



Project Number:	284941	Project Title:	Human Brain Project
Document Title:	HBP Data Management Plan		
Document Filename ⁽¹⁾ :	HBP_SP13_EPFL_14-0205_D13.3.2_Final.docx		
Deliverable Number:	D13.3.2		
Deliverable Type:	Planning Document		
Work Package(s):	WP 13.3		
Dissemination Level:	PP		
Planned Delivery Date:	M6 / 31 March 2014		
Actual Delivery Date:	M8 / 02 May 2014		
Author:	Nenad BUNCIC, EPFL (P1)	<i>Author</i>	
Author:	Felix SCHÜRMANN, EPFL (P1)	<i>Author</i>	
Editor:	Richard WALKER, EPFL P01	<i>Reviewer & Editor</i>	
Abstract:	This document describes the management life cycle for the data that will be collected, deposited, shared, processed, generated or derived by the Human Brain Project (HBP).		
Keywords:	data, data management, data management life cycle, data storage, data preservation, metadata		



Table of Contents

1. Executive Summary	3
2. Introduction	4
2.1 Background	4
2.2 Purpose	4
2.3 Scope	4
3. General Principles	5
3.1 Data Description	5
3.1.1 Data set Identification Card	5
3.1.2 Formats and Metadata	5
3.1.3 Quality Assurance	5
3.2 Access and Sharing	6
3.2.1 Access Procedures	6
3.2.2 Technical Access Methods	6
3.2.3 Access Charges	6
3.2.4 Restrictions on Access	6
3.3 Data Storage and Preservation	7
3.3.1 Storage, Backup, Replication, and Versioning	7
3.3.2 Provenance	7
3.3.3 Archiving and Preservation	7
3.4 Security	7
3.5 Intellectual Property Rights	7
3.5.1 Software Licensing Model	7
3.5.2 Data and Software Release Process	8
3.6 Ethics and Privacy	8
3.7 Adherence	9
Annex A: Data Classes in the HBP	10
Strategic Mouse Brain Data for the HBP Mouse Brain Atlas	10
Strategic Human Brain Data for the HBP Human Brain Atlas	10
Theoretical Neuroscience Data	11
Algorithms, Cognitive Models and Computing Principles for Implementation in HPC, Neuromorphic and Neurobotic Systems	11
Atlases and Brainpedia	12
Brain Simulation Data	12
High Performance Computing Data	13
Medical Informatics Data	13
Neuromorphic Data	14
Neurobotics Data	14
Applications Data	15
Ethics and Society Data	15
Software and Technical Documentation	16
Administrative and Management Documentation	16
Annex B: Data Set Identification Card	17
Annex C: Initial Tasks and Responsibilities	18
Annex D: References	19



1. Executive Summary

This document describes the management life cycle for the data that will be collected, deposited, shared, processed, generated or derived by the Human Brain Project (HBP). The HBP, a ten-year research project funded by the European Commission (EC), is laying the foundations for a new approach to brain research. The fields of neuroscience, medicine and information technology each have important roles to play in addressing this challenge, but the knowledge and data that each is generating are currently very fragmented. The HBP is driving integration of these different contributions and catalysing a community effort to achieve a new understanding of the brain, new treatments for brain disease, and new brain-like computing technologies.

To support this effort, the HBP is creating an integrated system of Information and Communication Technology (ICT) Platforms, offering services to neuroscientists, clinical researchers, information technology developers and roboticists. The total amount of data (experimental data, model data, simulation artefacts, analysis data, software, documentation) integrated and generated by the HBP will be in the petabyte to exabyte scale range.

All data and software will be accessible via the HBP Unified Portal to all registered and active HBP users. The data will be provided in the original file format. To describe the data, metadata will be either embedded or generated in XML format and stored alongside the data. The data and software accessible via the HBP Unified Portal will be used in the format and context in which it is provided.

We propose that the data and the HBP Software be licensed under a "dual licensing" model. Under this model, users may choose to use the data and/or the HBP Software under the free software/open source or under the commercial license terms. We believe a dual licensing model can sustain both innovation and growth, and that by adopting this licensing model, we can encourage freedom of development and contribution.



2. Introduction

2.1 Background

The goal of the Human Brain Project (HBP) is to build a completely new information and communication technology (ICT) infrastructure for neuroscience, and for brain-related research in medicine and computing. This ICT infrastructure will catalyse a global collaborative effort to understand the human brain and its diseases, and ultimately to emulate its computational capabilities.

To achieve the goal, The HBP is developing six ICT platforms dedicated to Neuroinformatics, Brain Simulation, High-Performance Computing (HPC), Medical Informatics, Neuromorphic Computing and Neurorobotics. The HBP Platforms will make it possible to federate neuroscience data from all over the world, to integrate the data into unified models and simulations of the brain, to validate the results against empirical data from biology and medicine, and to make the data available to the global scientific community.

Gathering all existing knowledge about the human brain, and reconstructing it in multi-scale models and supercomputer-based simulations of those models, will offer a fundamentally new understanding of the human brain and its diseases. It will also spur the development of novel, brain-like computing technologies.

The HBP is committed to having its data and software accessible via the HBP Unified Portal to all registered and active HBP users.

2.2 Purpose

The purpose of this plan is to describe the data management life cycle for the data that will be collected, deposited, shared, processed, generated or derived by the HBP.

2.3 Scope

The document outlines the data management policy that the HBP Platforms will use to organise and maintain all data generated by the project. This document will evolve over the lifetime of the HBP as the project expands, and as the strategy for HBP data management develops.



3. General Principles

3.1 Data Description

The HBP will generate and integrate large amounts of experimental data, model data, simulation artefacts, analysis data, software, and documentation – see [Annex A](#) for a list of HBP data classes. The total amount of data integrated and generated by the HBP will be in the petabyte to exabyte scale range.

3.1.1 Data set Identification Card

HBP data providers are required to produce a Data Set Identification Card (DIC) – see [Annex B](#) – for each data set that details the appropriate storage, backup, replication, versioning, archiving and preservation policies. DICs cover the points below on a data set-by-data set basis, and reflect the status of data production in the project as a whole. **Data set reference and name:** Identifier for the data set.

- **Data set description:** Description of the data generated or collected, its origin (in case it is collected), and its nature and scale.
- **Standards and metadata:** Reference to existing standards. If these do not exist, there should be an outline on how and what metadata will be created.
- **Data sharing:** Description of how data will be shared, including access procedures and embargo periods (if any).
- **Archiving and preservation:** Archiving and preservation (including storage and backup) policies.

3.1.2 Formats and Metadata

Data will be provided in the original file format. To describe the data, metadata will be either embedded or generated in XML format and stored alongside the data. For XML-stored metadata, the format must be compliant with the XML standards. XML compliance checking will ensure that the metadata use appropriate XML schemas.

The data providers will be responsible for ensuring that the metadata content is correct and maintained appropriately. Also, the data providers will implement systematic procedures for the periodic checking of metadata.

From the start of the project, the Dublin Core Metadata Standard¹ will be applied during metadata creation. Other metadata standards will be defined and applied accordingly.

3.1.3 Quality Assurance

The HBP will use internationally accepted ontologies and will develop and maintain collaborative workflows for data curation. A process for developing and curating neuroscience ontologies, defined by the Neuroscience Information Framework (NIF)², will drive the development, curation and dissemination of the ontologies needed to annotate and share neuroscience data and models. The ontologies will be distributed via the HBP Unified Portal.



3.2 Access and Sharing

3.2.1 Access Procedures

The HBP Platforms will provide access to all the data, software, workflows, and documents for all registered and active HBP users. The HBP users will access the data via a single point of access (the HBP Unified Portal) shared among all HBP Platforms. A review committee will vet the HBP user account applications to ensure that rare and valuable resources (i.e. supercomputing resources, high performance storage resources, etc.) are used in the most efficient and optimised way. SP6 will develop the HBP Unified Portal, and SP13 will provide the shared infrastructure that it will run on.

3.2.2 Technical Access Methods

SP5 (T5.1.1 - Shared Data Space) will provide infrastructure, software, and interfaces that will allow users to share and access data distributed to sites across the world. The Shared Data Space will allow neuroscientists to upload, access and analyse experimental data and computational models. The data will be discoverable and identifiable by means of a unique Digital Object Identifier (DOI).

A software ecosystem will be developed by SP6, with workflows and algorithms for model reconstructions (T6.1.1). Functionality will be implemented to query the HBP atlases during model reconstruction, and to import the data into the formats required for specific reconstruction tasks (T6.1.2).

SP7 (WP7.4) will design and develop technology to manage very large volumes of experimental and simulation data for analysis, model building and model validation. Novel spatial indexes, implemented by SP7, will allow scalable querying of massive neuroscience data sets located in external data centres, cloud systems and supercomputer storage (T7.4.1).

The SP8 (WP8.1) will build the tools necessary to federate clinical databases around the world, to effectively anonymise patient data, to make the data available to clinical researchers, and to mine the data for clinically relevant information. A software framework will be developed to allow researchers to query clinical data stored on hospital and laboratory servers, without moving the data from the servers where it resides (*in situ* querying) and without compromising patient privacy (T8.1.1-3, T8.1.4).

3.2.3 Access Charges

For a non-commercial and academic use, there will be no charge for the use of the HBP Unified Portal. For a commercial use, a commercial license will be required.

3.2.4 Restrictions on Access

To access the HBP Unified Portal all users will be required to register and to have a valid HBP account. The registration process will be open to all partners of the HBP. However, a review committee will vet the HBP user account applications to ensure that rare and valuable resources (i.e. supercomputing resources, high performance storage resources, etc.) are used in the most efficient and optimised way.

Access policies to the HBP Unified Portal for users external to the HBP will be defined in subsequent versions of this data management plan in time for the general release of the HBP Platforms (Month 30).



3.3 Data Storage and Preservation

3.3.1 *Storage, Backup, Replication, and Versioning*

Data providers, such as institutes and laboratories, and resource providers, such as HPC data centres, will be responsible for providing storage and backup resources for the raw data sets. SP13 (T13.3.1) will manage metadata storage, backup and replication. The metadata catalogue will be replicated across several zones to facilitate queries and to provide redundancy and resilience.

The data and resource providers are responsible for creating and implementing maintenance plans and policies, in accordance with the relevant industry standards and best practices, to manage the data and metadata backup and recovery process.

Specific storage, backup, replication and versioning policies will be described in more detail in the DICs.

3.3.2 *Provenance*

The HBP is committed to describing the generated and derived data in sufficient detail to facilitate reproduction and validation of results. Whenever possible, the chain of ownership will be recorded, as will the transformations the data have undergone.

3.3.3 *Archiving and Preservation*

The HBP Platforms will ensure that data are archived and migrated to new formats, platforms, and storage media in accordance with the good practices of the digital preservation community. The resource providers will address succession planning for digital assets and will designate a successor in the unlikely event that such a need arises.

Specific archiving and preservation policies will be described in more details in the DICs.

3.4 Security

Access to the data and software will be managed via the HBP Unified Portal. All users will have a project-wide HBP account. The Access Control Lists (ACL) will specify which users or system processes are granted access to objects, as well as what operations are allowed on given objects.

The HBP data providers are required to provide the DIC for each data set, detailing sharing and security policies. The DIC will be added to this Data Management Plan.

3.5 Intellectual Property Rights

Data owners and their institutions hold the copyright for the data they generate in accordance to the HBP Consortium Agreement. Access to these data for implementation and use by other HBP partners is similarly regulated in the HBP Consortium Agreement. Usage of the data by third parties may be subject to additional agreements.

3.5.1 *Software Licensing Model*

We propose that the HBP Software (software entirely generated within the Project) be licensed under a "dual licensing" model. Under this model, users may choose to use the HBP Software under the free software/open source or under the commercial license terms.



Under open source license, the HBP Software will be available free of charge, as long as the use adheres to the terms of the license. Commercial use of the HBP Software will be possible under the commercial license terms.

We believe a dual licensing model can sustain both innovation and growth, and that by adopting this licensing model, the HBP Software can remain open-source and encourage freedom of development and contribution.

Where appropriate, preference will be given to the use of HBP software through the HBP Unified Portal.

Different licensing models might be applicable for software generated outside the project, software existing prior to the project, or software generated partially within the project.

3.5.2 Data and Software Release Process

All data and software will be accessible via the HBP Unified Portal to all registered and active HBP users. The data and software accessible via the HBP Unified Portal will be used in the format and in the context in which it is provided. The data and software will follow five lifecycle phases:

- 1) **Creation:** The initial phase and the data or software origin point.
- 2) **Platform Distribution:** The phase and the process of managing the data or software once it has been created or received.
- 3) **Internal Use:** This takes place after data or software is distributed internally.
- 4) **IP Protection and Licensing:** This phase refers to the recognition of exclusive rights to data or software.
- 5) **Release:** The last phase when data or software will be released outside of the HBP, in accordance with the licensing model.

Exceptional Requests (requests prior to Release) for raw data and software binary or source code can be made, and will be considered under the following conditions:

- 1) Requested data or code does not include confidential data.
- 2) Purposes of the study/analysis/investigation are approved by the HBP.
- 3) Results and format of the study/analysis/investigation are presented to the HBP for further discussion, modification and acceptance prior to public release.
- 4) An agreement exists or is made between the HBP and the entity requesting data in terms of partnership and resource sharing to conduct data analyses.

3.6 Ethics and Privacy

Privacy issues arise when data (e.g. patient records, hospital information records, biological traits and genetic material) are collected and stored. The handling of digital personal data is a major concern because of the processing possibilities and the potential to link vast amounts of personal data. Privacy concerns exist with any data that—either alone or when linked—relate to an identifiable individual or individuals.

The HBP data providers will be responsible for rendering the data anonymous (when obvious individual identifiers are removed) and non-identifiable (when an individual cannot be identified either immediately from the data or when the data are combined with other data).



Where required, human subjects contributing data to the project (e.g. in SP2 and SP3) will be asked to explicitly consent to the sharing of their anonymised data within the HBP and within the research community.

3.7 Adherence

Responsibilities for implementation of individual Tasks are set out in [Annex C](#). Adherence to this data management plan will be controlled by SP13 - specifically T13.3.1 (Central Support - IT Services).



Annex A: Data Classes in the HBP

Strategic Mouse Brain Data for the HBP Mouse Brain Atlas

Roles and Responsibilities	SP1 (Strategic Mouse Brain Data) is responsible for generating strategic mouse brain data, and complementing or completing existing data on the structure of the mouse brain. T1.3.1 (University of Edinburgh) is responsible for the data aggregation and depositing into the HBP Mouse Brain Atlas. SP5 (Neuroinformatics Platform) is responsible for the data integration, federation, management and administration.
Data Types	Mouse brain transcriptomics and proteomics data; and mouse brain macrostructure, vasculature, cells and synapses data.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP5 (Neuroinformatics Platform) is responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.

Strategic Human Brain Data for the HBP Human Brain Atlas

Roles and Responsibilities	SP2 (Strategic Human Brain Data) is responsible for generating strategic human multi-level data that parallels the data collected for mouse to facilitate the use of mouse data to predict human data. T2.2.2 (Jülich) is responsible for the data aggregation and depositing into the HBP Human Brain Atlas. SP5 (Neuroinformatics Platform) T5.6.1 is responsible for the data integration, federation, management and administration.
Data Types	Structural and functional human brain data, neuron morphologies, numbers and distributions of neurons and glia, distribution of receptors in the human cerebral cortex, and connection maps.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP5 (Neuroinformatics Platform) is responsible for making the data available and accessible in the Shared Data Space (T5.1.1). SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.



Theoretical Neuroscience Data

Roles and Responsibilities	SP4 (Theoretical Neuroscience) is responsible for producing simplified models of complex brain structures and dynamics; rules linking learning and memory to synaptic plasticity; large-scale models creating a bridge between “high-level” behavioural and imaging data; and mathematical descriptions of neural computation at different levels of brain organisation. SP6 (Brain Simulation Platform) and SP13 (Management) T13.3.1 will share responsibility for the data integration, federation, management and administration.
Data Types	Simplified models and large-scale models.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP5 (Neuroinformatics Platform) is responsible for making the data available and accessible in the Shared Data Space (T5.1.1). SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.

Algorithms, Cognitive Models and Computing Principles for Implementation in HPC, Neuromorphic and Neurobotic Systems

Roles and Responsibilities	SP3 (Cognitive Architectures) is responsible for providing functional mapping data, cognitive architectures and models for the HBP Human Brain Atlas. SP4 (Theoretical Neuroscience) is responsible for providing algorithms, cognitive models and computing principles for the HBP Human Brain Atlas. SP5 (Neuroinformatics Platform) T5.6.1 is responsible for the data integration, federation, management and administration. The algorithms and computing principles will be used in HPC, neuromorphic and neurobotic systems.
Data Types	Functional mapping data, cognitive architectures and models, and algorithms and computing principles.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP5 (Neuroinformatics Platform) is responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.



Atlases and Brainpedia

Roles and Responsibilities	SP5 (Neuroinformatics Platform) is responsible for building the atlases, data integration, federation, management and administration.
Data Types	Mouse Brain Atlas, Human Brain Atlas, and Brainpedia.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP5 (Neuroinformatics Platform) T5.6.1 is responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.

Brain Simulation Data

Roles and Responsibilities	SP6 (Brain Simulation Platform) is responsible for delivering an Internet-accessible collaborative platform for data-driven predictive reconstruction and simulation of brain models.
Data Types	Brain models, simulation results, and validation tests.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	HPC Data Centres (Jülich, CINECA, CSCS) will provide the storage infrastructure. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the respective HPC Data Centres (Jülich, CINECA, CSCS).



High Performance Computing Data

Roles and Responsibilities	SP7 (High Performance Computing Platform) is responsible for providing supercomputing, Big Data and Cloud capabilities at the exascale level, as well as the system software, middleware, interactive computational steering and visualisation support necessary to create and simulate multi-scale brain models and to address the hard-scaling challenges of whole brain modelling.
Data Types	Simulation artefacts.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	HPC Data Centres (Jülich, CINECA, CSCS) will provide the storage infrastructure. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the respective HPC Data Centres (Jülich, CINECA, CSCS).

Medical Informatics Data

Roles and Responsibilities	SP8 (Medical Informatics Platform) is responsible for building the tools to federate clinical data, recruiting hospitals to use the system, and developing tools to extract unique biological signatures of disease making it possible to develop a new, comprehensive classification of brain diseases.
Data Types	Brain scans; electrophysiology data; electro-encephalography and genotyping data; metabolic, biochemical and haematological profiles; and clinical instruments data.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP8 (Medical Informatics Platform) T8.4.2 is responsible for building a federated data space for clinical data.
Data Storage, Backup, Archiving and Preservation	The data will be federated and anonymised (D8.6.1). <i>In-situ</i> distributed database querying (T8.1.1) will be developed to facilitate search and discovery of the data.



Neuromorphic Data

Roles and Responsibilities	SP9 (Neuromorphic Computing Platform) is responsible for allowing non-expert neuroscientists and engineers to perform experiments with configurable Neuromorphic Computing Systems (NCS) implementing simplified versions of brain models.
Data Types	Simplified brain models, neuromorphic circuits, configuration files, experimental environments, experiments and experimental results, and simulation results.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP9 (Neuromorphic Computing Platform) is responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.

Neurorobotics Data

Roles and Responsibilities	SP10 (Neurorobotics Platform) is responsible for allowing researchers to design and run simple experiments in cognitive neuroscience using simulated robots and simulated environments linked to simplified versions of HBP brain models.
Data Types	Simplified brain models, configuration files, experimental environments, experiments and experimental results, and simulation results.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	SP10 (Neurorobotics Platform) is responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.



Applications Data

Roles and Responsibilities	SP11 (Applications) is responsible for testing and refining the pre-release versions of the ICT platforms, providing early, small-scale demonstrations of their potential for research in neuroscience, medicine and computing and preparing for more ambitious research in the operational phase.□□
Data Types	Configuration files, experimental environments, and experiments and experimental results.
Standards and Metadata	The data will be provided in the original file format. Metadata will be either embedded or generated in XML format and stored alongside the data.
Access, Sharing and Privacy	The respective platform on which the data are run will be responsible for making the data available and accessible in the Shared Data Space. SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the Shared Data Space provided, managed, and administered by SP5 (Neuroinformatics Platform) T5.6.1. SP5 will implement, within the scope of the Shared Data Space (T5.1.1), a digital preservation process to ensure integrity and authenticity of the data.

Ethics and Society Data

Roles and Responsibilities	SP12 (Ethics And Society) is responsible for exploring the project’s social, ethical and philosophical implications, promoting engagement with decision-makers and the general public, fostering responsible research and innovation by raising social and ethical awareness among project participants, and ensuring that the project is governed in a way that ensures full compliance with relevant legal and ethical norms.
Data Types	Data showing all ethics applications by partners across the project.
Standards and Metadata	The data will be in the database.
Access, Sharing and Privacy	SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via HBP Unified Portal. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the database. SP13 (T13.3.1) will implement a process to ensure integrity and of the data.



Software and Technical Documentation

Roles and Responsibilities	SP5 (Neuroinformatics Platform), SP6 (Brain Simulation Platform), SP7 (High Performance Computing Platform), SP8 (Medical Informatics Platform), SP9 (Neuromorphic Computing Platform) and SP10 (Neurorobotics Platform) are responsible for delivering Internet-accessible collaborative platforms for reconstruction and simulation of brain models.
Data Types	Software, configuration files and validation tests.
Standards and Metadata	The data will be provided in the original file format
Access, Sharing and Privacy	All Platforms are responsible for making the software available and accessible in the Central Software Repository (Git). SP13 (Management) T13.3.1 will provide the infrastructure for running a Git Server. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving & Preservation	The data will be stored in the Central Software Repository provided, managed, and administered by SP13 (Management) T13.3.1.

Administrative and Management Documentation

Roles and Responsibilities	SP13 (Management) is responsible for supporting HBP decision-making, operating the management structure and European Research Programme, ensuring transparency and accountability toward funders and stakeholders, and maintaining standards of quality and performance.
Data Types	Agenda and minutes of meetings, deliverables, and quarterly reports.
Standards and Metadata	The data will be provided in the original file format.
Access, Sharing and Privacy	SP13 (Management) T13.3.1 will provide the infrastructure for accessing the data via EMDESK Research Project Management Platform. Access and sharing of the data will be granted in accordance with the HBP user account rights.
Data Storage, Backup, Archiving and Preservation	The data will be stored in the EMDESK Research Project Management Platform database provided, managed, and administered by SP13 (Management) T13.3.1.



Annex B: Data Set Identification Card

DATA SET IDENTIFICATION CARD	
Name	Unique name of the data set. The name should be in the following form: SPnnWPnnTnnDSnnnn_human_readable_name
SPID	Subproject ID
TID	Task ID
Description	Description of the data that will be generated or collected, format, its origin, nature and scale.
Standards and metadata	Reference to existing data and metadata standards. If these do not exist, an outline on how and what metadata will be created.
Access, sharing, and privacy	<p>Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).</p> <p>In case the data set cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).</p>
Storage and backup	Description of where the primary data resides as well as the data backup plan including backup frequency and backup retention (e.g. incremental and transaction log backups hourly, so that in the event of a catastrophic failure, researchers will lose no more than 1 hour of work. Retention period of 10 years to allow researchers to be able to restore the data set to any point in time within the last 10 years).
Archiving and preservation	Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, its approximated end volume, and the associated costs and how they will be covered.
Notes	



Annex C: Initial Tasks and Responsibilities

Task	Description	Leader	Institution
T1.3.1	Deposit data in the HBP mouse brain atlas	Douglas ARMSTRONG	University of Edinburgh
T2.2.2	Deposition of Human Brain data in the HBP Brain Atlas and Brainpedia	Katrin AMUNTS	Forschungszentrum Jülich
T5.1.1	Shared data space	Sten GRILLNER	Karolinska Institutet
T5.6.1	Neuroinformatics Platform integration, website construction, maintenance and administration	Sean HILL	EPFL
T6.5.1	Brain Simulation Platform integration, website construction, maintenance and administration	Henry MARKRAM	EPFL
T7.4.3	Data provenance and preservation	Peter BUNEMAN	University of Edinburgh
T7.5.7	High Performance Computing Platform website construction and maintenance	Thomas LIPPERT	Forschungszentrum Jülich
T8.1.2	Data provenance, preservation and integration	Vasilis VASSALOS	Athens University of Economics and Business
T8.4.1	Medical Informatics Platform integration, website construction, maintenance and administration	Richard FRACKOWIAK	Centre Hospitalier Universitaire Vaudois
T8.4.2	Setting up federated infrastructure	Richard FRACKOWIAK	Centre Hospitalier Universitaire Vaudois
T9.5.5	Neuromorphic Computing Platform website construction and maintenance	Karlheinz MEIER	Ruprecht-Karls-Universität Heidelberg
T10.3.1	Setting up and administration of the Neurorobotics Platform	Alois KNOLL	Technische Universität München
T13.3.1	IT Services	Nenad BUNCIC	EPFL



Annex D: References

¹ <http://dublincore.org/documents/dces/>

² <http://www.neuinfo.org>