EBRAINS Data and Knowledge services: FAIR Data and Models for the Neuroscience Community Status at M18 (D4.7 – SGA3)

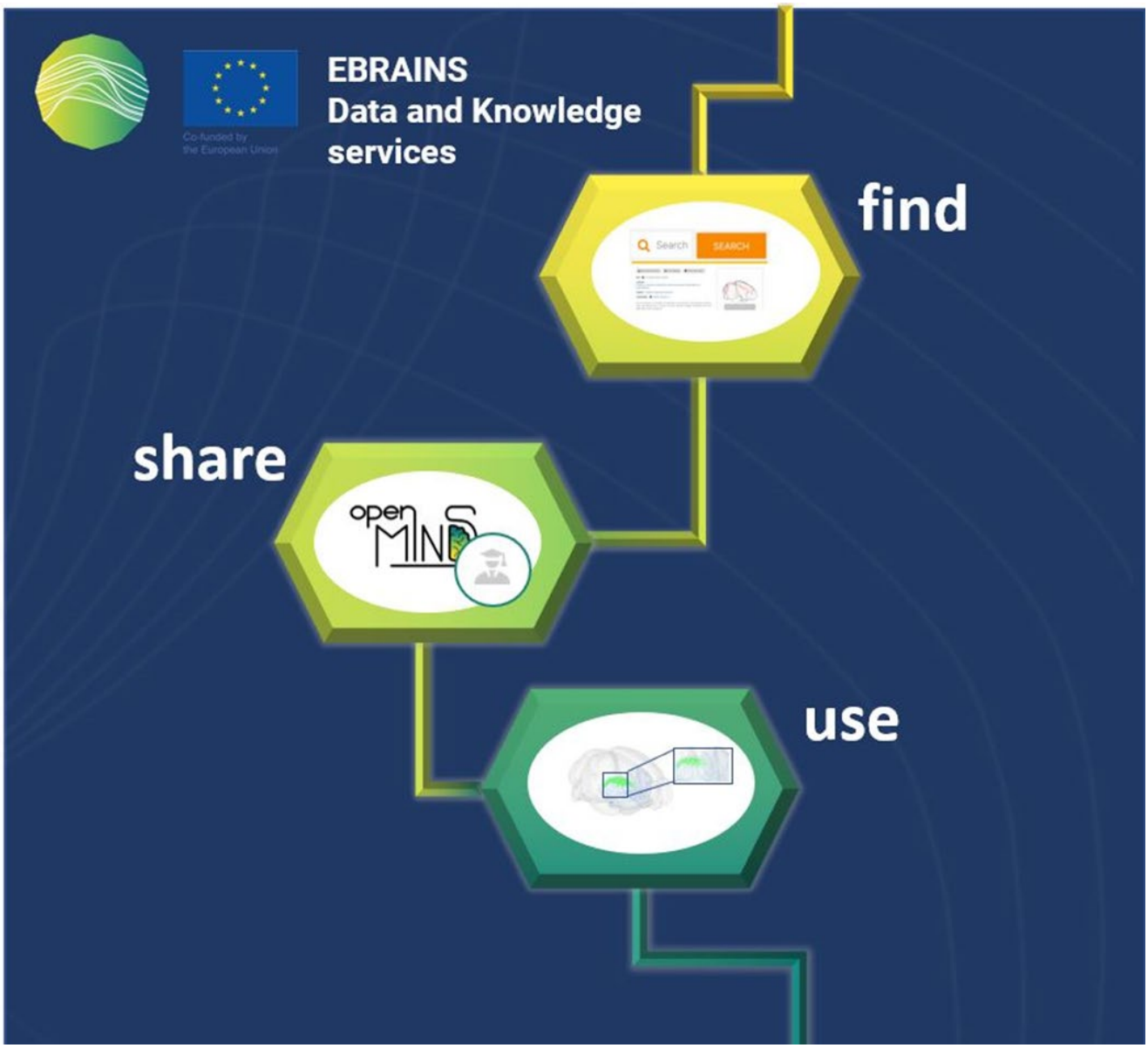

Figure 1: EBRAINS Data and Knowledge Service

The EBRAINS Data and Knowledge services targets the broader neuroscience community. Through the FIND and SHARE services, closely linked to other EBRAINS services, users have access to comprehensive tools and services for finding, sharing, and using shared data and computational models.

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	D4.7 EBRAINS FAIR data services (SC1) - status at M18
Document Filename:	D4.7 (D38) SGA3 M18 ACCEPTED 220520.docx
Deliverable Number:	SGA3 D4.7 (D38)
Deliverable Type:	Other
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA3 M18 / 31 Sep 2021
Actual Delivery Date:	SGA3 M20 / 30 Nov 2021; accepted 20 May 2022
Author(s):	Jan G. BJAALIE, UIO (P81), Andrew DAVISON, CNRS (P10), Ida AASEBØ, UIO (P81), Lyuba ZEHL, JUELICH (P20), Oliver SCHMID, EBRAINS (P1), Mathew ABRAMS, KI (P37), Simisola AKINTOYE, DMU (P16), William KNIGHT, DMU (P16), Damian EKE, DMU (P16)
Compiled by:	Martha Elisabeth BRIGG, Roman VOLCHENKOV, UIO (P81)
Contributor(s):	Shailesh APPUKUTTAN, CNRS (P10), Onur ATES, CNRS (P10), Timo DICKSCHEID, JUELICH (P20), Tom GILLESPIE (International Neuroinformatics Coordinating Facility), Xiao GUI, JUELICH (P20), Camilla HAGEN BLIXHAVN, UIO (P81), Anna HILVERLING, JUELICH (P20), Heidi KLEVEN, UIO (P81), Stefan KÖHNEN, JUELICH (P20), Trygve LEERGAARD, UIO (P81), Elodie LEGOUÉE, CNRS (P10), Glynis MATTHEISEN, CNRS (P10), Maja PUCHADES, UIO (P81), Ingrid REITEN, UIO (P81), Ulrike SCHLEGEL, UIO (P81), Benjamin WEYERS, UT(130), Sara ZAFARNIA, JUELICH (P20), Yann ZERLAUT, CNRS (P10), Martha Elisabeth BRIGG, UIO (P81)
WP QC Review:	Timo DICKSCHEID, JUELICH (P20)
WP Leader / Deputy Leader Sign Off:	Jan G. BJAALIE, UIO (P81)
T7.4 QC Review:	Martin TELEFONT (QC) and Annemieke MICHELS (editorial), EBRAINS (P1)
Description in GA:	Knowledge Graph release with improved functionality and extended content, and updated inventory of related FAIR data, tools and services, prepared for, or released through the EBRAINS portal.
Abstract:	The EBRAINS Data and Knowledge services facilitate neuroscience research and discovery by providing an online conduit for both sharing and easy access to research data, computational models, and software. These services revolve around an expert-driven Knowledge Graph which combines metadata ingestion pipelines, human user input, and multiple quality assurance processes to ensure consistency and quality to aid researchers who wish to contribute. Neuroscientists looking to make their data, models, and software FAIR (Findable, Accessible, Interoperable, and Reusable) can apply for user support for curation and annotation with standardised metadata facilitating discovery and reuse by the broader research community. Moreover, clearly defined Terms of use, responsible data compliance, protection, and governance make the EBRAINS Data and Knowledge services mutually attractive for researchers both depositing and consuming content. Neuroscientists seeking data or models to support their research can find and access data and models either through a straightforward online search or a programmatic API. Moreover, these shared resources are linked to software for visualising and analysing data and for performing computerised simulations of brain function. The ability for scientists to more easily share, find, integrate, analyse and simulate data,

	accelerates the global effort to understand human brain function and disease. This document is a follow up from Deliverable D4.1(D32) produced in October 2020.
Keywords:	Data sharing; data curation; data management; infrastructure; data compliance, ethics, ontology, EBRAINS.
Target Users/Readers:	Computational neuroscientists, HBP/EBRAINS users, Consortium members, funders, general public, policymakers, students

Table of Contents

Table of Tables

Table of Figures

1. Introduction

The EBRAINS Data and Knowledge services target the broader neuroscience community, making it easy for users to share and publish data and models, and to find heterogeneous data, models, and related software relevant for a broad range of research fields in basic and clinical neuroscience.

The services are available through the EBRAINS web portal and include:

- EBRAINS Knowledge Graph
- EBRAINS Curation
- Human Data Gateway
- Live papers
- EBRAINS Compliance Management, Data Protection, and Data Governance
- KnowledgeSpace

Two of the services, the Human Data Gateway and the Live Papers, are new since the report from October 2020 (D4.1). The Human Data Gateway is an invention created to allow safeguarded sharing of strongly pseudonymised human data. While the data are findable in the Knowledge Graph the access is controlled and coupled to a series of requirements. The Live Papers are a recent addition closely linked to the use of the Knowledge Graph. They serve to facilitate the interactive exploration and re-use of datasets, models and code in the context of the published article with which they are associated.

In this report, we also provide an outline of the extensive community engagement of the EBRAINS Data and Knowledge services, spearheaded by the curation team and with participation of all teams. These outreach and communication activities have served to promote Open Science in general and the EBRAINS Data and Knowledge services in particular.

2. Improvements and growth of services since October 2020

The following sections describe the changes implemented in each of the services since October 2020, as well as the user benefits resulting from these changes and the current work in progress for each service. A first description of the new Human Data Gateway and Live Papers services is also provided.

2.1 EBRAINS Knowledge Graph

The EBRAINS Knowledge Graph ((KG), https://kg.ebrains.eu) is the metadata management system of the EBRAINS Data and Knowledge services. It provides fundamental services for making neuroscientific data, models, and related software FAIR. The end-user accesses the Knowledge Graph via a UI (the FIND data, models, and software service; EBRAINS - Knowledge Graph Search[1]) or programmatically. Driven by the migration to the new metadata standard openMINDS (See EBRAINS Curation, below), the EBRAINS Knowledge Graph has been extended and improved on all its components.

2.1.1 Key components

Table 1 shows key components of the EBRAINS Knowledge Graph, with a description of changes and improvements since October 2020.

[1] https://search.kg.ebrains.eu/

Table 1: Changes and improvements of key components of the EBRAINS Knowledge Graph

Key Component	Description of changes / improvements
Core (T4.1)	The new EBRAINS Knowledge Graph v3 (KG v3) is in place and provided with appropriate data for all environments (test, pre-production and production). The KG Core API has been extended and adapted to all appearing needs of its dependent clients (such as KG Editor and KG Search). Several external services have started to integrate and substantial performance improvements have been achieved especially in terms of the structural reflection of the graph. (Metadata availability has continued to be determined by the access rights of the EBRAINS account holder.) Programmatic access has been simplified by cross-service token reuse as well as a new Python library for KG Core access which is now accessible (first commit on March 3, 2021): https://github.com/HumanBrainProject/kg-core-python/
Search UI (T4.2)	The search UI has been successfully extended to support incoming data from both the old (KG v2) and new (KG v3) Knowledge Graph versions. This is the foundation for the ongoing smooth migration from one system to the other. The Search logic has been optimised to improve appropriateness of search results for full text queries.
Editor (T4.1)	The KG Editor is now fully integrated with the KG Core functionality. It supports the access permission concept of "spaces" (also allowing Collaboratory users to write and read in private spaces in the KG), and also multi-type links, invitation mechanisms, and releasing structures. It has been released for use by the curation services and has proven to be fully functional.
Query Builder (T4.1)	The KG Query Builder has been heavily improved in terms of usability. It now supports features like "single item strategies", type filters, validations (e.g. visualising incomplete query structures) and many more; its functionality is kept in sync with the functionalities provided by the Query API. https://query.kg.ebrains.eu
Statistics (T4.1)	KG Statistics has been upgraded to visualise the content of KG v3 with all its aspects. It has been optimised in terms of performance and system load and has gained in importance as the tool for keeping the overview on what the Knowledge Graph provides. https://stats.kg.ebrains.eu
Automation (T4.1)	The automation systems (containing data ingestion pipelines, content validation and synchronisation with external services) have been updated to operate on the openMINDS metadata structures (DOI registration and synchronisation, automated file structure import, etc.). Additionally, automatically generated metadata can now be highlighted as such e.g. in the KG Editor to clarify the origin of the metadata for the curators.
Metadata monitoring (T4.1)	The metadata monitoring has – based on experiences with the first version – been extended. It is now also in use by the curators for KG v3. With the new version, not only datasets can be tracked but also other research products such as "models" and "software" with the option of further extensions. Additionally, the concept of GitLab boards has been adapted for a so-called "communication" board which allows to coordinate dissemination activities on top of individual KG entries. Here, curators can comment and highlight registered

	research products to be advertised and shared with the responsible persons for communication channels (such as Twitter accounts, public web page, newsletter, etc.). Both the metadata monitoring and the communication board provide an easy-to-use tool (incl. extensive notification mechanisms) to ensure that organisational aspects of the curation and dissemination workflows around the actual metadata in the Knowledge Graph are well integrated.

2.1.2 Associated tools

By the provision of a standardised metadata structure (openMINDS) and convenient APIs for the insertion and consumption, the EBRAINS Knowledge Graph provides various mechanisms to integrate with other services. These include the EBRAINS atlas viewers, which are open to all users and Jupyter Notebooks for scientific analysis or demonstrations / showcasing of methods and results, for users having an EBRAINS account. Furthermore, users can use their EBRAINS account for direct programmatic access to the EBRAINS Knowledge Graph API facilitating smart searches and the building of an individualised environment of integrated tools, including Python libraries like siibra-python[2] and fairgraph[3].

2.1.3 User benefits

The developments since October 2020 have resulted in several improvements for the end-users.

A key achievement is the improvement of the Knowledge Graph API. With the new integration with other EBRAINS services such as the JupyterLab, the token-reuse across systems and the Python libraries (e.g. https://github.com/HumanBrainProject/kg-core-python/) have simplified the overall use of the system for the end-users. Additionally, the simplified registration process for making use of the programmatic API has replaced the previous manual registration and activation of users. It now functions as a "self-service" (the user of the API can activate the programmatic access for dedicated spaces by herself/himself and without waiting time). Related to simplified access to the KG, the changes in the UI of the Query builder now allow exploration of the metadata and its consumption without requiring any kind of knowledge of graph query languages (such as SPARQL, GraphQL) from the end-user.

New possibilities have been created to let users work inside spaces of the Knowledge Graph. With the integration of "Collab spaces", users can register a space in the KG for their EBRAINS Collab. This allows them to manage the (read and write) access to the metadata registered in such a KG space by simply managing their "team" in the collab (viewers can read, editors and admins can write). For individual use, the KG space "myspace" allows any EBRAINS user who has accepted the terms of use of the KG to store and read metadata in a personal space.

The more user-friendly access permission system, and new possibilities to register metadata in small groups before the curation process (in "collab spaces" and "myspace", as outlined above) are key to meet the needs of groups of users or individual end-users. The needs of the curation team have been met with improved tools for the metadata registration and management (updated KG Editor and a new curation monitoring board) and its dissemination (the communication board). The curation team now also profit from more sophisticated automation systems which simplify previously resource-intensive work (e.g. for the registration of huge file repositories).

Finally, the optimisations of the search logic improve the appropriateness of the search results for full text queries.

[2] https://github.com/FZJ-INM1-BDA/siibra-python
[3] https://github.com/HumanBrainProject/fairgraph

2.1.4 Work in progress

Besides the maintenance of the overall system and its compatibility with the various integrating surrounding systems, the EBRAINS Knowledge Graph will mainly focus on the additional features as outlined in Table 2.

Table 2: Work in progress of key components of EBRAINS Knowledge Graph

Key Components	Description of changes / improvements
Core (T4.1)	The continuous integration of the openMINDS meta data structure leads to improvements of the handling of structural specification of the system. The results of the work invested into the back end will be exposed iteratively by extended views and connections in the Search UI, including more simplified programmatic access for end-users.
Automation (T4.1)	Scripts analysing the graph according to validation rules for consistency with the agreed conventions are built. These scripts report back issues to the curation monitoring boards and therefore trigger the process of validation by the responsible curators. Automation processes will be set up to analyse the graph for potential new connections between instances and to infer information from the given metadata.
Search UI (T4.2)	The KG Search service has been prepared to increase the extent of information for meta-data registered in openMINDS while maintaining the old representations. A lot of effort has been put into a more efficient and performant way of displaying file structures and for being prepared for “aspect-oriented” views of data content made possible thanks to the openMINDS structure “file bundles” which allow to specify alternative groupings other than hierarchical file representations (e.g. grouped by subject / study target / …). These additional benefits will be made available to the open public iteratively with the progress of the underlying openMINDS migration. Additional work is ongoing to improve the layout of the result representation to ensure good usability with the extended amount of information represented thanks to the new metadata structures and their additional interconnections.

2.2 EBRAINS Curation

EBRAINS Curation constantly oversees the deposition of metadata for research data, models, and software in the EBRAINS Knowledge Graph. The curation team provides expert neuroscience domain-specific assistance to the depositors at all stages of the curation. This includes restructuring of heterogeneous datasets and interpretation of methodologies in dialogue with the data providers, leading to optimisation and standardisation of the dataset presentation.

Since October 2020 (M7 in Figure 2), the overall number of datasets, models, and software curated and included in the EBRAINS Knowledge Graph has continued to increase. The growth rate was not as high as during the first months of the present phase of the HBP. This was expected, since many of the datasets submitted for curation originate from the HBP, and most of the data from the current phase of the Project are expected to arrive in 2022-2023.

The routines and workflows employed are continuously adjusted to accommodate user-feedback and to gradually increase efficiency while maintaining and improving quality. New developments and optimisations are based on interactions with the data providers as well as input from the broader research community acquired through extensive communication actions (see Section 3). As an open-source project, openMINDS is continuously optimised and collaboratively further developed by

members of the EBRAINS Curation team, the Knowledge Graph team, the KnowledgeSpace team, the Atlas team, as well as external contributors.

An overview of EBRAINS curation is provided in Deliverable D4.1, Section 2.2. Details of the developments since October 2020 are described below.

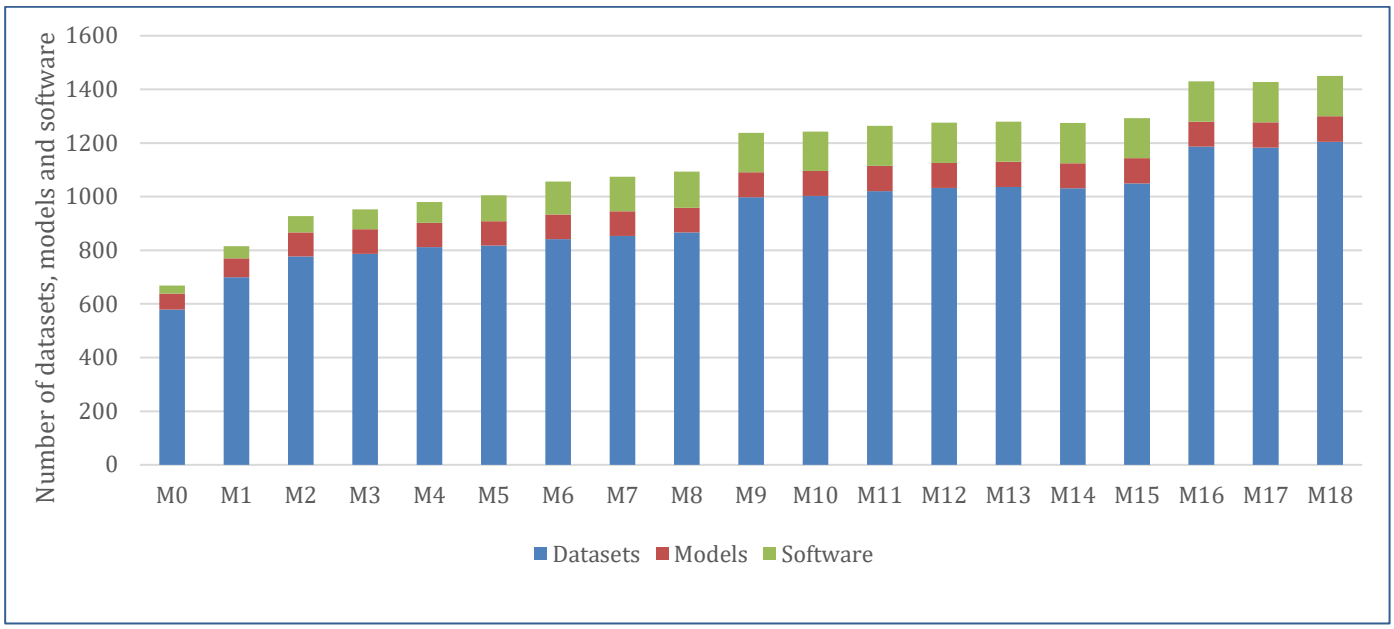

Figure 2: EBRAINS Knowledge Graph increased use

Increase in number of datasets, models and software integrated in the EBRAINS Knowledge Graph from M0 (March 2020) to M18 (September 2021)

2.2.1 Key components

2.2.1.1 Migration to openMINDS

The new Open Metadata Initiative for Neuroscience Data Structures (openMINDS v3.0) was released in November 2020. Since then, a significant effort has been invested in replacing the previously used HBP Minimum Information for Neuroscience Data Sets (HBP-MINDS) with openMINDS as the metadata construct of the EBRAINS Knowledge Graph.

As of November 2021, 85% (1037/1210; status at 24 Nov 2021) of the released datasets in the old Knowledge Graph 2.0 were correctly versioned and migrated to the new Knowledge Graph 3.0 using openMINDS. Most metadata elements of the old Knowledge Graph 2.0 were mapped by experts and automatically migrated. New metadata elements were manually added to the datasets. Starting from December 2021, the migrated datasets will be released via the new Knowledge Graph Search interface.

Key achievements during the transition to KG v3 include:

- Implementations of automatic migration routines for transfer of metadata from the old Knowledge Graph 2.0 to standardised openMINDS elements in the new Knowledge Graph 3.0, and establishment of standardised routines for manual migration of metadata that could not be automatically migrated, and routines for re-curation to adapt to openMINDS
- Expert verified mappings of non-standardised elements of the old Knowledge Graph 2.0 to standardised openMINDS elements in the new Knowledge Graph 3.0
- Implementation of multiple optimisations in openMINDS based on experiences with the migration of data and feedback from users (see below)
- Extension of the openMINDS library for standardised terms and brain atlases

- Alpha-releases of the openMINDS Python library (first release: 22 Feb 2021; latest release: 12 Oct, 2021) for dynamically using openMINDS metadata models and schemas in a Python application to generate openMINDS conformed metadata collections

2.2.1.2 Metadata schemas

openMINDS is managed and maintained on GitHub as an open-source, community-driven project: https://github.com/HumanBrainProject/openMINDShttps://github.com/HumanBrainProject/openMINDS_corehttps://github.com/HumanBrainProject/openMINDS_core. Developers from the EBRAINS team have further developed and optimised openMINDS in the past year together with external contributors. The next release (openMINDS v3) is coming up in December 2021. The changes introduced were driven by the migration process and external feedback.

Major new developments of openMINDS from version 2 to version 3 included:

- Revision of the core metadata model to extend the minimal model to fit better with the requirements introduced by the Common curation track (see below)
- Revision of the core metadata model to provide a better base for automated curation/annotation features of the Knowledge Graph Editor (content type annotation of files and description and interpretation of file repository structures)
- Introduction of a new schema to gather more information around “strains”
- Revision of the SANDS metadata model with optimised versioning of metadata information around brain atlases, the description of quantitative and qualitative relations among anatomical structures, and the overall robustness of the metadata structure. The SANDS schemas have now achieved a maturity level that allows them to be used as an interface for the atlas interface siibra (https://github.com/FZJ-INM1-BDA/siibra-python) to access data and metadata from the Knowledge Graph. The first robust connection is planned to be established in December 2021.
- Extension of the instance library for brain atlases and coordinate spaces (e.g. Julich-Brain Atlas, the Waxholm Space Atlas of the rat brain, or the MNI Colin 27 reference space) provided on GitHub
- Large extension of the instance library for content types (file formats), provided on GitHub
- Integration of the openMINDS_computation extension (see below)
- In-depth schemas for electrophysiology, stimulation methods, and computational provenance entirely rewritten as openMINDS extensions, increasing the degree of interconnection within the Knowledge Graph and eliminating redundancy.

Improvements through openMINDS of the description of EBRAINS research products (datasets, models, software) with visible effects for users of the EBRAINS Knowledge Graph (KG v3) include:

- A more robust and consistent metadata structure that will facilitate programmatically access EBRAINS research products (datasets, models, and software)
- Research products are versioned, and versions are grouped into conceptual (version-independent) landing pages in the Knowledge Graph Search
- ORCID and email as additional information available for data custodians
- Controlled vocabularies implemented for more than 2/3 of the terms used across research products (e.g. techniques, experimental approaches, species, age categories) via the openMINDS instance libraries. Controlled vocabularies can hold additional information such as definitions or synonyms, as well as links to preferred ontologies, the InterLex ontology database, or the KnowledgeSpace encyclopaedia
- Clear differentiation between subjects, subject groups, and tissue samples and tissue sample collections

- Increased data integration: direct links between datasets (pointing to versions or input data), direct links between models and datasets (input and output data), indirect relations between software and datasets or models via content types (input and output format), indirect relations between research products through controlled instances (e.g. age categories, cell types, brain atlases)
- Significantly enriched descriptions of individual files in a dataset, now allowing structured queries for file types and attributes

2.2.1.3 Ontologies

The ontologies reported in Deliverable D4.1 continue to be maintained in a collaboration between the EBRAINS teams and US Neuroscience Information Framework (NIF), thereby facilitating not only improvements for use by EBRAINS but also alignment of terminologies with other projects (e.g., the US BRAIN initiative). A major new development is a new Atlas Ontology Model (AtOM), which is in preparation for release (see below in Section 2.2.3 Work in progress).

2.2.1.4 Curation workflow

The curation workflow begins with the curation request from the data provider and ends with the release of the data, models, or software through the Knowledge Graph. A new version, referred to as the *Curation workflow 3.0*, has been developed during the migration phase. The new version incorporates the large changes in the data-sharing infrastructure that are derived from the migration to openMINDS and the release of the KG v3. It is well documented, with information for the data providers available at the EBRAINS SHARE web pages (https://ebrains.eu/services/data-knowledge/share-data/) and in the EBRAINS Curation Collab (https://wiki.ebrains.eu/bin/view/Collabs/data-curation).

The workflow consists of two tracks: *Common track*, capturing the most essential metadata, and *Advanced track*, providing extensions to capture increasing amounts of metadata. The Common track prioritises speed, efficiency, and high usability, while ensuring a high degree of FAIR compliance. The Advanced track provides options for capturing additional metadata according to the complete openMINDS metadata construct. All data providers are guided through the Common track and will receive advice and suggestions for possible additions. The Advanced track is especially suited for detailed programmatic curation from highly standardised dataset forms (such as BIDS). It is also aimed at well-resourced research projects with the highest ambitions for interpretability and re-use of data.

For the Common track, the EBRAINS Data sharing excel template is used to collect metadata. This template serves the needs of data providers that do not have programmatic skills and wish to have an overview in the curation planning phase. Two new methods for collecting metadata are being developed to provide more opportunities for the data providers (see Section 2.2.3, Work in progress).

All inputs collected are quality checked and revised by a data curation scientist in the curation team.

In the past, the data uploading to the EBRAINS long term data storage suffered from low usability and frequent need for time-consuming assistance from the curation team. These experiences motivated the development of the *Data proxy*. This new online tool greatly simplifies the uploading of data and has been in use from November 2021. It allows data providers with an EBRAINS account to access data in Fenix storage without having to request a Fenix account. The solution is based on a service account which stores the data in Fenix object-storage. The access control to these data is that of the Collaboratory 2, with access permissions to an object container (aka bucket) managed as those of a collab Drive. For information, users are referred to the Data Proxy Collab (see Figure 3): https://wiki.ebrains.eu/bin/view/Collabs/data-proxy/. Also, an API (https://data-proxy.ebrains.eu/api/docs) has been developed, for programmatically interacting with the Data proxy.

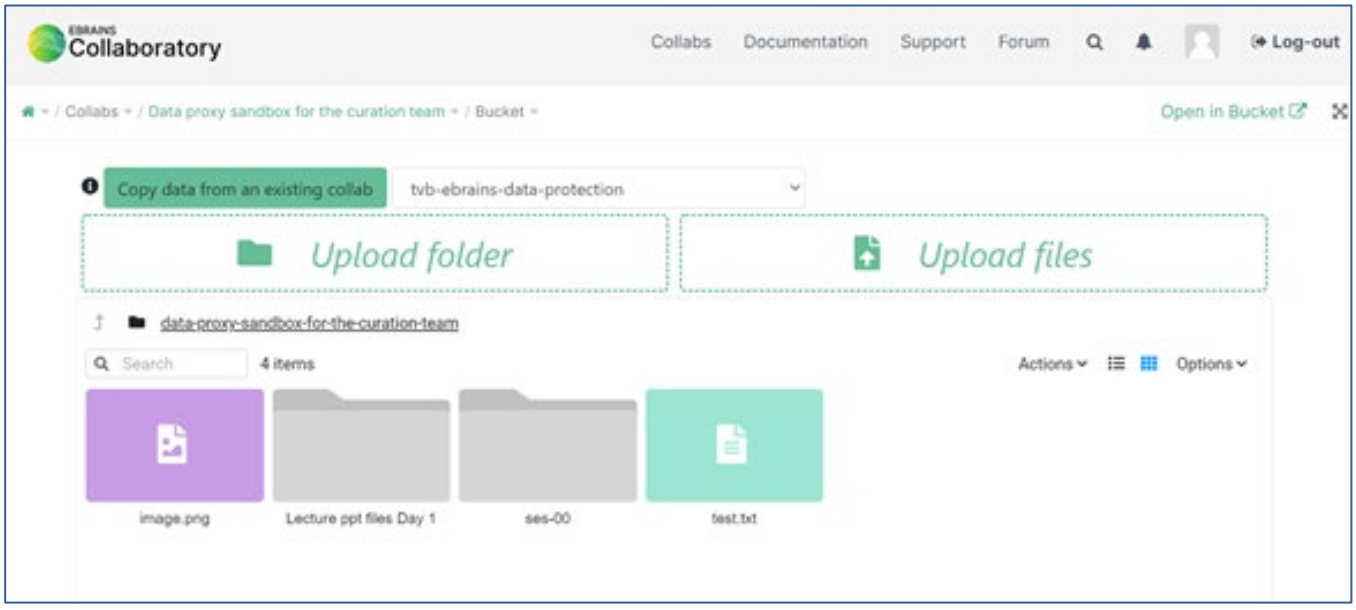

Figure 3: Screenshot of the user-facing end of the Data Proxy upload tool

Finally, the improved *Curation monitoring board* for the new KGv3 (see Section 2.1.1 Key components: Metadata monitoring), provides an overview of the stages of curation and assists the curators in following up of the curation of individual datasets.

2.2.2 User benefits

The most significant improvements available to end-users as a result of developments since October 2020 are summarised below:

The migration to the openMINDS-based Knowledge Graph provides users with improved search opportunities, with access to richer and more uniform metadata. Furthermore, the improved openMINDS documentation provides researchers with guidance in the planning phase, in particular on how to structure the collection of metadata during data acquisition. Such preparations make data sharing easier and allow for adding more metadata, using the Modular or Advanced tracks as outlined above (Section 2.2.1.4).

The new Curation workflow offers several advantages for the data providers. The two new tracks, Common and Advanced track, allow users to make an informed choice about the depth of curation, managing expectations regarding the amount of time they are willing to invest in integrating their research in the Knowledge Graph. The Common track captures the most essential metadata and is therefore a fast and convenient track for data providers who may not have a primary interest in the data sharing. The three alternative methods for metadata collection are tailored to meet the different needs of the large majority of users with limited coding skills to those that can benefit from the use of a programmatic tool.

The Data proxy for simplified uploading of content to the EBRAINS long-term data storage represents a huge improvement for the data providers. It has now been included in the Curation workflow, providing a user-friendly browser-based drag-and-drop upload mechanism that has undergone considerable testing and refinements. The tool has been troubleshooted to ensure minimal file transfer interruptions and corruptions as well as resuming uploads, for a variety of folder and file sizes.

2.2.3 Work in progress

Several openMINDS extensions are under development.

- The openMINDS_computation extension was published on GitHub (first draft: Feb 2021; latest commit: Aug 2021). It covers the description of provenance for simulations, analysis and visualisations, and will be part of the openMINDS v3 release in December 2021.
- The openMINDS_ephys extension was published on GitHub (first draft: Oct 2021). It covers the description of provenance of electrophysiological experiments.
- The integration with all other openMINDS metadata models will follow in 2022. Finally, the openMINDS_publication extension is currently under review by the openMINDS development team. It will cover more detailed metadata around research publications, such as the live papers (see below).
- The remaining in-depth metadata schemas from KG version 2 are being rewritten as openMINDS extensions, and new extensions for fMRI metadata are being developed, taking care to ensure compatibility with the BIDS and NIDM standards.

For the Common track of the curation workflow, two additional methods for metadata collection are being developed:

- The EBRAINS Data sharing web-form (wizard), an alternative to the excel template for collecting metadata, will be introduced in December 2021. This browser-based form is aimed at the large majority of data providers.
- The EBRAINS Data sharing Python library, to be added in 2022, is a programmatic tool, targeting data providers with Python-based local databases. It will be added during the first half of 2022.

The Atlas Ontology Model (AtOM) is being developed to provide more precise specifications about the atlases to which data have been registered. AtOM will define key constituents of reference atlases and will include an ontology model specifying the elements and their relations. It will be applied to mouse, rat, and human brain reference atlases. As for the other ontology developments, AtOM is developed in collaboration with the US Neuroscience Information Framework and is expected to be taken up also by the US Brain Initiative and other major projects globally.

2.3 Human Data Gateway

The Human Data Gateway (HDG) is a new service that allows the sharing of strongly pseudonymised human data on a set of formalised conditions. The metadata of the data included in HDG are openly available in the EBRAINS Knowledge Graph, but the data sets are protected by HDG.

HDG was introduced in February 2021 and developed across multiple teams in the HBP. The initiative to create the service and the initial design originated from EBRAINS curation in close collaboration with the Data compliance team and the HBP Data Governance Working Group. HDG is a response to the needs of multiple data providers who are bringing data of human origin to EBRAINS. HDG covers the sharing of a limited range of data of human origin, i.e. data without direct identifiers and with very few indirect identifiers (strongly pseudonymised, de-identified). It is an extension of the existing services and does not replace the future EBRAINS Service for sensitive data (planned for 2023) which is outside the domain of the current EBRAINS Data and Knowledge services.

Message to users requesting access to data through the Human Data Gateway:

You have requested to access a pseudonymized human dataset shared via EBRAINS. We hereby remind you that the EBRAINS Data Use Agreement defines your responsibilities in accessing this dataset and forbids you from sharing its content aside from sharing the link to the dataset in the EBRAINS Knowledge Graph.

An email has been sent to you to validate your identity and for you to keep as a reminder of the above information. The email contains a link for you to confirm that you consent to the EBRAINS Data Use Agreement.

Your consent and access to the dataset are logged by EBRAINS for audit reasons. The link in the email gives you access to the data for 24 hours. You can request access as many times as you need.

Figure 4: Message to users requesting access to data through the Human Data Gateway

The safeguarding of the data shared under HDG consists of numerous requirements and steps. Data can only be included in the HDG if agreed by the relevant Data Protection Officers, to ensure that the necessary permissions are in place and that the data are suitable for sharing under the conditions provided by HDG. While the existence of the data is public information (the data are discoverable in the EBRAINS Knowledge Graph), access is only given to EBRAINS account holders. The user requesting access to the data will receive the message shown in Figure 4, outlining a series of conditions:

Access is possible through the KG Search UI (see Figure 5) and programmatically, as shown in the following public Collab:

https://wiki.ebrains.eu/bin/view/Collabs/data-proxy/Human%20Data%20Gateway/

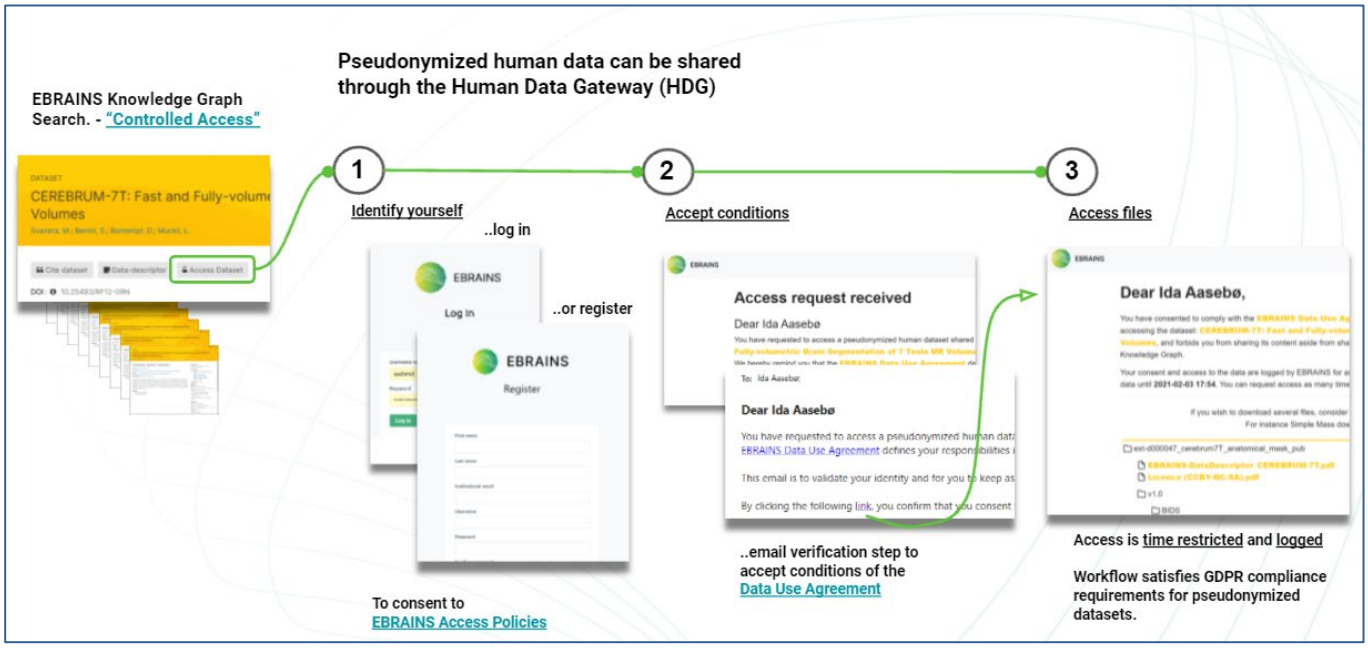

Figure 5: A user perspective on the Human Data Gateway. Step by Step process.

Step wise process from the discovery of a datasets through the KG Search UI, followed by the request to be identified as a user through EBRAINS account login, acceptance of conditions for access, to time-restricted access. Users are thus tracked and the information about their request is stored by EBRAINS.

2.3.1 User benefits

A GDPR-compliant service for sharing of pseudonymised human data to identified users is much needed in the community. From the user perspective, the Human Data Gateway is fast and convenient. It covers a part of the needs of data providers and data users, for a limited group of data referred to as strongly pseudonymised.

2.3.2 Work in progress

Efforts are being made to evaluate data encryption as a possible addition to the Human Data Gateway. A feature of this kind would reduce even further the risks related to the sharing of strongly pseudonymised human data and could potentially allow for a broader group of data to be shared through this mechanism, depending on the individual evaluations of the relevant Data Protection Officers.

2.4 Live Papers

EBRAINS Live Papers is a new service introduced in October 2021. Live Papers are structured and interactive online documents that complement published scientific articles. Live papers include text, figures, links to publicly available research products (datasets, models, or code), and integrated interactive tools that allow users to explore and visualise models and datasets, and to run simulations and Jupyter notebooks.

The Live Papers service is based on the hand-built live papers for modelling and simulation that were developed in the previous phase of the Human Brain Project. This previous incarnation required live paper authors to have experience with web development and was only loosely integrated with the Knowledge Graph.

The new service includes a "Live Paper builder" app accessible with an EBRAINS account through the "Publish your own Live Paper" button at https://ebrains.eu/service/live-papers, and a live paper viewer which is publicly accessible through the "Browse and search Live Papers" button on the same page (all live papers are published with open access licences).

The Live Paper builder allows users to add links to research products from several public repositories, currently EBRAINS Knowledge Graph, the Allen Brain Atlas, NeuroMorpho.org, ModelDB, OpenSourceBrain, and BioModels (see Figure 6). For the Knowledge Graph, a search interface is available; for the other repositories a dataset or model identifier must be entered. Links to Jupyter notebooks in the EBRAINS Lab may also be added, and formatted text and figures can be added using Markdown syntax to give context and explanation. In-development live papers may be downloaded as a JSON document and/or saved to the Knowledge Graph into the private "myspace" or a "collab space" (see Section 2.1.3). The latter option allows in-development live papers to be shared with/edited by collaborators using the standard EBRAINS Collaboratory access controls.

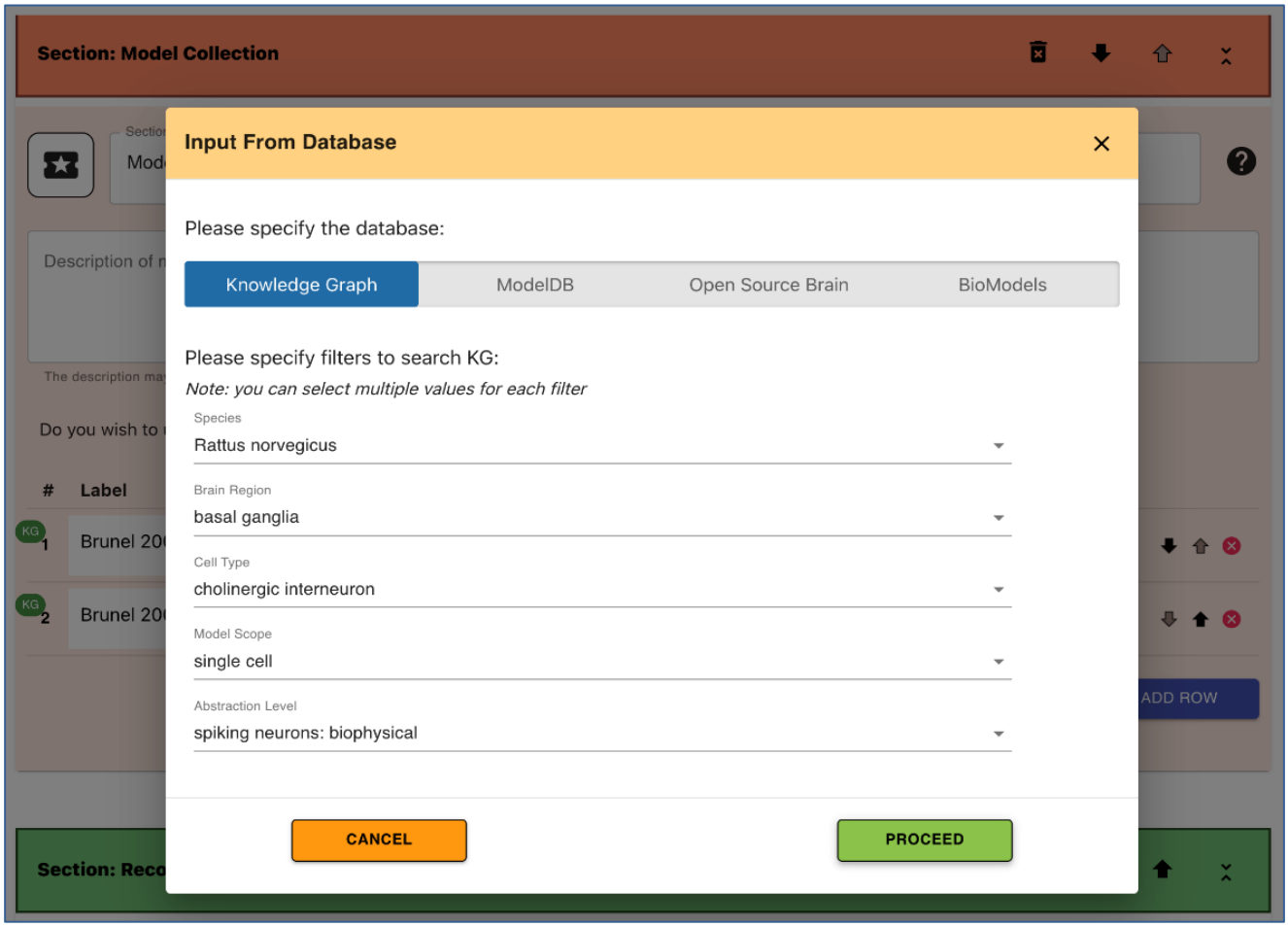

Figure 6. A dialog box from the Live Paper Builder app.

The dialog box allows the user to link to models in public repositories. Similar dialogs are available for electrophysiology data and for morphological reconstructions.

When a live paper is ready for publication the authors may request curation through EBRAINS support. The curation process checks that all linked resources are publicly available in suitable repositories with compatible licences. If any resources are not stored in suitable repositories, the authors are invited to request curation of these resources through the EBRAINS Curation service. Once curation is complete, a DOI is attributed, and the live paper is made available online.

Where the live paper serves as supplementary material to a traditional manuscript, the authors may share it anonymously with reviewers prior to publication by setting a password that can be communicated to the journal editor.

The Live Papers viewer automatically detects resources for which an associated EBRAINS tool is available and either embeds the tool in page or provides a button to launch it in a new browser tab (see Figure 7). Currently supported tools include the NeoViewer widget for electrophysiology data, the HBP Neuron Morphology Viewer for neuronal reconstructions, and the BlueNAAS in-browser simulation tool for models of neurons.

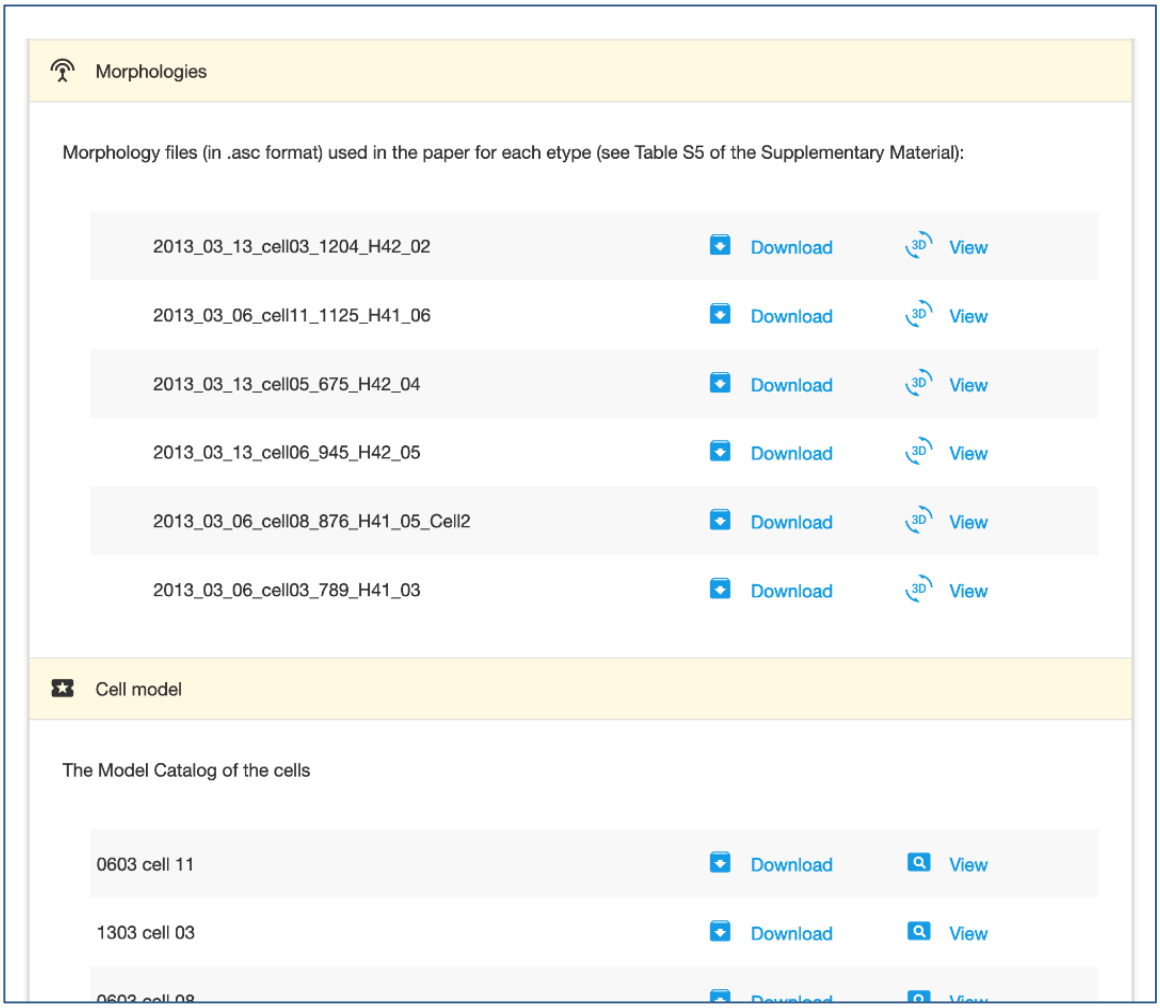

Figure 7. Lists of resources in a published live paper, with buttons to open interactive viewers.

2.4.1 User benefits

By sharing their data, models and code through EBRAINS, scientists satisfy data-sharing mandates according to FAIR principles, and in a way that maximises the potential impact of their work. Since many traditional publications are based on multiple datasets and/or multiple models, Live Papers provide an additional mechanism that groups and contextualises related datasets, code and models, and enables interactive exploration and re-use of these research products.

Live Papers also allow the grouping of resources across different public repositories, not only EBRAINS but also the Allen Brain Atlas, NeuroMorpho.org, ModelDB and others.

Individual published live papers are citable and can be accessed directly through their DOI.

In most cases, Live Papers functions as interactive supplementary material to a published manuscript. While a small number of publishers such as eLife have started publishing their own live papers, the EBRAINS Live Papers service allows authors to associate interactive, reproducible documents with publications in any journal, in a publisher-agnostic way.

2.4.2 Work in progress

The Live Papers service was built on KG v2. An openMINDS extension schema for live papers and other publications (https://github.com/HumanBrainProject/openMINDS_publications), has been developed, and a migration is underway of the API that underlies the builder and viewer apps, to use openMINDS schemas and KG v3. Additional EBRAINS interactive tools will be added as they become available

2.5 EBRAINS Compliance Management, Data Protection, and Data Governance

EBRAINS Compliance Management, Data Protection and Data Governance provide expert advice and support to researchers and EBRAINS users on the topics of Ethics compliance, the governance of both human and animal data, the protection of personal data and research integrity.

In October 2020, we introduced a number of policies and procedures to assist data users and data providers in finding, sharing and using data on EBRAINS platforms and to ensure that all personal data stored on EBRAINS platforms are secure. These included the EBRAINS Access policy[4], EBRAINS Data Use Agreement[5] and EBRAINS General terms of use[6]. Since then, Task T4.5 has introduced the following policies and procedures which further enhance the ability for EBRAINS to deliver:

- The Data Provision Protocol (DPP). This guidance document details all the information that a data provider will need to ensure that their data meet the requirements for sharing on EBRAINS. The DPP includes details of ethical and legal requirements, as well as the technical requirements asked of data providers. This document is shared with EBRAINS data providers as part of the data curation process.
- The EBRAINS Informed Consent Template. To assist researchers in collecting data in a manner which complies with data protection regulations, EBRAINS has developed an Informed Consent Template. This template provides tips, guidance and important data protection advice which can be localised and applied to researchers' own informed consent forms.

Additionally, representatives of the EBRAINS Compliance Management, Data Protection and Data Governance have led international efforts towards responsible data governance through the International Brain Initiative (IBI). Leadership roles in the IBI Strategy committee and the IBI Data Governance Taskforce has led to a publication in Neuron (Eke et al., in press) highlighting the urgent need to address ethical, legal, technical and socio-cultural challenges to data governance in Neuroscience. This is a particularly important piece as technological advancements in AI and machine learning continue to expand the nature, scope and utility of neuroscience data requiring a harmonised international data governance framework. User Benefits

EBRAINS compliance Task continues to work with EBRAINS data providers and data users to advise and support their usage of EBRAINS services. Compliance management works closely with data providers to ensure that the appropriate compliance documentation is collected. The Data Protection Officer supports and advises researchers with any queries they might have with regards to any personal data they are either sharing or looking to access through EBRAINS.

EBRAINS users can, therefore, be assured that data they access on EBRAINS platforms is collected and shared responsibly. Data providers, similarly, can be assured that the data they share through EBRAINS is handled appropriately, that the rights of their data subjects are respected and any queries they might have on towards ethics compliance, data protection or data governance are addressed by experts in those fields.

2.5.1 Work in progress

This service will work with other EBRAINS tools and service representatives to conduct a series of Data Protection Impact Assessments (DPIA). These assessments will ensure that any data processed on EBRAINS platforms or services is processed in a GDPR-compliant manner.

We are also working on a proposal for the governance of animal data. This process should allow EBRAINS to ensure ethical and legal compliance, whilst also encouraging the sharing of a wide variety of animal data from anywhere in the world.

[4] https://ebrains.eu/uploads/EBRAINS_access_policy_v1_1_ed4b84ee9e.pdf
[5] https://ebrains.eu/uploads/EBRAINS_Data_Use_Agreement_90858e7836.pdf
[6] https://ebrains.eu/uploads/EBRAINS_General_Terms_of_use_e457353c1a.pdf

2.6 KnowledgeSpace

The KnowledgeSpace (KS; http://knowledge-space.org) is a community-based encyclopaedia and data catalogue for neuroscience that provides a unique, global interface between brain research concepts and the data, models, and literature that support or weaken their definition. Today, KS contains over 90,000 ontology terms and over 700,000 datasets (excluding literature results from PubMed). As a service for EBRAINS, KS provides Knowledge Graph users with a catalogue of data and models generated outside of the HBP — thus increasing the volume and heterogeneity of the data and models available to EBRAINS users.

An overview of the functionalities of the KnowledgeSpace is provided in Deliverable D4.1, Section 2.4. Details of the developments since October 2020 are described below.

New features of KS include:

- updated user interface
- expanded range of data and models catalogued
- new APIs for 3rd party resources to programmatically access KS

The KS user interface was updated in response to community feedback, solicited at three neuroscience conferences (Faculty for Undergraduate Neuroscience, INCF Assembly, and Society for Neuroscience). Users now have three options for searching for data and models in KS:

1) Dataset-only
2) Literature-only
3) Encyclopaedia (combines dataset and literature with a Wikipedia-like description) with the ability to receive aggregated results from across data/model sources

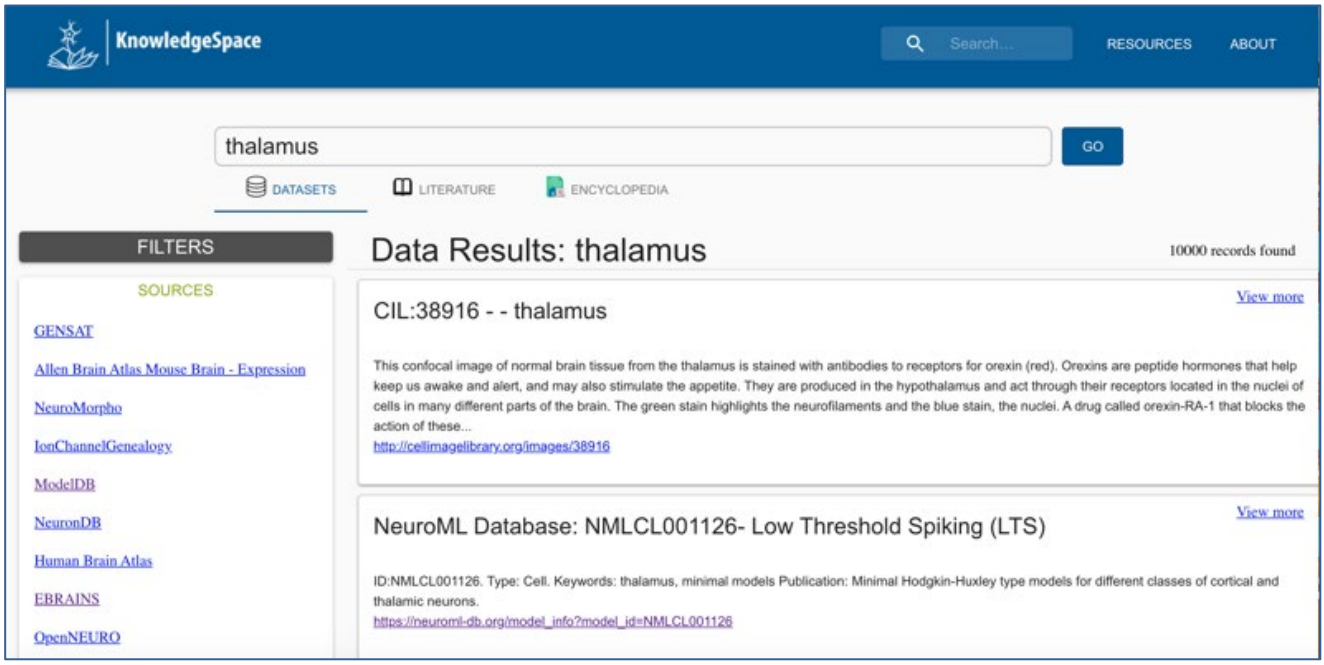

Figure 8: Example of source-specific search result in the KnowledgeSpace

Example of search results. Below the search bar, users have the option to view search results in a dataset-only, literature-only, or encyclopaedia view. In the dataset-only view, users can refine their searches to particular data sources using the filers in the left panel. Also visible are the new data cards which provide users with a summary description of datasets with the ability to expand for a full description.

Previously, users were only able to receive results through the encyclopaedia, which allowed source-specific search returns (see Figure 8). The new data set search enables users to bypass the encyclopaedia and search through single data sources or across all data sources either by a structure name or modality. From an EBRAINS perspective, this new functionality introduces a tighter integration with the Knowledge Graph data and models. The new literature search function enables

users to perform keyword searches in a much simpler fashion than through the previous format through the encyclopaedia.

In addition to the newly implemented three search modes, users can refine their searches by either filtering by data source or by adding additional text in the search bar.

New data cards have also been implemented to facilitate dataset discovery and selection by users (see Figure 9). The new data cards include a summary which is visible in the dataset-only search function with the option to expand which provides users with more detailed metadata about each of the datasets.

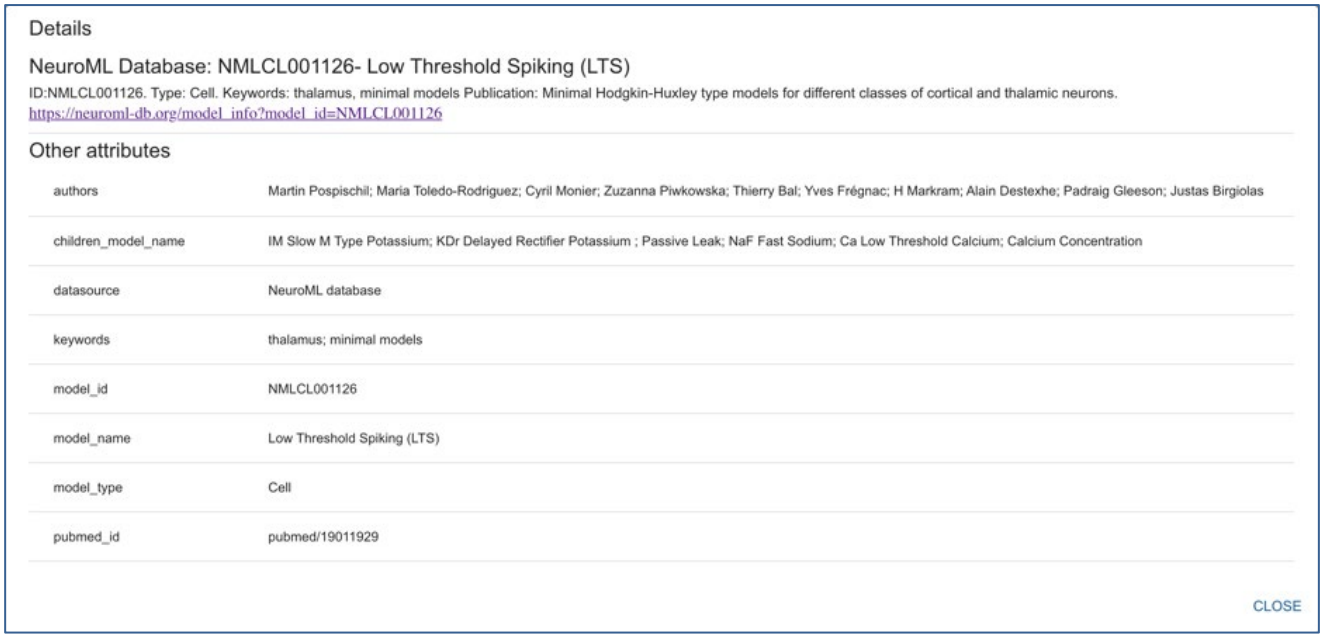

Figure 9: Example of expanded view of a data card

In terms of data, with the addition of the NIH-funded SPARC (Stimulating Peripheral Activity to Relieve Conditions) Project, KS users can now find datasets not only from the central nervous system but also from the peripheral nervous system. Importantly, KS now provides users with a single point of entry to access the datasets generated by four of the world's large-scale brain projects: the HBP, the US BRAIN Initiative, Japan Brain/MINDS, and the Canadian Open Neuroscience Platform (CONP), in addition to the 13 neuroscience repositories that were previously indexed. New APIs have been developed and made publicly available. The range of species (bold indicates new species) indexed in KS now includes: mouse, rat, **human, and macaque**. The range of modalities (bold indicates new types) indexed in KS now includes: gene expression, morphology, models, physiology, **connectivity, histology, microscopy,** and **MRI** in the CNS and **PNS.** The new repositories include:

- OpenNeuro (example search: https://knowledge-space.org/dataspace/scr_005031_openneuro?q=*&term=*)
- DANDI Archive (example search: https://knowledge-space.org/dataspace/scr_017571_dandi?q=*&term=*)
- Canadian Open Neuroscience Platform Portal (CONP; example search: https://knowledge-space.org/dataspace/scr_016433_conp?q=*&term=*)
- Brain/MINDS Portal (example search: https://knowledge-space.org/dataspace/scr_005069_brainminds?q=*&term=*)
- SPARC (example search: https://knowledge-space.org/dataspace/scr_017041_sparc?q=*&term=*)

New APIs have been developed and are made available through Swagger which provides a UI to test the APIs and documentation for accessing the APIs.

2.6.1 User benefits

KS provides users with a single point of entry catalogue for some of the world's leading neuroscience data repositories. Today, KS contains over 90,000 ontology terms and over 700,000 datasets (excluding literature results from PubMed) covering both the central and peripheral nervous system in mouse, rat, human, and macaque over a wide range of modalities (gene expression, morphology, models, physiology, connectivity, histology, microscopy, and MRI). In addition to providing users with a one-stop shop catalogue of datasets from many sources, KS also enables repositories to demonstrate that they are adhering to the FAIR Principles since only repositories adhering to the FAIR principles are indexed in KS.

2.6.2 Work in progress

We will continue to increase the range and types of data and models, improve the user experience, and achieve a tighter integration with the EBRAINS Knowledge Graph, as outlined below.

Expanding the range of data and model types: As new data and models are added to the existing data sources, they will be indexed in KS. We are currently in discussions with the stewards of the Virtual Fly Brain Project and plan to have their data indexed in KS by April 2022.

Improving the user experience:

1) Enhanced search capabilities: The data and model panels in the encyclopaedia will be updated so that users can benefit from the data set search functionality from the encyclopaedia (expanded modalities, inter/intra source search).
2) Curation of terms in the encyclopaedia: This work is ongoing and will be a priority in 2022. We are also in discussion with the US BRAIN Initiative Cell Census Network (BICCN) to include their cell type cards as part of the descriptions of cell types in the encyclopaedia.
3) Tutorials: New multimedia content (including the updated *how to* guide and video tutorial) will be embedded into the KS website.

Achieving tighter integration with the EBRAINS Knowledge Graph: We will explore how components of KnowledgeSpace (descriptions of neuroscience concepts and literature) can be incorporated into EBRAINS. APIs that will enable EBRAINS Knowledge Graph search of KS data and models have been made available to the KG Team in Swagger which provides a UI to test the APIs and documentation for accessing the APIs. A plan has been devised between the KG and KS teams for how to make KS data and models discoverable through KG.

3. Education, Outreach and other interactions with users of the services

The EBRAINS Data and Knowledge services accept data from all domains of neuroscience and therefore address the needs of the entire neuroscience community that wants to deposit or consume data and that looks, amongst other things, for clearly defined terms of use, responsible data compliance, data protection provisions and governance mechanisms.

One of the main priorities of the outreach and dissemination effort has been to clearly communicate the benefits that derive from using the EBRAINS Data and Knowledge services. The EBRAINS Data and Knowledge team focuses particularly on reaching out to data providers and transmitting benefits such as increased impact and visibility of research, public benefits, transparency and opportunities for re-use, proper accreditation when sharing data, and being able to meet requirements from institutions or funders. Data consumers who are motivated by benefits such as validation of own findings, complementing existing data, avoiding duplication of efforts, and fostering collaboration are targeted.

The period covering M8 to M18 has been very active with respect to dissemination events (although mostly virtual due to Covid-19). The next sections provide concrete examples (Sections 3.1, 3.2, 3.3 and 3.4) of the Data and Knowledge team´s engagement:

- Resources have been allocated to preparing better communication tools (online courses and tutorials), that maximise the time investment and make better use of existing collaborations that can assist in making these tools visible to a larger community of neuroscientists. This allows for reuse of high-quality materials and maximises the impact of each dissemination.
- The creation and launch of the new EBRAINS Data Sharing and Knowledge Community Space (publicly launched October 2021).The Community will be a platform to engage with users in a more relaxed way and to inform researchers of the benefits of being part of the larger EBRAINS Community. This effort is also complemented by increased presence in social media platforms like Twitter.

It was originally planned that European Research Council (ERC) grant winners would also be approached by the HBP, with an offer to make EBRAINS resources available to store and manage their data, thus contributing to and using the EBRAINS Data and Knowledge services. However, the COVID-19 pandemic has led to new priorities, and the budget earmarked for ERC grant winners was reallocated to a Call for research into COVID-19-related brain disorders and mental health issues.

3.1 EBRAINS Knowledge Graph

A new collab has been developed that provides users with technical documentation of the EBRAINS KG v3: https://wiki.ebrains.eu/bin/view/Collabs/knowledge-graph

In addition, direct communication (video calls) and hands-on tutorials have been held with various interested parties (e.g. service providers, partners from within the HBP, etc.).

Other forms of interactions with larger audience include presentations at congresses and meetings such as:

- 14 July 2021: NFDI Community Workshop
- 2 September 2021: INCF Workshop

3.2 EBRAINS Curation

Five tutorials and instruction videos have been prepared, which have been included in the new designated space for Curation service video promotions available at the INCF Training Space (https://training.incf.org/course/open-data-neuroscience-data-sharing-ebrains). The different videos cover the following topics:

- What is Open data and EBRAINS data sharing?
- Preparing your data: Data organisation
- Preparing your data: Metadata organisation
- Preparing your data: the Data Descriptor
- Finding data in EBRAINS

A new section on the Community space for EBRAINS Data and Knowledge has been created (https://community.ebrains.eu/_communities/-MeZ8Zm1DT9v5iAjfjsJ/feed, which will allow for increased interaction with the community. This is taking place in addition to increased interactions and greater visibility and presence in different social media channel:

- EBRAINS Curation services Twitter account (https://twitter.com/EBRAINScuration) regularly posts updates, announcements and #DebunkingDataSharingMyths.

- HBP Neuroinformatics platform Twitter account, which started in the SGA2 phase of the Human Brain Project and which currently counts above 1,250 followers and is accessible here: https://twitter.com/HBPNeuroinf

In addition, user documentation for the new and improved openMINDS is now available at https://wiki.ebrains.eu/bin/view/Collabs/openminds and https://github.com/HumanBrainProject/openMINDS

Below is a list with additional contributions to events that took place between M8 and M19:

- 19-23 April 2021: INCF General assembly - booth
- 4 May 2021: Model Validation workshop
- 6 July 2021: CNS*2021 Workshop 10: Training Resources for Cross Initiative Data-driven Modelling Workflow
- 13-14 July 2021: 5th NFDI Neuroscience Community Workshop
- 2 September 2021: INCF Workshop on openMINDS metadata initiative
- 6 September 2021: Mediterranean Seminar for Consciousness (MESEC)
- 7 September 2021: NFDI-Neuro Webinar
- 8 October 2021: HBP Summit

3.3 EBRAINS Compliance Management, Data Protection, and Data Governance

Below is a list of outreach opportunities where Compliance, Data protection and Data governance have been represented:

- Data Protection Impact Assessment Workshop with EBRAINS Service Representatives (29 June 2021)
- Introduction to Data Governance - Capacity Building Workshop (April 2021)
- Contributions to Society for Neuroscience (SfN) and International Neuroinformatics Coordinating Facility (INCF) events on the topics of Data Protection and Ethics Compliance.
- Publication: Damian Eke, Ida E.J. Aasebø, Simisola Akintoye, William Knight, Alexandros Karakasidis, Ezequiel Mikulan, Paschal Ochang, George Ogoh, Robert Oostenveld, Andrea Pigorini, Bernd Carsten Stahl, Tonya White, Lyuba Zehl, Pseudonymization of neuroimages and data protection: Increasing access to data while retaining scientific utility, Neuroimage: Reports, 2021 Volume 1, Issue 4, DOI: 10.1016/j.ynirp.2021.100053 (Publication ID: P2963)
- Publication, prepared in collaboration with the US BRAIN Initiative, Japan Brain MINDS and other leading initiatives in the International Brain Initiative: Damian Eke, Amy Bernard, Jan G. Bjaalie, Ricardo Chavarriaga, Takashi Hanakawa, Anthony Hannan, Sean Hill, Maryann Elizabeth Martone, Agnes McMahon, Oliver Ruebel, Sharon Crook, Edda Thiels, Franco Pestilli: International Data Governance for Neuroscience, Neuron (in press). Preprint: https://psyarxiv.com/esz9b/ (Publication ID: P2968; The HBP acknowledgment is not in the preprint but will be included in the final peer-reviewed publication to be aligned with the EC guideline)

3.4 KnowledgeSpace

The KnowledgeSpace has been presented to the Organization for Computational Neuroscience[7] (OCNS), the Federation of European Neuroscience Societies[8] (FENS), the International Brain Initiative[9] (IBI), and the International Neuroinformatics Coordinating Facility[10] (INCF) Communities.

KS is currently under consideration for serving as the back end for the IBI data catalogue, and discussions are underway with MathWorks to use KS as a data source for a summer coding programme. Interactions with other services have been limited to the repositories of the large-scale brain projects in N. America and Japan (BRAIN, CONP, and Brain/MINDS) where efforts focused on bi-directional indexing of data.

In addition, efforts are underway to refine KS integration in the INCF TrainingSpace[11], an open access hub of multimedia educational resources and tool tutorials.

[7] https://www.cnsorg.org/
[8] https://www.fens.org/
[9] https://www.internationalbraininitiative.org/
[10] https://www.incf.org/
[11] https://training.incf.org/

4. Annex 1: Listing of known problems, delays and risks

This Annex provides an updated risk evaluation building on the risks previously reported in Deliverable D4.1. A status update is provided for each risk, indicating whether or not the risk materialised, mitigation actions, and any foreseen changes related to the risk.

- Risk identified in D4.1: Following the release of the openMINDS core, the integration into the EBRAINS Knowledge Graph has started. This integration consists of a number of steps and requires extensive interactions among the partners delivering the EBRAINS Data and Knowledge services. The full integration is a major effort requiring prioritisation and careful monitoring of the required tasks.

  Status update November 2021: As expected, the migration from the old Knowledge Graph v2 to the new openMINDS-based Knowledge Graph v3 was a major effort, requiring prioritisation and careful monitoring of the required tasks. The migration focused on the visible and stable elements of the old system which could be translated into the new system without involvement of the data providers. Embracing the full power of the new system will require further work in 2022, with further optimisations, extensions and clean-ups, in some cases in collaboration with respective data providers. Any data (a small fraction) that would still remain in the old system will be migrated early 2022, and the hybrid-visualisation of the old and new system via the Knowledge Graph Search guarantees a smooth transition until the migration is completed. The migration remains defined as requiring prioritisation and careful monitoring until fully completed in early 2022.

- Risk identified in D4.1: Several KnowledgeSpace functionalities are depending on developments happening outside of the control of KnowledgeSpace developers. Any delay in the integration of the openMINDS core into the Knowledge Graph will lead to a delay in the integration of data and models from the EBRAINS Knowledge Graph into KnowledgeSpace. Issues or delays in the development of APIs by the US BRAIN Initiative and Brain MINDS would lead to a delay in the indexing of data/models from these initiatives in KnowledgeSpace.

  Status update November 2021: The openMINDS schema was mapped to the KnowledgeSpace data model without delay. This allowed the Knowledge Space data model to be updated, thereby enabling KnowledgeSpace developments to continue while waiting for the integration of the openMINDS core into the Knowledge Graph (the completion of the migration process). Furthermore, following indexing of keywords, a search view for EBRAINS data and models was introduced. While this action does not provide full integration on the ontological level (encyclopaedia), it does provide a mechanism (less specificity) for EBRAINS data and models to be queried in KnowledgeSpace using keywords. Further mitigation actions are so far not deemed necessary.

- Risk identified in D4.1: With a high intensity of efforts focused on developments “under the hood”, and intensive follow up of individual support and curation requests, the much-required efforts of the curation team in promoting the overall EBRAINS Data and Knowledge services may suffer. The resources provided by the Inclusive Community Building team and the Outreach team will be utilised in full to deliver targeted and efficient promotion actions.

  Status update November 2021: This risk did not materialise. We refer to Section 3 in this Deliverable, which describes our outreach and dissemination effort.

- Risk identified in D4.1: The construction of an advanced EBRAINS service for data sharing will have to strike the balance between building the perfect system and creating a user friendly, “good enough”, solution. If this balance is skewed too much toward the perfect system at the cost of user friendliness, potential users may decide to use other services. Communicating the added value of using the service and choosing simplicity, when possible, in the design of the services will be critical.

  Status update November 2021: This risk has not materialised. User friendliness has been prioritised on the Common track for curation, as outlined in Section 2.2.1.4.

- Risk identified in D4.1: The GDPR gives rise to a number of challenges related to the sharing of human data. EBRAINS Data and Knowledge services have, in close collaboration with the Data Governance Working Group of the HBP, created solutions that defines the space that the services will work in. A close collaboration with the Communication team of the project will be critical to ensure optimal visibility and clarity of information in this intricate domain.

  Status update November 2021: Regarding human data, the present EBRAINS Data and Knowledge service is primarily hosting anonymised data. A solution for sharing of strongly pseudonymised data has been introduced and is being carefully monitored.