Table 1

Impact and usage of Blue Brain Nexus at Blue Brain

Data-driven Science Cycle	Challenge	Impact
Data Discovery	Blue Brain organises large amounts of highly heterogeneous data, across the different levels of organisation of the brain. The project requires the ability to search for data to constrain and validate data-driven computational models of ion channels, neurons, synapses, circuits and brain models.	Nexus provides access-control, implementation of a common metadata model for discovery of heterogeneous data, and the adoption of domain specific vocabularies.
Data Acquisition	Data sizes range both in terms of numbers of datasets O(10^6), size of datasets O(100TB), and the need to keep track of data origin and license.	Support for distributed data storage and multiple storage backends (including NFS, S3 and GPFS).
Data Preparation	Data integration of complex and diverse data from biological experiments. Challenging definitions of quality, in which data quality depends on the specific use case.	Versioned data schemas can be revised, open standards readily adopted, extensible metadata annotations, integrated provenance tracking for all data.
Knowledge Discovery	Collaborative and integrative computational modeling and simulation-based neuroscience project involving multiple teams of scientists, engineers, project managers. Requires support for project tracking, scientific objectives of model building, simulation, validation, analysis, visualization and publication via data portals.	Unified schema that integrates project structure, data, workflows, provenance and scientific output. Flexible schema management enables cross-discipline schemas. Dynamic indexing mechanisms enable high performance search adapted to web-based data portals and interactive brain atlases applications.
Data and Knowledge Sharing	Need for publishing data, models and simulation results, interactive web-based data portals and 3D brain atlases.	Custom and extensible indexing using Elasticsearch custom event streams enables high performance data search, all resources have persistent identifiers.

Table 2

Impact and usage of Blue Brain Nexus at the Human Brain Project

Data-driven Science Cycle	Challenge	Impact
Data Discovery	Need to organise large amounts of highly heterogeneous data, across the different levels of organisation of the brain. Data is contributed by scientists at more than 130 universities, teaching hospitals, and research centres across from over 20 countries across Europe. Support for data embargoes, ethics approval and licenses are compulsory.	BBN enables implementation of a common metadata model for discovery of heterogeneous data, and the adoption of domain specific vocabularies. The security features of BBN allow curators to control the access to embargoed data.
Data Acquisition	Data sizes range both in terms of numbers of files O(10^6); and the need to keep track of data origin and license.	Support for distributed data on a federated data and compute infrastructure (FENIX), comprising five high performance computing (HPC) centers across Europe.
Data Preparation	Data curation in HBP proceeds in multiple stages: Tier 1) Basic metadata, making data findable in the EBRAIN Knowledge Graph. Tier 2) Location metadata, increasing data visibility in HBP atlas viewers and exploitation through HBP analytic workflows. Tier 3) Method specific, deep metadata, optimising data accessibility and reusability	Versioned data schemas can be revised, open standards readily adopted, extensible metadata annotations, integrated provenance tracking for all data. BBN enables seamless integration of any vocabularies (e.g. schema.org, neuroshapes.org) to integrate (meta)data in a unifying semantic metadata layer.
Knowledge Discovery	The HBP Collaboratory supports collaboration between teams of neuroscientists, students, and researchers across Europe. The EBRAIN Knowledge Graph provides the unified data search and access for curated data in the HBP.	Unified schema that integrates project structure, data, workflows, provenance and scientific output. Flexible schema management enables cross-discipline schemas. Dynamic indexing mechanisms enable high performance search adapted to web-based data portals and interactive brain atlases applications.
Data and Knowledge Sharing	The EBRAIN Knowledge Graph serves as a primary search engine for data across the project for both project members and the public. It also serves as a foundation for a rich, domain specific software ecosystem.	BBN provides the knowledge graph management capability to HBP and facilitates the creation and maintenance of (meta)data. Its REST API provides a standard interface to create specialized user applications.

Table 3

Impact and usage of BBN at KCNI

Data-driven Science Cycle	Challenge	Impact
Data Discovery	Clinical and research data are stored in separate and sometimes proprietary systems that make it challenging to search and integrate into a common patient-centric view.	Nexus enables secure integration of both clinical health record and research data in a unified data schema.
Data Acquisition	Data includes demographics, vital measures, clinical assessments, pharmacogenomics, brain imaging, EEG, and actigraphy data that are integrated from data-type specific data sources.	Support for distributed data storage; multiple storage backends (including NFS, S3 and GPFS).
Data Preparation	Complex data integration of diverse data from biological experiments. Challenging definitions of quality.	Versioned data schemas can be revised, open standards readily adopted, extensible metadata annotations, integrated provenance tracking for all data.
Knowledge Discovery	Clinical decision support tools require integrated discovery of all patient-related data. Finding patterns and trends in data can improve diagnosis, interventions and outcomes for patients.	Nexus enables a unified view of all patient-related data that can directly be accessed by business-intelligence and machine learning tools to integrate research and care.
Data and Knowledge Sharing	Clinicians require visual decision support dashboards. Patient facing dashboards can support treatment adherence and engagement. A web-based cohort explorer portal can enable third parties to explore aggregate data.	Custom and extensible indexing using Elasticsearch and server sent event streams enables high performance data search, anonymization, data analysis, and aggregation.

Table 4

Impact and usage of Nexus in the Research Data Connectome Project

Data-driven Science Cycle	Challenge	Impact
Data Discovery, Acquisition and Preparation	Finding and reusing the vast amount of dataset generated by researchers and from non-academic sources across Switzerland is a key challenge tackled by the Research Data Connectome specially if the dataset are : i) about different disciplines; ii) stored in fragmented repositories exposing different APIs; iii) serialised in different formats and described using different metadata standards if any. The result is a complex set of siloed repositories complex for researchers to grasp and search from for reuse.	Nexus Forge was used to implement linked data pipelines: i) extracting data from repositories; ii) transforming and validating them so that they are mapped to the project's common SHACL schemas and OWL ontology; iii) storing the resulting normalised, structured and validated data in Nexus Delta. The stored dataset are then made accessible programmatically through Nexus Forge or visually through Nexus Fusion.
Data and Knowledge Sharing	The ability to share in a FAIR, scalable and secured way increasingly large amount of research results generated by Switzerland laboratories and universities to a wide audience is one of the main challenge the Research Data Connectome project is addressing.	BBN enables the project's users to readily create custom views and web-based data portals exposing selected datasets and knowledge to a narrow or broad audience that can then search, access and further share them. The project also developed a Connectome API offering additional points of access to the knowledge graph and insights from it and custom web UIs consuming Nexus Delta REST API.

Table 5

Blue Brain Nexus enables a comprehensive implementation of the FAIR guiding principles for scientific data management. This table focused on how BBN implements the Findable, Accessible and Interoperable principles.

FAIR Principles		**Enabling BBN Features**
Findable	F1. (meta)data are assigned a globally unique and persistent identifier	-To each resource and upon creation, BBN assigns an HTTP(S) based globally unique identifier either provided by the user (e.g a persistent DOI) or automatically generated in the form of a URI combining the BBN instance HTTP(S) address as a prefix as well as UUID as a fragment.
	F2. data are described with rich metadata	- BBN describes each resource with default metadata related to data identification (@type, @id), auditing (e.g. _createdBy, _createdAt, _updatedBy) and management (e.g _rev, _deprecated, _constrainedBy,...) - BBN supports a wide range of metadata description expressiveness provided by users and ranging from plain JSON to full RDF and SHACL enabling metadata to rely on ontologies and be constrained for better quality.
	F3. (meta)data clearly and explicitly include the identifier of the data it describes	- BBN uses JSON-LD as metadata exchange format and each resource identifier is retrieved as value of an always present @id property.
	F4. (meta)data are registered or indexed in a searchable resource	- metadata are indexed as documents in Elasticsearch to support full text and faceted search and as triples in Blazegraph to support graph based queries through SPARQL.
Accessible	A1,(meta)data are retrievable by their identifier using a standardized communications protocol	- Each (meta)data resource is retrievable by its identifier over HTTP(S) protocol via a RESTful API
	A1.1 the protocol is open, free, and universally implementable.	- BBN exposes (meta)data over HTTP(S) protocol which is open, free and is one of the foundation of the world wide web - users can use the Server Sent Event protocol over HTTP to discover in real time new (meta)data added to the knowledge graph
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary.	- HTTP(S) in combination with OpenID Connect and OAuth are used to manage authentication and authorization.
	A2 metadata are accessible, even when the data are no longer available.	- (meta)data can be deprecated in BBN and if so they can't be changed anymore but remain accessible along with their history.
Interoperable	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	- BBN uses RDF and W3C SHACL as metadata and constraints language respectively. RDF and W3C SHACL are W3C open standards with long term support within many open source systems.
	I2. (meta)data use vocabularies that follow FAIR principles.	- the vocabulary used by the BBN technology (eg. audit) follows the FAIR principles - users are encouraged to use open vocabularies that are FAIR compliant such as schema.org, W3C PROV, ...
	I3. (meta)data include qualified references to other (meta)data.	- BBN allows qualified references between (meta)data thanks to its usage of JSON-LD linking capabilities.

Table 6

Blue Brain Nexus enables a comprehensive implementation of the FAIR guiding principles for scientific data management. This table focused on how BBN implements the Reusable principles.

FAIR Principles		Enabling BBN Features
Reusable	R1. meta(data) have a plurality of accurate and relevant attributes.	- (meta)data with many different attributes can be managed within BBN. Values of attributes can be defined in ontologies enabling accurate and fine grained definitions.
	R1.1. (meta)data are released with a clear and accessible data usage license.	- BBN supports (meta)data release through the tag resource enabling users to obtain an immutable reference to a state of their (meta)data that can include user provided license. Users can add license information as resource in BBN and reference them from other metadata resources.
	R1.2. (meta)data are associated with their provenance.	- BBN describes each (meta)data with default minimal provenance information useful for audit: e.g. _createdBy, _createdAt, _updatedBy, _updatedAt, _rev - Furthermore users can leverage BBN flexibility and (meta)data language expressiveness to store and associate (meta)data provenance using standards such as W3C PROV-O.
	R1.3. (meta)data meet domain-relevant community standards.	- BBN is based on open standards for (meta)data representation (RDF) and constraints (W3C SHACL) language as well as exchange (JSON-LD). Those standards are normalised at W3C level and widely adopted within many domains for (meta)data management.