentation data**

Tool for generating separate mesh files for labels present in a segmented image volume. Standalone JAR executable. Supported input formats are NIfTI; ITK-Snap label files; ILF (Integrated Label File, MBAT compatible); Brain-Map XML files. The tool employs a variant of Constrained Elastic Surface Nets [CESN], which has been modified to avoid intersection between neighbouring meshes. The tool is used to generate mesh sets representing brain structures from the Waxholm Space (mouse and rat) or Allen mouse brain reference atlases.

RUP 5.1.6 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76 [No further developments planned for SGA1.]

**Product 79-5 MeshView, from RUP: Online 3-D surface and custom slice viewer for reference atlas mesh data**

Web-application for interactive viewing a volumetric vector-based meshes generated with MeshGen. The purpose is to allow 3-D surface rendering of brain structures from the available reference atlases and opportunity for cutting the atlas volumes in arbitrary, user defined planes, providing customized atlas plates.

RUP 5.1.6 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76 [Further development under Product 83-9.]

**Product 80-6 Mesh set for Waxholm Space rat brain atlas (from RUP)**

Data set consisting of meshes for brain structures from the Waxholm Space rat brain reference atlas.

RUP 5.1.6 – UC 4, 13, 17, 23, 38, 46 [Initially produced in RUP. New mesh sets will be generated when updated atlas versions become available.]

**Product 81-7 Mesh set for Allen mouse brain reference atlas (from RUP)**

Data set consisting of meshes for brain structures from the Allen mouse brain reference atlas.

RUP 5.1.6 – UC 1, 3, 5, 13, 17, 23, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76 [Initially produced in RUP. New mesh sets will be generated when updated atlas versions become available.]

**Product 82-8 Mesh set for Waxholm Space mouse brain reference atlas (from RUP)**

Data set consisting of meshes for brain structures from the Waxholm Space mouse brain reference atlas.

RUP 5.1.6 – UC 1, 3, 5, 13, 17, 23, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76 [Initially produced in RUP. New mesh sets will be generated when updated atlas versions become available.]

**Product 83-9 QuickNII v 2.0: updated functionality and new procedures for propagation of anchoring information through large series of images**

The next generation of the QuickNII software (v 2.0) will be optimized for faster anchoring of large series of 2D images to the reference atlases. The new version will build on the version from the RUP (product 76-2.) and will include new functionality for propagating spatial transformations across series of sections following anchoring of selected images.

SGA1 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 84-10 Non-linear warping of anchored 2D images to a reference atlas**

Anchoring (affine registration) of section images to a reference atlas is performed using QuickNII (product 83-9). This product will provide a new functionality for enhancing the registration to the reference atlas by applying a 2D to 2D non-linear approach. These non-linear techniques will be based on image intensity features and/or user defined landmarks features. The possibility of using a non-linear technique and the choice of the non-linear technique (intensity-based or landmark-based or both) will be dependent on the type of material and the contrast in the section images. The product will be considered for inclusion in an extended version of QuickNII.

SGA1 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 85-11 LocaliZoom: viewer for series of 2D images with reference atlas superimposed**

Web-viewer tool for viewing of series of 2D images that have been anchored to reference atlases. The tool allows display of the relevant reference atlas cut planes superimposed on the images at a user-defined level of transparency. The tool will have additional functions for graphical and semantic annotation functionality and reading of spatial coordinates (Waxholm Space and Bregma coordinates) for points-of-interest in the images. Spatial coordinates can be exported to MeshView v 2.0 (product 86-12).

SGA1 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 86-12 MeshView v2.0: updated functionality, viewing of annotations from LocaliZoom**

The next generation of the MeshView web-viewer for interactive viewing of volumetric vector-based meshes from reference atlases and cutting of the reference atlas volumes in arbitrary, user-defined planes, providing customized atlas plates. The new version will provide functionality for viewing of annotations from LocaliZoom (product 85-11). MeshView v2.0 thus delivers results aggregated from series of 2D images, anchored to reference atlas using QuickNII and annotated in LocaliZoom.

RUP 5.1.6 – UC 1, 3, 4, 5, 13, 17, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73, 76

**Product 87-13 Non-linear warping of whole brain 3D volumes to reference atlas**

Tool for intensity-based non-linear registration of whole brain 3D volumes to the 3D reference atlases, aimed at managing data sets from rodent brain. For visualizing and controlling the registration, the front-end developed as Product 60-1 and 60-2 of task 5.3.3 for human brain data will be co-developed, integrated and used.

SGA1 – UC 1, 13, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73

**Product 88-14 Transformation inverter**

This software methodology will invert the transformations established by affine and non-linear spatial registration of customized 2D atlas slices to images (using Products 83-9 and 84-10) and apply these to transform 2D images to reference atlas space.

SGA1 – UC 1, 13, 23, 38, 39, 40, 41, 46, 62, 63, 70, 71, 72, 73

Task 5.4.3 - Development of 3D High-Volumetric Interactive Atlas Viewer (JUELICH (36), EPFL (3), M01 - M24):

Provision of a fully web-based, responsive 3D viewer for high-resolution volumetric data that can serve as the default visual interface for exploring high-resolution template models, initially represented by the Big Brain. The viewer will provide:

- navigation through the three planes of a high-resolution dataset at different resolutions;
- semitransparent overlay of the default contrast volume with a parcellation;
- an additional lower resolution 3D view of the complete dataset with a visual indication of the currently displayed cutting planes.

It will initially allow to display image volumes stored on FENIX infrastructure, and ultimately connect to the HBP image service.

In its final form, the 3D viewer will also provide the possibility to perform visual queries by selecting rectangular ROIs, and trigger a spatial query to the knowledge graph according to a previously selected data type (e.g. search for available receptor densities in the areas touched by the selection).

Since the complete dataset could be too large to fit into the client's memory, the solution cannot be purely client-side. The client will only hold a copy of the low-resolution whole-brain overview, and then keep track of the 3D region of interest that the user is looking at in detail (location, size, resolution). Since this region is changing according to user interactions (pan, zoom, etc.), the client will constantly request the minimal amount of image data from the server that is needed to update

the displayed data (multi-resolution tiling and streaming). To allow for a responsive user experience, the infrastructure on the back-end site has to scale with the number of expected users on the client side. Since SGA1 is devoted to the implementation and setup of the client and back-end service however, this Task will only provide operation support for a restricted number of users. Nevertheless, it will take appropriate scalability into account in the technology design. SGA2 will need dedicated resources to setup and run scalable IT operations for the back-end image service.

DPIT recommendation: The researchers will evaluate the alternatives as to own development or using external software by sub-contracting with owners of existing suites like "Atelier3D"/"BrainBrowser" of McGill & NRC-CNRS in the first 2 months.

**Product 89-1 Web-based 3D viewer for navigating the Big Brain in three planes at different resolutions**

Web based viewer for high-resolution Big Brain data with capabilities for interactively panning and zooming the image data in three different planes.

SGA1 – UC 69

**Product 89-2 Extension of web-based 3D viewer for selecting and displaying a parcellation as a semitransparent overlay**

In this extended version, the user will be able to choose a parcellation from a list, which is then displayed as a semi-transparent overlay on top of the original contrast. We assume the parcellation to be given as a labelled (integer) volume dataset.

**Product 90-3 Extension of web-based 3D viewer to perform search queries by interactive selection of a rectangular region of interest**

In this extended version, the user should be able to interactively select a rectangular 3D region of interest, and trigger a spatial search for a specific datatype on the search index. The datatype will be preselected, and the search results will be provided as a list of URLs similar to the NIP search mask. This assumes that a spatial search API is available, otherwise a simple mockup service will be used as a temporary replacement.

**Product 90-4 Integration of web-based 3D viewer into the NIP atlas website**

At the end of SGA1, the viewer will be integrated into the overall atlas website, adopting its look and feel as much as possible to the related web-based tools and connecting it to the start page of the atlases.

Task 5.4.4 - Integration of Neuroimaging Tools (JUELICH (32), M01 - M24):

Provide the pre-processing pipelines necessary to bring large volumes or raw MRI-data into a format that is useful for further analysis.

Make neuroimaging tools available on an HCP environment, integrating them with the spatially anchored datasets and the KnowledgeGraph.

Using these tools on the datasets aligned in WP5.3 will allow production of new derivative data adding information on structure, function and connectivity to the parcellations available in the aligned brain atlases, providing information on inter-individual variability and relationship to behaviour.

Provide query tools that allow researchers in the neuroscience SPs to address their specific questions while at the same time registering the output into the KnowledgeGraph.

This Task will create the necessary software environment and evaluate different pipeline solutions for data pre-processing and initial analytical questions. These first use-cases will help to establish the general framework serving a dual purpose. On the one hand, neuroscience researchers can use the HBP resources to address individual scientific questions. On the other, the derivatives will be integrated into the NIP as well.

**Product 91-1 Pre-processing pipeline for raw neuroimaging data**

Given that most data shared by institutions around the world are provided as raw (DICOM or NIfTi) data and that processing should be standardized across datasets, we will set up the necessary pipelines to process these into datasets that are ready to be used by researchers in the HBP and beyond. This includes implementing pipelines that cover surface- and volume-based structural imaging, resting- and Task-state fMRI as well as diffusion-weighted imaging (both standard and HARDI).

SGA1 – UC 25, 70, 71, 73, 74

**Product 92-2 Implementation of atlas-based analysis pipelines**

We will implement the HCP-based extraction of features from on the Human Connectome data for

| |
|---|
| regions defined by one of the HBP atlases. We expect the first instalment of this to be not yet fully linked to other datasets through the KnowledgeGraph, giving ongoing developments in other Tasks, but will set up the necessary infrastructure that can be linked with the rest of the NIP in SGA2.<br>SGA1 – UC 25, 70, 71, 73, 74 |

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.4.1 | List of KPIs for internal and external user adoption of Products/services offered by WP5.4 | 81 | M10 | |
| MS5.4.2 | Initial workflow integration of spatial curation tools with KnowledgeGraph | 1 | M15 | |
| MS5.4.3 | Tools for connecting data elements to images in atlas space | 81 | M18 | |
| MS5.4.4 | 3D high resolution viewer | 1 | M24 | |
| MS5.4.5 | High-resolution viewer and manual registration tool | 81 | M15 | |
| MS5.4.6 | Preliminary NIP image format standardization document outlining obvious standards and open questions where standards are not clear. Validated by the CDPs and SP6 (T5.3.1 T5.1.1, T5.2.2, T5.4.2, T5.4.3, T5.4.4, T5.5.4) | 20 | M15 | This milestone is postponed to M15 due to prioritizing work on the SP5 work plan and technology assessments for viewer architecture and image service. |
| MS5.4.7 | NIP image format standardization document defining standards for HBP. Validated by the CDPs and SPs 1, 2, 3, 6, 7 and 8 (T5.3.1, T5.3.2, T5.1.1, T5.2.2, T5.4.2, T5.4.3, T5.4.4 and T5.5.4) | 20 | M24 | |

Table 5: Description for Work Package WP5.5

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.5 | **Lead Participant** | Ecole Polytechnique Federale de Lausanne (P1 EPFL) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Community-Driven Neuroinformatics Platform and Infrastructure Operations | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 1 | | | **Total** | | | |
| **Participant short name** | EPFL | | | | | | |
| **Person-Months** | 147.1 + 70 in-kind | | | 147.1+70 (in-kind) | | | |
| **Person-Years** | 12.3+ 5.8 | | | 12.3 + 5.8 | | | |

**Objectives**

The focus of this Work Package is on designing, implementing and operating production-standard core services which the users of the NIP in SP5 and the rest of HBP rely upon. Users will access the Platform services directly or through the Collaboratory. The NIP will consist of a collection of Tools, APIs and Data Repositories. For SGA1, the focus will be on providing production-standard data repositories hosted by HBP data services in FedApp/FENIX and on ensuring that those services support Productive collaborations to curate, analyse, explore and access multi-level brain atlases and datasets which have been integrated therein.

The NIP key services are:

1) The KnowledgeGraph, a set of tools to contribute data and a database storing HBP-Provenance information about curated datasets and simulations;
2) Repositories to store working datasets;
3) Repositories to store released, archival datasets;
4) A search app and web service to discover, navigate and interact with datasets;
5) A catalogue and entry point to assist discovery of tools for atlas building;
6) Standardized image services, which allow interaction and analysis of volumetric and time-series datasets which are too large to move to the user in their entirety.

Description of work and role of participant

Task 5.5.1 - Development and Maintenance of KnowledgeGraph (EPFL (55.2), M01 - M24):

The KnowledgeGraph is a database storing experimental provenance metadata as well as spatial anchoring of datasets into atlas templates. The database is accessed through REST APIs. The work will ensure continuous development of the API and the underlying database.

Initial priority will be given to changes required to accelerate the integration of data produced in the RUP by data integration teams in WP5.1, WP5.2, and WP5.3. Subsequently, prioritization will shift to simplifying SGA1 and community data integration activities.

As data integration Use Cases are addressed, priority will then shift to adding generalized workflow provenance.

Part of the work is dedicated to the maintenance of the Production database and to verify data quality and consistency, as well as monitoring performances and optimize the software when necessary. Additionally, architectural changes will be undertaken to prepare for a federated knowledge graph in SGA2.

Following the 2-phase development recommendations of DPIT this work will initially evaluate the

KnowledgeGraph Product developed in the RUP by EPFL. This system is a combination of a traditional RDBMS for transactional data ingest. A specialized Extract, Transform, Load (ETL) daemon moves data into both a file-based archive repository and constructs Elastic Search index database. Regardless of the outcome of the first DPIT-prescribed phase, the KnowledgeGraph will continue to be a hybrid system with a database (relational or NoSQL) for initial ingest and non-interactive queries, with Elastic Search as a search index for queries where interactive performance is required.

If the RUP Knowledge Graph architecture is deemed appropriate to continue, the work will focus on extending the data model handled by the ETL daemon and the Elastic Search indexes to support:

- ACLs to control read, write, delete and update semantics on key entities and key properties in the data model;
- workflow provenance;
- metadata to support community repositories;
- metadata for community trust metrics;
- spatial anchoring information;
- additional ontological flexibility needed to achieve the flexibility objectives of tools developed in 5.4;
- Conditional on policy development happening in the Data Governance Working group, the data model may also include metadata needed for Data Governance policy compliance.

If the RUP Knowledge Graph is abandoned, the data models update above will be re-implemented in the new system along with the following:

- Experimental provenance with ontology support.

**Product 93-1 Knowledge Graph Service**

The Knowledge Graph Service is a REST service with a database (relational or NoSQL) storing-provenance metadata as well as spatial anchoring of datasets into atlas templates. This database will be used for data ingest as well as non-interactive queries (the latter may be restricted to NIP internal users and services).

RUP and SGA1 – UC 4, 27, 37

**Product 93-2 Knowledge Graph Elastic Search Index Service**

The Knowledge Graph Elastic Search Index Service as a clustered, scalable search index for queries where interactive performance is required. This Task is responsible for installation, optimization and maintenance of the Elastic Search cluster required to support interactive queries. This service is private and is only publicly accessible through Product 95-1- KG Search API.

RUP and SGA1 – UC 4, 27, 37

**Product 93-3 Knowledge Graph Evaluation Report**

This report will address:

1) Architectural flexibility of the RUP Knowledge Graph;
2) Progress towards transferring operations ownership to the HBP;
3) MoU regarding shared development between HBP and BBP.

**Product 93-4 KnowledgeGraph Federation Architecture Report**

Federating a system like the KnowledgeGraph is no trivial matter, especially at a global scale. This report will be the outcome of a prototyping and evaluation activity undertaken by T5.5.1 and T5.5.2 along with database expertise from T5.5.8. This report will set the architecture and roadmap for development of the federated Knowledge Graph based on concrete use cases and detailed technical assessments. This is crucial groundwork for implementation of such a system in future SGAs when a federated Knowledge Graph is deemed strategically important.

**Product 94-2 Knowledge Graph Python API**

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 Knowledge graph inside and outside the HBP. This API will focus on developer-friendly data preparation of metadata ingestion into the KnowledgeGraph. This library will be open-sourced to facilitate community building.

RUP and SGA1 – UC 4, 27, 37

DPIT Comment:

The Assessment process of DPIT for the KnowledgeGraph is described in the DPIT Notes 1, 2 and 3, the resulting recommendation is described in DPIT WD 3. All documents are listed in the Appendix A1.

Task 5.5.2 - Search Application (EPFL (31.9), M01 - M24):

The Search application provides a REST API to all indexed data, i.e. KnowledgeGraph, Collaboratory, Software Catalog and selected third party repositories. The REST API will provide a customized view of an Elastic Search system which will enforce necessary ACLs. Spatial queries are expected to be limited to bounding box queries with work happening in T5.5.7 to ensure that these queries are scalable and performant. It is expected that further analysis of data registered in the KnowledgeGraph will be done by individual scientific users using scripted analysis workflows leveraging per-use-case analysis libraries. To support such analytics and scripted workflows, these REST APIs will be complemented by Python client APIs.

The other key output planned for this Task is an intuitive search UI for the data stored by repositories throughout the HBP which have been registered in the KnowledgeGraph. The goal will be to iteratively improve the search interface over the RUP to facilitate ever better discoverability of data being curated in WP5.1, WP5.2, and WP5.3.

This Task will work closely with T5.5.1 to ensure that KnowledgeGraph data models are supported by the KG Search Indexer, the KG Search API and the Collaboratory Integrated Search UI.

**Product 95-1 KG Search API**

REST API to discover datasets, collabs, collab contents, people and ontology terms. For public consumption of data stored in Product 93-2, a secure, authenticated, ACL aware, application specific REST API must be developed.

RUP 5.6.1 – UC all except 4, 27, 37

**Product 96-2 Collaboratory Integrated Search UI**

A Collaboratory Integrated UI to allow search for entities in the KG Search API and to navigate to locations where the user can find more information on each of the entities returned by the search. Significant effort will be expended to ensure that the UI is usable and productive for power users and non-power users alike.

SGA1 – UC all except 4, 27, 37

**Product 97-3 KG Search Python API**

To facilitate automation and scripted workflows using the KG this Product will provide a Production grade Python client API for the KG Search API. This API is expected to see heavy use in the Collaboratory's Jupyter notebooks to allow interactive programmatic exploration of the data visible through the KnowledgeGraph Search API.

SGA – UC all except 4, 27, 37

**Product 98-4 KG Search Indexer**

To populate the KnowledgeGraph, a collection of indexers will be delivered for the data sources including, but not limited to:

- Collab service;
- Collab Storage service;
- Knowledge Graph Service.

SGA1 – UC all except 4, 27, 37

Task 5.5.3 - Development and Maintenance of Web Front End to the Platform Services (EPFL (12), M12 - M24):

This Task will provide frontend and back-end web development to ensure high-quality customization and integration of various solutions developed and deployed throughout SP5 into the Collaboratory.

This is necessary because many of the solutions planned for SGA1 will be based on existing 3rd solutions which are tailored to Use Cases other than the Use Cases planned in the HBP. To ensure these 3rd party solutions work together as a productive Platform will require much integration and customization work on both the back-end and front-end. This Task will make use of user testing and monitor web analytics to maintain and improve Platform usability throughout the SGA1.

**Product 99-1 Web site**

A Neuroinformatics public facing website with anonymous access areas will be rebuilt to ensure that documentation, software catalogs, and applications to search for data, as well as atlas viewers can be accessed without registering for an HBP Collaboratory account. Special care will be needed to integrate with the public facing www.humanbrainproject.eu and ensure easy access to the NIP for visitors who are new to the HBP.

RUP 5.6.1 – UC 34, 65, 66, 76, 78

**Product 100-2 UI development support**

For all web applications developed in SP5. This Task will provide UI development support to ensure a consistent look and feel and to ensure usability across SP5's various web interfaces.

SGA1 – UC: contributes to Products of all Tasks in WP5.1 and 5.5 (and selected software Tasks in 5.2, 5.3, 5.4)

Task 5.5.4 - Data and Image Services (EPFL (6), M18 - M24):

This Task will perform the work required to provide a standard image service for all atlasing viewer Use Cases. The image service will deliver volumetric, image and time-series data (2-4D data sets) and geometric data needed to allow T5.4.2 and T5.4.3 to provide Productive atlas exploration tools. The image service will be accessible via REST API and will be deployed at data centres hosting large atlasing datasets. The Image service will also deliver image, volumetric or geometry datasets registered to the reference atlases datasets to the image viewers.

An effort will be made to integrate this service into atlas building tools and workflows, but this is considered a prototype effort in SGA1 and will be subject to data volumes processed by the atlasing workflows, as well as data locality optimizations available in the implemented image service.

This Task will also manage a subcontract to build a prototype Big Data image query and processing capabilities on top of Apache Spark. The target will be atlas building and rich analytic query processing.

**Product 101-1 Image Service**

REST APIs delivering rich URL-indexed image, voxel, time series and geometric data. Transitioning this service to a high TRL service in SGA1 will require delivering a well-documented specification for the API provided by the service. This will also be crucial to allowing groups outside the HBP to build their own clients and to allow more loosely coupled development between the Image Service and the HBP atlases.

RUP and SGA1 – UC 1, 3, 5-8, 10, 11, 13, 16, 17, 20-23, 32, 33, 38-41, 43, 60-64, 66-77, 79

**Product 102-2 Python Image Service Client**

Python client APIs for the Image service to enable workflow scripting and analytic Use Cases. This Product depends in part on a stable, documented rest API as described in Product 1 of this Task.

RUP and SGA1 – UC 1, 3, 5-8, 10, 11, 13, 16, 17, 20-23, 32, 33, 38-41, 43, 60-64, 66-77, 79

**Product 103-3 Data Analytics Service**

REST APIs delivering rich URL-indexed image, voxel, time series and geometric data. This was provided by the Voxel Brain API in the RUP but will be developed in part with an external contractor in the SGA1.

RUP and SGA1 – UC 1,3,5-8,10,11,13,16,17,20-23,32,33,38-41,43,60-64,66-77,79

**Product 104-4 Interactive Data Analytics Notebook Service**

A prototype Big Data image query and processing system on top of Apache Spark. The target will be to prototype work needed for atlas building and rich analytic query processing in SGA2.

SGA1 – UC 1, 3, 5-8, 10, 11, 13, 16, 17, 20-23, 32, 33, 38-41, 43, 60-64, 66-77, 79

Task 5.5.5 - Collaboratory (EPFL (14 + 70 in-kind), M01 - M24):

The HBP Platforms are largely independent collections of software and services. Bringing these together into a Productive ecosystem for scientific projects is the responsibility of the Collaboratory, which is developed and maintained by this Task. To this end, this Task will provide a

basic set of common web services and a central portal for the Collaboratory and Platforms.

One of the key functions requested by Platform developers throughout the HBP is a central web-accessible storage location with REST API access and a web UI. This storage is intended for collaboration and scientific workflow support.

This Task is responsible for feature development, maintenance and extension of this capability in response to feature requests and service usage growth.

Storage capacity will be provided via FENIX/SP7 including European storage providers like EGI. (Very heavy data transfers will be performed in HPC-centric computing jobs in SP7 strictly under FENIX).

Another key function will be the Collaboratory integration of Jupyter notebooks. Jupyter notebooks are a collaboration technology that provides a web-based literate programming tool with execution of Python workloads (other languages are supported, but will not be the focus of this Task). These are used today in the HBP for documentation, interactive instruction, automation and exploration. This Task will seamlessly integrate Jupyter notebooks as Collaboratory Apps and viewers. This Task also configures computing resources for running the execution kernel of the Jupyter notebooks. The following apps and services are the basis for the Collaboratory ecosystem and supported for HBP internal and external developers.

- HBP Identity - OpenID Connect
- Collaboratory
- Chat
- Navigation
- Wiki
- Documentation hosting service
- Team Profiles
- Software and service catalogue for software and service Components developed and deployed throughout the HBP infrastructure
- Additional capabilities may be added in form of independent apps

**Product 105-1 Collaboratory**

REST APIs delivering rich URL-indexed image, voxel, time series and geometric data. Transitioning this service to a high TRL service in SGA1 will require delivering a well-documented specification for the API provided by the service. This will also be crucial to allowing groups outside the HBP to build their own clients and to allow more loosely coupled development between the Image Service and the HBP atlases.

RUP and SGA1 – UC ALL

Task 5.5.6 - Platform Administration, Operation and Validation (EPFL (10), M01 - M24):

The goal of this Task is to operate and support the NIP services in Production and to provide services and support to improve SP5's ability to deliver those reliable services into the Production environment in a predictable way.

To accomplish this goal this Task will provide DevOps services to the Platform development teams in SP5. These services include but are not limited to:

1) Support for SP5's use of VM base image configuration and Docker container infrastructure;
2) Service monitoring infrastructure for SP5 offered services;
3) Security reviews and security patches on base software;
4) Common database services with security and backup policies established;
5) Support for continuous integration services related to software release and deployment;
6) Architectural recommendations to minimize cost and maximize scalability and service resilience;
7) Platform validation.

SP5 will develop and implement a variety of ICT software technologies, with the goal of enabling neuroscience research by integrating these technologies into the NIP. The Platform also comprises architectures and Components from other ICT solution providers. In addition, this Task has the goal of ensuring the usability of the Platform for the concerned research area. The strategy for achieving this goal is to establish a validation process for a selected set of relevant Use Cases.

Validation is achieved by implementing end-to-end solutions for these Use Cases. This approach will not only stimulate the use of the Platforms by the neuroscience Subprojects, but also provide important guidance for the R&D activities within SP5 as well as the on-going build-up and operation of the NIP.

The work will be organized in the following sub-Tasks:

- Selection of Use Cases and detailed definition of requirements;
- Implementation of Use Cases on the NIP;
- Evaluation of the results.

**Product 106-1 HBP Standard Deployment service**

IaaS offering to provide, configure and deploy VMs at the respective sites with optimized access to appropriate Data Services.

SGA1 – UC ALL

Task 5.5.7 - Spatial Search Application (EPFL (18), M01 - M24):

There is a significant amount of spatial analysis involved in the KnowledgeGraph (KG). Simple examples include spatial registration of datasets produced by HBP to common brain atlas templates, annotating neurons with their brain regions, extracting brain (sub-volumes, etc.). Performing spatial search in KG means searching through all available (meta) information that have spatial data associated with it using a query that also expresses an interest in spatial data. For example, given an ad hoc brain region (shape), retrieve every dataset (or its metadata) that is associated with a brain area that overlaps or intersects the region. To make spatial querying scalable with increasing amounts of data, this Task will extend KG with spatial indexing to support the core spatial functionality efficiently. This includes the 3-dimensional range (interval) query that can be used to implement other type of queries (e.g., distance and *k*-nearest neighbour queries). Using the core functionality, NIP users will be able to perform more sophisticated spatial queries via the developed REST API.

**Product 111-1 Spatial Search API**

REST API to perform ad hoc spatial queries in KG.

SGA1 – UC 29,73

**Product 111-2 Spatial Index for KG**

Spatial index (plugin or a separate component in KG) responsible for spatial data organization.

SGA1 – UC 29,73

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.5.1 | List of KPIs for internal and external user adoption of Products/services offered by WP5.5 | 1 | M10 | |
| MS5.5.2 | Workflow provenance support available in Knowledge Graph | 1 | M18 | |
| MS5.5.3 | Knowledge Graph Evaluation Report | 1 | M12 | See product 93-3 description for further details of the intention and |

| | | | | |
|---|---|---|---|---|
| | | | | planned content for this milestone. |
| MS5.5.4 | Knowledge Graph Federation Architecture Report | 1 | M24 | See product 93-4 description for further details of the intention and planned content for this milestone. |
| MS5.5.5 | PERT chart showing global HBP data-related work breakdown structure and prioritized workflows for SP5 in interaction with the entire HBP | 1 | M12 | |
| MS5.5.6 | Update of PERT chart showing global HBP data-related work breakdown structure and prioritized workflows for SP5 in interaction with the entire HBP | 1 | M18 | |

Table 6: Description for Work Package WP5.6

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.6 | **Lead Participant** | Ruprecht-Karls-Universitaet Heidelberg (P47 UHEI) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Data Mining and Analysis Neuroinformatics Capabilities | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 1 | | 45 | | **Total** | | |
| **Participant short name** | EPFL | | UHEI | | | | |
| **Person-Months** | 24 | | 36 | | 60 | | |
| **Person-Years** | 2.0 | | 3.0 | | 5.0 | | |

**Objectives**

The main goal of WP5.6 is to bring learning-based image analysis to the neuroscientists of the HBP. While generic standalone tools for such analysis are already available, lack of integration into the NIP limits their use in the HBP to pilot studies. This is especially true for large-scale processing, where the sheer amount of bookkeeping very quickly becomes prohibitive. Besides, the vast majority of recent algorithms developed by the image processing community are not distributed in a form accessible to experts from other fields.

The ilastik toolkit is a laudable exception to the status quo: it is a simple, user-friendly tool for interactive image classification, segmentation and analysis. Most important analysis operations are performed lazily, which enables targeted interactive processing of data sub-volumes, followed by complete volume analysis in offline batch mode. Using it requires no experience in image processing. While ilastik has been used successfully for analysis of images, ranging from electron micrographs to telescope signals, one of its main target applications has always been neural image processing. For instance, the first algorithms for synapse detection in Electron Microscopy images have been implemented as ilastik workflows. The FlyEM project extensively uses ilastik for semantic segmentation of EM image stacks.

With this work, we plan to enable the HBP neuroscientists, with or without image processing and parallel computing experience, to analyse enormous data volumes produced by modern imaging techniques.

We propose to tackle this problem by fully and seamlessly integrating ilastik with the existing HBP infrastructure and extending it with several cutting-edge algorithms, particularly well-suited for the task of neural image analysis.

Expected outcomes:

- ilastik will be able to read and write spatially anchored data and preserve HBP metadata throughout the processing, eliminating the need for bookkeeping on the user side. Besides the standard image output, we will extract objects and statistics specific for the HBP Use Cases and make them interoperable with the 3D viewer (5.4.3).
- User annotations, provided to train ilastik classifiers, will be stored in the KnowledgeGraph, along with all necessary provenance tags. This will allow re-use of annotations in other machine learning-related WPs.
- Interactive training of ilastik classifiers will be made available through the NIP web-based image viewer. This will greatly simplify exploratory image analysis, since users will no longer have to find a representative part of the data or transfer large amounts of data to local storage.

- Trained classifiers will be spatially anchored and stored along with the image data. The users will then be able to build on the work of others and reduce the amount of time dedicated to manual annotation.
- Algorithms for automated segmentation of tree-like structures and multi-scale centreline detection will be made available as ilastik workflows. Following ilastik philosophy, they will be made as "lazy" as possible.

Description of work and role of participant

Task 5.6.1 - Interactive Feature Classification and Extraction for Spatial Images (UHEI (24), M01 - M24):

Currently, ilastik is either used on a local user machine in an interactive session or on a large compute server or cluster in the so-called "headless" mode. The former – interactive – strategy is generally used for training, while the latter is most convenient for processing large amounts of data with an already trained model. With this Task, we will extend the existing functionality and enable the users to run ilastik via the NIP. The Task can be separated into two parts: processing only and interactive training.

For the first part, the NIP will offer the user a choice of already existing classifiers for a given dataset and allow upload of a new classifier. The algorithm will then process the dataset remotely, near the location of the data, using and building on the resources and tools for parallel processing, provided by SP7 and FeDaPP.

For the second part, we will add the option to replace the "volumina" viewer of ilastik by the web-based viewer, developed as a part of NIP (Tasks 5.4.2 and 5.4.3). This approach offers many advantages over the status quo: since the viewer is multi-scale, the user does not have to send the full-resolution data over the network or copy it locally. They can instead annotate low-resolution data and perform the algorithm training with full resolution, on a machine physically close to the data storage. Since the HBP viewer architecture is not yet finalized, this task will require close coordination with other viewer-related tasks in SP5. The programmers on this task be involved in the development of the Viewer Architecture Document.

Storage and further use of ilastik results, as well as of the user annotations, is the subject of T5.6.2.

**Product 109-1 Connection of ilastik to HBP 2D and 3D viewers**

SGA1 – UC 3, 5, 76

Task 5.6.2 - Workflow for Populating Brain Atlases with Features, Automatically Extracted by ilastik (UHEI (12), M12 - M24):

Currently, ilastik results are returned in the form of images. For the Object Classification and Tracking workflows it is also possible to export the properties of resulting objects. To achieve seamless integration with the other tools of the Platform, additional post-processing will have to be developed to convert labelled images of ilastik into objects, expected by the Knowledge Graph. Similarly, new pre-processing applets will ensure that ilastik takes the dataset spatial anchor into account, maintains it all through the workflow pipelines and transfers the relevant properties into the output objects. This integration is also required in order to connect ilastik with the data serving and computing services of SP7 and FeDaPP.

Besides the conversion and storage of ilastik results, this Work Package will deal with the storage of user annotations. These annotations can serve multiple classifiers and algorithms, such as those developed in WPs 5.1.3, 5.1.4, 2.7.6 or 2.6.4. The second goal of this WP is to ensure that expensive manual annotations are used in all cases where they could be applicable, by tracking provenance of data and correctly anchoring it in the KnowledgeGraph.

**Product 110-1 Connection of ilastik to other HBP services**

As a result, ilastik will be able to work with HBP data and to save its own results and the manual user annotations in a way to be used by the related Tasks.

SGA1 – UC 3, 5, 76

DPIT recommendation: The Task "Analytics Workflow for Large-Scale Volumetric Data, Using ilastik" is recommended to begin in SGA2.

Task 5.6.3 - Integrating Feature Extractors and Classifiers for Neuroscience (EPFL (24), M01 - M24):

Automated detection, segmentation and counting of structures such as cells or synapses, along with delineation of neuron fibres and vasculature in Light Microscopy data, are priorities. While off-the-shelf solutions for some of these problems are already available, state-of-the-art approaches rely on Machine Learning techniques such as Deep Convolutional Neural Networks or Active Learning, for which we have working algorithms. We will tailor them to the needs of HBP Use Cases and make them available to the community in the form of ilastik plug-ins.

In particular, while tools for semi-automated segmentation and object detection are already available, counting is a fundamentally different problem with its own challenges. First, counting by detection is difficult in areas with high cell densities because their cluttered and noisy appearance makes it almost impossible to tell objects apart. Second, even though segmentation can provide an estimate of the volume occupied by groups of objects, it cannot accurately measure densities inside this volume, which is required for an actual count. We have therefore developed regression methods that predict local density based on appropriately designed features. The desired counts can then be obtained by integrating the resulting densities over regions of interest. The current implementations are written in Python. Integrating this code into ilastik will require porting some critical parts to C++ for speed and writing the corresponding Python wrappers.

These methods rely on annotations to learn the models. While semi-automated tools exist, they still require inordinate amounts of manual intervention to handle biologically relevant volumes. We have recently shown that Active Learning techniques can considerably speed up this bottleneck, generating training data with a minimum of manual annotation. Our algorithms show state-of-the-art results on both segmentation and delineation Tasks. We will identify Use Cases that will benefit from these technologies and investigate their integration into ilastik as plug-ins, going towards SGA2.

**Product 112-1 Segmentation tools**

We developed new, context-based tools for the automatic and semi-automatic segmentation of synapses and mitochondria in EM images. These algorithms where delivered as a plug-in for ESPINA, which was further improved following the requests and recommendations of the neuroscientists at UPM. We also upgraded our plugin for two subsequent versions of ESPINA that had a new, backward-incompatible API. This work was described in RUP T5.2.2 and advanced as part of EMDigest (RUP T5.2.1, ESPINA Component). This Task is not a direct continuation of the RUP Task.

RUP 5.2.1 – UC 3

SGA1- UC 3,5,33,76

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.6.1 | List of KPIs for internal and external user adoption of Products/services offered by WP5.6 | 47 | M10 | |
| MS5.6.2 | ilastik integration with KnowledgeGraph | 47 | M24 | |
| MS5.6.3 | Retrieve and store classifiers in KnowledgeGraph | 47 | M24 | |
| MS5.6.4 | Integration of feature extractors in ilastik | 47 | M24 | |
| MS5.6.5 | Scientific data service deployed in Production | 1 | M24 | |

Table 7: Description for Work Package WP5.7

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.7 | **Lead Participant** | Forschungszentrum Julich Gmbh (P20 JUELICH) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Tools and Curation for Integrated Parallelized Analysis of Activity Data | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 10 | | 20 | **Total** | | | |
| **Participant short name** | CNRS | | JUELICH | | | | |
| **Person-Months** | 15 | | 25 + 9 from SP9 | 39 + 9 from SP9 | | | |
| **Person-Years** | 1.25 | | 2.1 + 0,75 from SP9 | 3.35 + 0.75 from SP9 | | | |

**Objectives**

The development of novel methods for the analysis of activity data such as massively parallel spike data or local field potentials (LFPs) is an ongoing research endeavour driven by the need to uncover the principles by which neurons in the network coordinate their activity. A large range of analysis approaches for data recorded on multiple spatial and temporal scales have been proposed, many of them tuned to the specificities of either simulated or experimental data.

This Work Package will provide the framework to host analytic tools. It will continue the development of the Tools for the Analysis of Functional Data, which feature the Electrophysiology Analysis Toolkit (Elephant) as a central Component.

Particular focus will be placed on providing support for advanced analysis methods that are frequently used in the analysis of experimental and simulated data sets. Furthermore, the team will begin to implement the integration of the tools into the parallelization frameworks made available by the HBP infrastructure. To this end, the Task will work closely with T7.5.4 “Application Software Services”.

The temporal dimension (neural dynamics) has been largely lacking in the Neuroinformatics Components of HBP. It cannot be sufficiently addressed via other SPs. The experimentalists (SP1, SP2, SP3, and possibly also SP8) will post data collected on neural activity, and these need to be curated. The modellers (essentially SP4 but also SP6 and partly SP3) will post their models and the activity maps they produce. The work under this Task will enable comparisons to the activity maps recorded experimentally and by in-depth analysis. There will also be a contribution from the neuromorphic researchers (SP9) and neuroroboticists (SP10), who may post neuron-like activities simulated by their devices. The goal will to establish a dataflow all the way from experiments to models, along with implementations in neuromorphic and neurorobotic systems.

Elephant will be the analysis component in use cases built in conjunction with electrophysiology data components provided by T5.7.2. Methods contained in the Elephant tool are adequate to analyse data that arise in SGA1, in particular in the context of T4.5.1 and T9.1.5.

Description of work and role of participant

Task 5.7.1 - Elephant (JUELICH (12 + 9 from SP9), M01 - M24):

This Task will deliver maintenance, continued development and integration of available and new Tools for the Analysis of Activity Data. Development will focus on making available additional advanced methods for the analysis of spike and mesoscopic/macroscopic time series, providing experimental support for parallelization of selected analysis methods on HPC hardware, and adding functionality to aid the handling and manipulation of the corresponding data objects during the

analysis workflow.

Tasks validating models against other models from network simulation and robotics (e.g., T6.2.7 and T9.1.5 and WP10.1), models against experimental data (e.g. WP4.2 and WP4.5 or also WP3.1, 3.2, 3.3, 3.4 and 3.5), and by analysis routines required for experimental data in CDP4 and CDP5.

This Task will develop data interfaces to file formats that enable the tools to access simulation data supplied by Use Cases of T6.2.7, T9.1.5, and WP10.1, as well as UCs, e.g. from SP3 like UC 8-24 and 79 and experimental data, in particular the data supplied in CDP4 and data used for validation in other Tasks (T4.5.1) and CDP5. In particular, the Task will build an interface to access data from the NEST simulation engine (T6.2.7) or the data coming from neuromorphic hardware (T9.5.1) to the Tools for the advanced Analysis of Activity Data.

Support to integrate the Tools for the Analysis of Activity Data in multiple Use Cases (in particular, co-design with CDP4, CDP5 and UC 8-24 and 79). To this end, the tools will be fully integrated into the Collaboratory (so far T11.5.3 and T11.5.5, in future T5.5.5) and into HPC-enabled analysis processes based on the UNICORE framework (T7.5.4 and T7.5.6) in a fashion that satisfies the analysis requirements set forth by the Use Cases.

**Product 113-1 Elephant**

Elephant data analysis library. The source code and tests are contained in a Git repository hosted on GitHub.

RUP 5.3.1 – UC 8-24, 25, 26, 28, 54, 60, 61, 79

**Product 114-2 Elephant Code Repository**

The source code and tests for the Elephant data analysis library are contained in a Git repository hosted on GitHub.

RUP 5.3.1 – UC 8-24, 25, 26, 28, 54, 60, 61, 79

**Product 115-3 Expanded and improved release of Elephant**

Elephant data analysis library with additional functionality created during SGA1 and with experimental support for parallelization of selected analysis methods. The source code and tests are contained in a Git repository hosted on GitHub.

SGA1 – UC 8-24, 25, 26, 28, 54, 60, 61, 79

Task 5.7.2 – Neural Activity Resource (CNRS (15), JUELICH (13), M01 - M24):

Task 5.7.2 will set up a "neural activity" resource in SP5, to the development of which all of HBP is expected to contribute. This has the background that physiological and functional data have to be integrated into databases and analysed, but also linked to HBP atlases considering the specific topography of the structural correlates of such activity data. These activity data cover different temporal and spatial scales: from milliseconds to hours and from the cellular to systems level. Presently, HBP atlases include those for mouse, rat and human brains with first projects preparing atlases of other species (e.g. monkey, ferret).

- To initiate the deep integration of neural activity data, this Task will make use of a basic database of T5.4.1, adapted by T5.7.2 to handle and analyse neuronal activity maps or signals, at all spatial and temporal scales, from the cellular (single-unit), networks and ensembles, i.e. groups of single-units, recorded with multi-channel Ephys probes, (voltage-sensitive dyes, calcium dyes, local field potentials), ensembles in multiple areas, up to large scales (ECoG, cFOS, fMRI, EEG, MEG, behaviour and cognition).
- Secondly, it will link these data to existing and future HBP atlases.
- Thirdly, it will develop and apply a scenario of how incoming and existing activity data will be stored and curated in such a way that they can be included into HBP databases and atlases.

As a proof of concept, fMRI data will be linked to the human brain HBP atlas, electrophysiological and c-Fos data to the mouse brain atlas. Production level will be achieved in SGA2.

Due to the high level of heterogeneity of data types and experimental scenarios, it is imperative that data are accompanied by machine-readable, comprehensible and detailed descriptions of the metadata accumulated during the experiment or the simulations. To identify technical and conceptual solutions to this problem, a second part of this Task is to supply selected datasets posted to the database with a comprehensive metadata collection using emerging community standards (e.g. open metadata mark-up language (odML) for electrophysiological data). Metadata organization for simulated data needs to be newly developed. Furthermore, the Task will make use

of or provide interfaces that enable access of metadata from the registered datasets through the data curation and Knowledgegraph frameworks.

**Product 116-1 Neural activity resource**

For integration and analysis of neural dynamics data (ECoG, fMRI, cellular, ensembles, model results, behaviour, etc.).

SGA1 – UC (for SP3) 8-24 and 79, (for SP4) 25-28, 62, 63, 79

**Product 117-2 Linking NAR to human and mouse brain atlas**

Linkage between Neural activity resource to Human and Mouse brain atlas. SGA1 – UC (for SP3) 8-24 and 79, (for SP4) 25-28, 62, 63, 79

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.7.1 | List of KPIs for internal and external user adoption of Products/services offered by WP5.7 | 20 | M10 | |
| MS5.7.2 | Prototype version of neural activity resource | 10 | M24 | |
| MS5.7.3 | Additional functionality of Elephant validated | 20 | M24 | |

Table 8: Description for Work Package WP5.8

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Work Package No.** | WP5.8 | **Lead Participant** | Universitetet I Oslo (P81 UIO) | **Starting:** | M01 | **Ending:** | M24 |
| **Work Package title** | Management and Coordination | | | | | | |
| **Activity Type** | Research & Innovation | | | | | | |
| **Participant number** | 1 | 36 | 81 | **Total** | | | |
| **Participant short name** | EPFL | KI | UIO | | | | |
| **Person-Months** | 5.9 | 39 | 36 | 80.9 | | | |
| **Person-Years** | 0.5 | 3.3 | 3 | 6.7 | | | |

**Objectives**

This WP manages SP5, the NIP. It coordinates and validates the software and infrastructure development in the Subproject, ensuring that (i) the work is aligned with the overall objectives of the Project, (ii) meets actual user needs, (iii) is efficiently organised and documented, and (iv) takes care for specific Neuroinformatics collaboration and community building.

Description of work and role of participant

Task 5.8.1 - Subproject Management and Technical Coordination (UIO (36), EPFL (5.9), KI (3) M01 - M24):

This Management and Technical Coordination Task (MTC) manages SP5, the NIP. It will be carried out by the Subproject management team in close collaboration with the responsible Task for "Platform Administration and Operation", T5.5.6.

The MTC is responsible for resource allocation and use, performance and risk management, internal review and quality control.

The MTC coordinates the technology and infrastructure development in SP7. These responsibilities include:

- Assessment of TRLs of software and services;
- Monitoring and reporting of technical and infrastructure progress to STC;
- Documentation and dissemination of the NIP's technical and operational standards;
- Testing and quality control of software and infrastructure developed in SP5;
- Implementation of infrastructure Components according to Project Lifecycle as described in the FPA;
- Representation of the SP on the Technical Coordinators Committee and the Infrastructure Coordinators Committee;
- Interaction with FENIX and European RI providers;
- Attendance of conferences/events on behalf of the HBP, participation in HBP booths.

The MTC triggers and verifies reporting and Deliverable writing.

The MTC is responsible for organizing the internal communication within the Subproject (through the WP leader and all-hands meetings, regular audio/video conferences, mailing lists and internal web sites), as well as between the Subproject and the central Project Coordination Office and other relevant bodies of the HBP governance.

The MTC monitors scientific progress in SP5.

The MTC organizes SP outreach and communication activities and liaises with the central

communications team to produce dissemination material relevant for all HBP activities. It also coordinates the planning of SP5 activities for the next project phase under SGA2.

The MTC coordinates with the External Relations Team on issues related to innovation.

The MTC provides input from SP5 for innovation reports and liaises with the central innovation management.

The MTC also acts as ethics rapporteur for SP5, who reports technical, social or ethical issues to the HBP Ethics Manager or to the Ethics Advisory Board, and provides support to Partners on issues related to administration, innovation and ethics.

Task 5.8.2 - Neuroinformatics Community Building (KI (12), M01 - M24):

This Task will be responsible for strengthening the relationships with the other brain initiatives and international brain research organisations, and facilitation of information exchange between them. The Task will coordinate and support the operational community building conducted in the NIP, the execution of which is the responsibility of the WP leaders of 5.1-7. The work will include:

- Acting as liaison between HBP and INCF concerning in particular KnowledgeSpace;
- Organization of workshops to promote the broad usage of Neuroinformatics in the community;
- Coordinating communications towards participation of different neuroscience communities in Europe;
- Assisting the SIB chair for communications worldwide including interaction with the other brain initiatives.

Task 5.8.3 - Development and Maintenance of KnowledgeSpace (KI (24) M01 - M24):

KnowledgeSpace is a community encyclopaedia and consists of interactive interface and data-driven taxonomies, Production ontology services, data type.

**Product 118-1 Knowledge Space – Service**

RUP – UC n/a

**Product 119-2 Knowledge Space – Software**

RUP – UC n/a

**Product 120-3 Cell type ontologies**

Per brain area: hippocampus, cerebellum, basal ganglia, cortex - including key molecular markers and property-based cell ontology for each region.

SGA1 – UC 36, 37, 38, 39, 40, 41, 43

**Product 121-4 Integrated data sources**

Integration of additional federated data sources to support the modelling Use Cases; Search results for current federated data sources tuned to support modelling Use Cases; Subscribe/notify for concepts - receive notifications when new data becomes available in the searchable dataspace

SGA1 – UC 36, 37, 38, 39, 40, 41, 43

Schedule of relevant milestones

| Milestone number | Milestone name | Lead participant number | Delivery date from Annex I[11] | Comments |
|---|---|---|---|---|
| MS5.8.1 | List of KPIs for performance evaluation of coordination, outreach and adoption of Products/services offered by WP5.8 | 81 | M10 | |
| MS5.8.3 | SP5 roadmap for SGA2 | 81 | M13 | |
| MS5.8.4 | Report on community building | | M18 | Report of communications towards the HBP community, European neuroscience |

| | | | | |
|---|---|---|---|---|
| | | | | communities and worldwide to promote the understanding of the goals of HBP and engagement for the usage of the NIP. |

Table **9**: Deliverables of SP5 in SGA1

| List of WP Deliverables | | | | | | | |
|---|---|---|---|---|---|---|---|
| Deliverable Number | Deliverable Title | Lead Partner ID | Estimated indicative person-months | Type | Dissemination level | Delivery date | |
| SGA1 D5.6.1 | SP5 work plan for SGA1 (DPIT) | DPIT & UIO (P81) | n/a | R | PU | M06/ M10 | Originally submitted 10 Oct 2016. EC Feedback from Review of 24 Oct 2016 required resubmission by M10 (this document) |
| SGA1 D5.6.2 | HBP overall Platforms Architecture (DPIT) | DPIT & EPFL (P1) | n/a | R | PU | M06 | Originally submitted 10 Oct 2016. EC Feedback from Review of 24 Oct 2016 required resubmission by M12 |
| SGA1 D5.8.1 | Strategy and Architecture for HBP data viewers | UIO (P81) | n/a | R | PU | M12 | New Deliverable requested by EC Feedback from Review of 24 Oct 2016. |
| SGA1 D5.8.2 | SP5 Neuroinformatic Platform - Results for SGA1 Period 1 | UIO (P81) | n/a | R | PU | M12 | Formerly SGA1 D5.6.3, renumbered to reflect new SP5 structure introduced by this document.<br>Summary of SP Results for M1-12, broken down by Task:<br>• Actual vs. planned results (due & done/due & not done)<br>• Links to Components which the Task contributes to<br>• Results passed to another Task, WP or SP (due & done/due & not done)<br>• CDP products (due & done/due & not done)<br>• Milestones due and achieved, how validated, by whom<br>• Milestones due but not achieved, why, impact, corrective action<br>• Publications<br>• Per-Component status reports with linkages to contributing tasks (PLA export)<br>a. Summary of improvements made<br>b. Input Components used by this Component<br>c. Co-design projects using this |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | Component<br>d. TRL status<br>NOTE: In order to report fully on the above topics, the Deliverable may need to include information that we would not wish to make public. This material will be included in the Deliverable for the EC and Reviewers, but not in the version published on the HBP public website. COULD INCLUDE: Research results, tools and methods and contribution to initiate the platforms (for Period 1) |
| SGA1 D5.8.3 | SP5 Neuroinformatics Platform - Results for SGA1 Period 2 | UIO (P81) | n/a | R | PU | M24 | Formerly SGA1 D5.6.3, renumbered to reflect new SP5 structure introduced by this document.<br>NOTE: This document will now cover the topics previously assigned to Deliverable SGA1 D11.3.4 (Summary of Collaboratory improvements made in SGA1). D11.3.4 is now deleted from the SGA1 work plan.<br>Summary of SP Results for M13-24, broken down by Task:<br>• Actual vs. planned results (due & done/due & not done)<br>• Links to Components which the Task contributes to<br>• Results passed to another Task, WP or SP (due & done/due & not done)<br>• CDP products (due & done/due & not done)<br>• Milestones due and achieved, how validated, by whom<br>• Milestones due but not achieved, why, impact, corrective action<br>• Publications<br>• Per-Component status reports with linkages to contributing tasks (PLA export)<br>a. Summary of improvements made<br>b. Input Components used by this Component<br>c. Co-design projects using this Component<br>d. TRL status<br>NOTE: In order to report fully on the above topics, the Deliverable may need to include information that we would not wish to make public. This material will be included in the |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | Deliverable for the EC and Reviewers, but not in the version published on the HBP public website. INCLUDE: Research results, tools and methods and contribution to initiate the platforms (for Period 2) |

Table 10: Previous SP11 Deliverables planned for resubmission by SP5

| List of WP Deliverables | | | | | | | |
|---|---|---|---|---|---|---|---|
| Deliverable Number | Deliverable Title | Lead Partner ID | Estimated indicative person-months | Type | Dissemination level | Delivery date | |
| SGA1 D11.3.1 | HBP Collaboratory Architecture | EPFL (P1) | | R | PU | M06/ M12 | Originally submitted 10 Oct 2016. EC Feedback from Review of 24 Oct 2016 required resubmission by M12. Under the current SP5 work plan, this Deliverable is now the responsibility of Task T5.5.5 (Collaboratory) |

# 3. Product Breakdown

The following list is a complete product breakdown of SP5. The list follows the serial number of products as given in section two. The list of products is matched with the list of use cases from DPIT WD 2 Comprehensive Collection of HBP Data Use Cases" as given in DPIT WD 4 "Use Case to Product Matrix".

| | | | |
|---|---|---|---|
| 1-1 | Meta-Data Enriched RUP/SGA1 data and models | T5.1.1 | WP5.1 |
| 2-2 | Living report on adoption of T5.1.1 tools by HBP and external Researchers | | |
| 3-1 | Identification of HBP users' Use Cases | T5.1.2 | |
| 4-2 | Ontology engineering | | |
| 6-1 | Support for data upload and download | T5.1.3 | |
| 7-2 | Support for data transfers | | |
| 8-3 | Support for data searches | | |
| 9-1 | Waxholm Space | T5.2.1 | WP5.2 |
| 10-2 | Waxholm Space rat brain reference atlas, v1.0 | | |
| 11-3 | Waxholm Space rat brain reference atlas, v1.0.1 | | |
| 12-4 | Waxholm Space rat brain reference atlas, v2.0 | | |
| 13-5 | Enriched Waxholm Space rat brain reference atlas: PLI / M2 data | | |
| 14-6 | Atlas of rat brain cyto-and myeloarchitecture | | |
| 15-7 | Whole brain connectivity atlas of the rat | | |
| 16-8 | Volumetric Allen mouse brain reference atlas | | |
| 17-9 | Update of Allen mouse brain reference atlas | | |
| 18-10 | Three-plane high resolution architectonic atlas of the rat hippocampal region | | |
| 19-11 | Standardized anatomical landmarks for registration of whole brain imaging datasets to mouse brain Waxholm Space | | |
| 20-12 | Tg-Arc-Swe atlas | | |
| 21-13 | Neurotransporter atlas: GLT1 | | |
| 22-14 | Neurotransporter atlas: VGLUT3 | | |
| 23-15 | tTA mouse brain atlas: Nop | | |
| 24-16 | tTA mouse brain atlas: PrP | | |
| 25-17 | tTA mouse brain atlas: PrP-CamKII | | |

| | | | |
|---|---|---|---|
| 26-18 | Database on rat cerebro-cerebellar connectivity | | |
| 27-19 | Customized versions of Allen mouse brain atlas tailored for different analyses | | |
| 28-20 | Allen mouse brain reference atlas with white matter structures parcellated | | |
| 29-21 | Waxholm Space rat brain reference atlas enriched with additional receptor data | | |
| 30-22 | Expanded and improved Waxholm Space rat brain reference atlas (with additional and corrected structure delineations) based on registered multimodal data | | |
| 31-1 | Tutorials, training and supervision in assignment of spatial metadata | T5.2.2 | |
| 32-2 | Validation and approval of spatial metadata before final entry in KnowledgeGraph | | |
| 33-3 | Optimized procedure for anchoring of 2D image data to reference atlas | | |
| 34-4 | Procedure for anchoring of 3D image data to reference atlas | | |
| 35-5 | Procedure for non-linear warping of 2D image data to reference atlas | | |
| 36-6 | Procedure for non-linear warping of 3D image data to reference atlas | | |
| 37-7 | Curation of semantic spatial metadata delivered in T5.4.1. | | |
| 38-1 | Metadata and links to Allen Institute mouse brain data repositories (https://www.alleninstitute.org/) | T5.2.3 | |
| 39-2 | Metadata and links to data shared via the mouse connectome project (www.mouseconnectome.org) | | |
| 40-3 | Registration of rat brain hippocampal connectivity descriptions from temporal lobe database (www.temporal-lobe.com) | | |
| 41-4 | Registration of rat brain connectivity descriptions from the BAMS database (https://bams1.org/) | | |
| 42-1 | Region-wise key characteristics of neuronal morphologies | T5.2.4 | |
| 43-2 | Region-wise key characteristics of cellular elements | | |
| 44-3 | Region-wise key characteristics of subcellular elements | | |
| 45-4 | Region-wise key characteristics of connectivity | | |
| 46-1 | Synapse Generator Pipeline | T5.2.5 | |
| 47-2 | The connectomic composition predictor | | |
| 48-1 | Whole Human Brain Cytoarchitectonic and Maximum Probability Maps registered and curated | T5.3.1 | WP5.3 |

| | | | |
|---|---|---|---|
| 49-2 | Human V1 laminar profiles from RUP registered and curated | | |
| 50-3 | Big Brain Release 2015 registered and curated | | |
| 51-4 | Infant atlas and major tracts in infant brains registered and curated | | |
| 52-5 | Quantitative human receptor data in selected areas registered and curated | | |
| 53-6 | Morphologies of selected human neurons registered and curated | | |
| 54-7 | Whole brain connectivity atlas registered and curated | | |
| 55-8 | Human Intracranial Database accessible in the NIP | | |
| 56-9 | Wistar rat brain fibre orientation model registered and curated | | |
| 57-1 | Software for external datasets import/integration | T5.3.2 | |
| 58-2 | Human Connectome Project data searchable in NeuroInformatic Platform | | |
| 59-3 | Allen Human Brain Atlas data searchable in NeuroInformatic Platform | | |
| 60-1 | Web-based multi-resolution three-planar viewer for large image volumes | T5.3.3 | |
| 61-1 | Selection, management and navigation of many landmarks | | |
| 61-2 | Affine transformation estimation from landmarks | | |
| 61-3 | Iterative workflow loop for landmark adjustment | | |
| 61-4 | Connection and interoperability NIP services | | |
| 62-1 | 3DSynapsesSA | T5.3.4 | |
| 63-1 | 3DSomaMS | | |
| 64-3 | 3DPyrStructure | | |
| 65-5 | Single cell arborisation model | | |
| 66-5 | Incomplete basal dendrites repairing model | | |
| 67-6 | Spine morphology feature extractor | | |
| 68-7 | Spine morphology clustering | | |
| 69-8 | Spine morphology simulator | | |
| 70-9 | Morpho-electrical neuron types Component | | |
| 71-1 | Registration in KnowledgeGraph | T5.4.1 | WP5.4 |
| 72-2 | Data Workbench (API, WebApp, MetaData DB) | | |
| 73-3 | Parameter Workbench (API, WebApp, MetaData DB) | | |

| | | | |
|---|---|---|---|
| 74-4 | Neural Activity Resource Development (API, WebApp, MetaData DB) | | |
| 75-1 | AligNII, from RUP: online tool for anchoring of 2D experimental image data to 3D reference atlases | T5.4.2 | |
| 76-2 | QuickNII, from RUP: standalone tool for anchoring of 2D experimental image data to 3D reference atlases | | |
| 77-3 | CutNII, from RUP: custom-angle slicer for 3D reference atlases | | |
| 78-4 | MeshGen, from RUP: Tool for generating smooth surface meshes from volumetric segmentation data | | |
| 79-5 | MeshView, from RUP: Online 3-D surface and custom slice viewer for reference atlas mesh data | | |
| 80-6 | Mesh set for Waxholm Space rat brain atlas (from RUP) | | |
| 81-7 | Mesh set for Allen mouse brain reference atlas (from RUP) | | |
| 82-8 | Mesh set for Waxholm Space mouse brain reference atlas (from RUP) | | |
| 83-9 | QuickNII v 2.0: updated functionality and new procedures for propagation of anchoring information through large series of images | | |
| 84-10 | Non-linear warping of anchored 2D images to a reference atlas | | |
| 85-11 | LocaliZoom: viewer for series of 2D images with reference atlas superimposed | | |
| 86-12 | MeshView v2.0: updated functionality, viewing of annotations from LocaliZoom | | |
| 87-13 | Non-linear warping of whole brain 3D volumes to reference atlas | | |
| 88-14 | Transformation inverter | | |
| 89-1 | Web-based 3D viewer for navigating the Big Brain in three planes at different resolutions | T5.4.3 | |
| 89-2 | Extension of web-based 3D viewer for selecting and displaying a parcellation as a semitransparent overlay | | |
| 90-3 | Extension of web-based 3D viewer to perform search queries by interactive selection of a rectangular region of interest | | |
| 90-4 | Integration of web-based 3D viewer into the NIP atlas website | | |
| 91-1 | Pre-processing pipeline for raw neuroimaging data | T5.4.4 | |
| 92-2 | Implementation of atlas-based analysis pipelines | | |
| 93-1 | Knowledge Graph Service | T5.5.1 | WP5.5 |
| 93-2 | Knowledge Graph Elastic Search Index Service | | |
| 93-3 | Knowledge Graph Evaluation Report | | |

| | | | |
|---|---|---|---|
| 93-4 | Knowledge Graph Federation Architecture Report | | |
| 94-2 | Knowledge Graph Python API | | |
| 95-1 | KG Search API | T5.5.2 | |
| 96-2 | Collaboratory Integrated Search UI | | |
| 97-3 | KG Search Python API | | |
| 98-4 | KG Search Indexer | | |
| 99-1 | Web site | T5.5.3 | |
| 100-2 | UI development support | | |
| 101-1 | Image Service | T5.5.4 | |
| 102-2 | Python Image Service Client | | |
| 103-3 | Data Analytics Service | | |
| 104-4 | Interactive Data Analytics Notebook Service | | |
| 105-1 | Collaboratory | T5.5.5 | |
| 106-1 | HBP Standard Deployment service | T5.5.6 | |
| 111-1 | Spatial Search API | T5.5.7 | |
| 111-2 | Spatial Index for KG | | |
| 109-1 | Connection of ilastik to HBP 2D and 3D viewers | T5.6.1 | WP5.6 |
| 110-1 | Connection of ilastik to other HBP services | T5.6.2 | |
| 112-1 | Segmentation tools | T5.6.3 | |
| 113-1 | Elephant | T5.7.1 | WP5.7 |
| 114-2 | Elephant Code Repository | | |
| 115-3 | Expanded and improved release of Elephant | | |
| 116-1 | Neural activity resource | T5.7.2 | |
| 117-2 | Linking NAR to human and mouse brain atlas | | |
| 118-1 | Knowledge Space – Service | T5.8.3 | WP5.8 |
| 119-2 | Knowledge Space – Software | | |
| 120-3 | Cell type ontologies | | |
| 121-4 | Integrated data sources | | |

# 4. Use Case Breakdown

| DPIT ID | SP | MultiSP / CDP | Name |
|---|---|---|---|
| **1** | 1 | | SP1 - 3D reconstruction and visualization of microanatomical parameters of pyramidal cells: PyramidalExplorer 1.2 |
| **2** | 1 | | SP1 - Biophysical model of LTP and LTD in wild type and mutant hippocampal CA1 synapses |
| **3** | 1 | | SP1 - Connectivity skeletons to explore local synaptic rules |
| **4** | 1 | | SP1 - GABAergic interneurons characterization |
| **5** | 1 | | SP1 - High-resolution synaptic maps in the mouse neocortex and hippocampus |
| **6** | 1 | | SP1 - Spatial models along networks allowing to place synapses locations over the dendritic shaft |
| **7** | 2 | | SP2 - Cell types quantitative characterization in the human brain |
| **8** | 3 | | SP3-UC-001 Neuronal ensemble activity in brain areas during learning, navigation and multisensory integration |
| **9** | 3 | | SP3-UC-002 The Comprehensive Cluster Cutter |
| **10** | 3 | | SP3-UC-003a Robotic demonstration of visuotactile integration and spatial memory |
| **11** | 3 | | SP3-UC-003b Systems-level neural model of tactile pattern detection and episodic memory |
| **12** | 3 | | SP3-UC-003c (connected to 003a,b) Simulation of spatial navigation and multisensory object recognition *in silico* and robotics using behavioural neurophysiology data |
| **13** | 3 | | SP3-UC-005a A perturbation-based atlas |
| **14** | 3 | | SP3-UC-005b (SP3-UC-WP3.2-b) Repository of slow-wave models, simulators and observables |
| **15** | 3 | | SP3-UC-005c Spontaneous and light-evoked slow wave activity recordings and simulations |
| **16** | 3 | | SP3-UC-006 Spontaneous cortical activity in rodents and humans in wake, sleep, and general anaesthesia |
| **17** | 3 | | SP3-UC-007 Perturbational Complexity Index and related measurements in disorders of consciousness |
| **18** | 3 | | SP3-UC-008a Perceptual learning and plasticity (a) |

| | | | |
|---|---|---|---|
| **19** | 3 | | SP3-UC-008b Perceptual learning and plasticity (b) |
| **20** | 3 | | SP3-UC-009a Integrating fluorescence data of neural activity throughout the cortical column |
| **21** | 3 | | SP3-UC-009b Population-level GCaMP Calcium imaging data in sensory processing, visual perception and alterations of conscious state. |
| **22** | 3 | | SP3-UC-010 (connected to SP3-UC-006) Computation of global measures of neural connectivity, perceptual integration and consciousness |
| **23** | 3 | | SP3-UC-011 (to be worked out): High-resolution functional MRI database of visual cortex, consisting of area and layer-specific resolution for comparison between species and computational models. |
| **24** | 3 | | SP3-UC-012 (to be worked out): Layered deep cortical architectures for predictive coding based on information theoretic principles of pyramidal cells |
| **25** | 4 | | SP4 - CDP4 - Integrative Loop – comparative analysis of experimental and simulated data |
| **26** | 4 | | SP4 - Collaborative analysis of experimental data (WP4.5) |
| **27** | 4 | | SP4 - Generic models and algorithms |
| **28** | 4 | | SP4 - Generic Neuronal Network Simulation |
| **29** | 5 | | SP5 - Annotating a neuron with the tag of the region it belongs to |
| **30** | 5 | | SP5 - Dendritic spines morphological classification |
| **31** | 5 | | SP5 - Morpho-electrical neuron types using unsupervised classification and multi-output regression |
| **32** | 5 | | SP5 - Repairing models of incomplete dendrites |
| **33** | 5 | | SP5- Basal arborisation single-cell hierarchical models |
| **34** | 6 | | SP6 - *In silico* microcircuit experimentation |
| **35** | 6 | | SP6-001 Molecular |
| **36** | 6 | | SP6-002 Subcellular (Jeanette/Dan) |
| **37** | 6 | | SP6-003 Human Cells |
| **38** | 6 | | SP6-004 Somatosensory Cortex |
| **39** | 6 | | SP6-005 Cerebellum |
| **40** | 6 | | SP6-006 Hippocampus |
| **41** | 6 | | SP6-007 Basal Ganglia |
| **42** | 6 | | SP6-00x Simplification |

| | | | |
|---|---|---|---|
| **43** | 6 | | SP6-00x Whole mouse brain |
| **44** | 8 | | SP8 - Baseline healthy control data |
| **45** | 8 | | SP8 - Candidate brain regions or genetics pathways from animal studies |
| **46** | 8 | | SP8 - Generative models for brain data |
| **47** | 8 | | SP8 - High Performance Deep Phenotyping for large scale machine learning |
| **48** | 8 | | SP8 - Microservices to enhance collaboration with HBP platforms |
| **49** | 9 | | SP9-UC-001 - A single run of a simple network model |
| **50** | 9 | | SP9-UC-002 A scripted run of a complex network model with input data and parameter files |
| **51** | 9 | | SP9-UC-003 Using the Neuromorphic Computing Platform through the Collaboratory and Brain Simulation Platform |
| **52** | 9 | | SP9-UC-004 Parameter sweeps |
| **53** | 9 | | SP9-UC-005 Closed-loop experiment involving a virtual environment |
| **54** | 9 | | SP9-UC-006 Validation of neuromorphic results with respect to software simulations |
| **55** | 9 | | SP9-UC-006b Validation of neuromorphic results with respect to software simulations using central storage and HPC |
| **56** | 9 | | SP9-UC-007 Reproducible research |
| **57** | 9 | | SP9-UC-008 Long-term simulation with STDP |
| **58** | 9 | | SP9-UC-009 Evaluate platform performance |
| **59** | 9 | | SP9-UC-010 NeuroML import |
| **60** | 10 | | SP10 - Manipulation experiments with humanoid robots and human avatars |
| **61** | 10 | | SP10 - Mouse rehabilitation experiment in the Neurorobotics Platform |
| **62** | | CDP1 | CDP1-01 - Spatio-temporal coordinated activity during motor learning |
| **63** | | CDP1 | CDP1-02 - Robotic platform for the study of rehabilitation-induced cortical remapping after stroke |
| **64** | | CDP1 | CDP1/SP8 - Predicting motor performance in humans after stroke |
| **65** | | CDP2 | CDP2-UC-001 - Single cell modelling |
| **66** | | CDP2 | CDP2-UC-002 - Multi-scale validation |
| **67** | | CDP2 | CDP2-UC-003 - *In silico* microcircuit experimentation |
| **68** | | CDP3 | CDP3-P1 - 2D interactive atlas viewer with annotation capabilities |

| | | | |
|---|---|---|---|
| **69** | | CDP3 | CDP3-P2 - 3D interactive big data viewer |
| **70** | | CDP3 | CDP3-P3 - Provision and maintenance of template datasets with labelled parcellations in the Human Brain Atlas |
| **71** | | CDP3 | CDP3-P4 - Enrichment of the Human Brain Atlas with qualitative and quantitative datasets |
| **72** | | CDP3 | CDP3-P5 - Interactive spatial alignment tools for human brain data |
| **73** | | CDP3 | CDP3-P6 - Interactive region-based analysis of different modalities |
| **74** | | CDP3 | CDP3-P7 - Revision of human brain ontologies to reflect intersubject variability |
| **75** | | CDP3 | CDP3-P8 - Modelling and model validation using human quantitative data |
| **76** | | SPs 1 & 5 | SP1-SP5 - Prediction of axonal termination patterns of long-range projection neurons (LRPN) |
| **77** | | SPs 1 & 5 | SP1-SP5 - Use of publicly available gene expression data to predict cell density, cellular composition and connectivity of the mouse brain. |
| **78** | | SPs 1 & 6 | SP1/SP6 - fitting individual synaptic events |
| **79** | | SPs 2 & 3 | SP2-SP3 - Encoding models of complex stimuli into local brain activity |

# 5. Budget Breakdown

Table 11: SP5 budget breakdown for revised D5.6.1.

NOTE: The budget breakdown information is confidential and for use within the Project only.

# 6. Proposal for an Ontology Definition Team (ODT)

*This section is kept unchanged from D5.6.1 of 10 October 2016.*

DPIT proposes creating an Ontology Definition Team (ODT), led by the SP5 leader. The ODT should become an overarching activity of the HBP. The ODT should further consist of 3 technical WP leaders from SP5. The ODT should include high ranked representatives from all SPs, including SP12.

The ODT's work is assisted by the WP leader of WP5.1, Task 5.1.1 in the "Data Curation Support Lab".

The ODT regularly scans the SPs in advance of the meetings for proposals of definitions and adaptions of data standards, standardized coordinate systems of the HBP atlas templates, data categories as well as new standards. T5.1.2 is responsible for the knowledge engineering of the ontologies required to support the data driven model building. This concerns all data and model activities of the HBP, while T5.1.1 is training the users to adopt the formats and ontologies proposed.

In addition, the ODT monitors the international standardization activities and actively contributes to such work.

The leader of the ODT works in close contact to the HBP data governance group that is concerned with the ethical and legal aspects.

The leader of the ODT regularly reports to the SIB.

# 7. Data Flows and Platform Integration

*This section is kept unchanged from D5.6.1 of 10 October 2016.*

The HBP is a prolific producer of data and models in all contexts of brain research. The data and the models are generated in multiple SPs and under several CDP umbrellas. Coupled with the improving toolsets in various SPs for building and simulating every more realistic models, it is not surprising that the complexity of these activities increases exponentially. It is also evident that no single person or a disjoint set of groups can ever create a complete knowledge of all data flows in the project, rather creation of this knowledge needs a genuinely governed collaborative effort requiring sound monitoring as well as management of the HBP data flows.

It is important to document clearly:

- who creates models or data;
- who uses models and data created in the HBP;
- which models are brought in from where;
- how the impact of a failure of delivery can be mitigated.

Paraphrasing the Consortium Agreement (AC) of the HBP, there it is primarily considered the responsibility of the data producer to keep data accessible. HBP partners may share this responsibility if they provide data repository services for use by data producers. When data producers use 3rd parties outside of the HBP to store their data produced under HBP funding they are responsible for ensuring that the 3rd party complies with appropriate data and system engineering practices. In practice, however, SP5's previous attempts to create "partner federated storage" have failed to deliver a viable solution for data producers to store their data in a reliable data repository. As a consequence, the majority of partners cannot currently secure data archival except through highly heterogeneous services offered by their local institutions. Under these conditions data archival is addressed but there is no standard mechanism for HBP or 3rd parties to access data stored with a partner. As a result, data accessibility remains a critical unsolved problem.

It is therefore equally important

a. to know which domains (clustered in the SPs) are connecting with each other to get consensus on standard data formats, term usage and definitions for ontology work (see section 6 on the ODT;
b. to guarantee the effectiveness of the flow of data between "Producers" and "Modellers" in the HBP;
c. to achieve the integration of the NIP services with the HBP federated data infrastructure strategy.

**Ad a. and b.** DPIT has carried out a Data Flow Survey. The results of this survey will provide the necessary input to both the ODT and T5.1.2 for ontology planning as well as give important information as to capacity planning for the primary archive platform of the HBP and the large-scale data repositories (see FENIX) as well as for the HBP high speed interconnect on top of the PRACE data infrastructure. The results of the survey are given in the document "HBP Data Flows" (A1, XII).

**Ad c.** DPIT recognizes the need for transcending the previous HBP ad-hoc concept of "partner federated storage" that led to lots of small data stored in the HBP Collab or being hidden at partner sites while large data is mostly stored un-federated at supercompute centres.

DPIT/SP5 together with SP7/FENIX agreed to introduce a new strategy for the integration of the NIP services with the HBP federated data infrastructure strategy. The latter is pursued by SP7. Through SP7, a federation of HBP's supercomputing partner centres is right on its way, parallel to the activities of DPIT. The planned large-scale data service platform called FENIX (Federated ENgine for Information Exchange) plays the role of the

underlying basis infrastructure, i.e. as IaaS (Infrastructure as a Service) for the NIP, which provides both aspects, SaaS and PaaS (Software as a Service and Platform as Service), both through the HBP Collaboratory. The IaaS aspect pursued by FENIX concerns both the issue of 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. The details will be described in the parallel Deliverables D11.3.1 (HBP Collaboratory Architecture) and D5.6.2 (HBP Overall Platforms Architecture).

# Appendix

## 1. Mapping Tasks D5.6.1 Oct 10, 2016 → present

Tables 12 and 13 in D5.6.1 of 10 October 2016 show the reordering of WPs and Tasks from earlier work plan versions up to the version presented in the DPIT work plan.

The mapping of Tasks from the DPIT work plan onto the present revised work plan is shown below.

Table 12: Re-ordering of tasks from DPIT work plan (D5.6.1 of October 10, 2016) to the present Second revised work plan.

| WP | Task | Task Title | Changes to Second Revised Workplan |
|---|---|---|---|
| 5.1 | 5.1.1 | Data and Parameter Workbench Support and Curation Team | |
| | 5.1.2 | Ontologies for the integration of Models in the KnowledgeGraph | |
| | 5.1.3 | Integration of SP6 models and supporting datasets | Suppressed |
| | 5.1.4 | Large-Scale Federated Data Accessibility Services | 5.1.3 |
| 5.2 | 5.2.1 | Maintenance of Rodent Atlases | |
| | 5.2.2 | Curation of Rodent Atlas Data | |
| | 5.2.3 | Harvesting and Curation of Strategic Metadata from Existing Third Party Data Repositories | |
| | 5.2.4 | Strategic Mining of Data Anchored to Rodent Atlases | |
| | 5.2.5 | Prediction of Cellular, Synaptic and Connectomic Composition, Distributions and Properties of the Rodent Brain | |
| 5.3 | 5.3.1 | Curation of Human Atlas Data | |
| | 5.3.2 | Alignment and Import of Strategic Data from External Repositories | |
| | 5.3.3 | Cross-scale Interactive Spatial Alignment Tools for Partial Volumes | |

| | | | |
|---|---|---|---|
| | 5.3.3 | Cross-scale Interactive Spatial Alignment Tools for Partial Volumes | |
| | 5.3.4 | Machine Learning and Statistical Methods for Modeling Cellular and Subcellular Morphologies | |
| 5.4 | 5.4.1 | Tools for Metadata Curation | |
| | 5.4.2 | Integrating 2D Atlas Viewers and Manual Spatial Registration Tools | |
| | 5.4.3 | Development of 3D High-Volumetric Interactive Atlas Viewer | |
| | 5.4.4 | Integration of Neuroimaging Tools | |
| 5.5 | 5.5.1 | Development and Maintenance of KnowledgeGraph | |
| | 5.5.2 | Search Application | |
| | 5.5.3 | Development and Maintenance of Web Front End to the Platform Services | |
| | 5.5.4 | Data and Image Services | |
| | 5.5.5 | Collaboratory | |
| | 5.5.6 | Platform Administration, Operation and Validation | |
| | 5.5.7 | HPAC Platform Service Integration Support | Suppressed |
| 5.6 | 5.6.1 | Interactive Feature Classification and Extraction for Spatial Images | |
| | 5.6.2 | Workflow for Populating Brain Atlases with Features, Automatically Extracted by ilastik | |
| | 5.6.3 | Spatial Search Application | 5.5.7 |
| | 5.6.4 | Integrating Feature Extractors and Classifiers for Neuroscience | 5.6.3 |
| 5.7 | 5.7.1 | Elephant | |
| | 5.7.2 | Neural Activity Resource | |
| 5.8 | 5.8.1 | Subproject Management and Technical Coordination | |
| | 5.8.2 | Neuroinformatics Community Building | |
| | 5.8.3 | Development and Maintenance of KnowledgeSpace | |