NeuroImage 251 (2022) 118973

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Brain simulation as a cloud service: The Virtual Brain on EBRAINS

Michael Schirner[a,b,c,d,e,*], Lia Domide[f], Dionysios Perdikis[a,b], Paul Triebkorn[a,b,g], Leon Stefanovski[a,b], Roopa Pai[a,b], Paula Prodan[f], Bogdan Valean[f], Jessica Palmer[a,b], Chloê Langford[a,b], André Blickensdörfer[a,b], Michiel van der Vlag[h], Sandra Diaz-Pier[h], Alexander Peyser[h], Wouter Klijn[h], Dirk Pleiter[i,j], Anne Nahm[i], Oliver Schmid[k], Marmaduke Woodman[g], Lyuba Zehl[l], Jan Fousek[g], Spase Petkoski[g], Lionel Kusch[g], Meysam Hashemi[g], Daniele Marinazzo[m,n], Jean-François Mangin[o], Agnes Flöel[p,q], Simisola Akintoye[r], Bernd Carsten Stahl[s], Michael Cepic[t], Emily Johnson[t], Gustavo Deco[u,v,w,x], Anthony R. McIntosh[y], Claus C. Hilgetag[z,aa], Marc Morgan[k], Bernd Schuller[i], Alex Upton[bb], Colin McMurtrie[bb], Timo Dickscheid[l], Jan G. Bjaalie[cc], Katrin Amunts[l,dd], Jochen Mersmann[ee], Viktor Jirsa[g], Petra Ritter[a,b,c,d,e,*]

[a] *Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany*
[b] *Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Department of Neurology with Experimental Neurology, Charitéplatz 1, 10117 Berlin, Germany*
[c] *Bernstein Focus State Dependencies of Learning & Bernstein Center for Computational Neuroscience Berlin, Germany*
[d] *Einstein Center for Neuroscience Berlin, Charitéplatz 1, 10117 Berlin, Germany*
[e] *Einstein Center Digital Future, Wilhelmstraße 67, 10117 Berlin, Germany*
[f] *Codemart, Str. Petofi Sandor, Cluj, Napoca, Romania*
[g] *Institut de Neurosciences des Systèmes, Aix Marseille Université, Marseille, France*
[h] *SimLab Neuroscience, Jülich Supercomputing Centre (JSC), Institute for Advanced Simulation, JARA, Forschungszentrum Jülich GmbH, Jülich, Germany*
[i] *Jülich Supercomputer Centre, Forschungszentrum Jülich, Jülich, Germany*
[j] *Institut für Theoretische Physik, Universität Regensburg, Regensburg, Germany*
[k] *Swiss Federal Institute of Technology Lausanne (EPFL), Campus Biotech, Geneva, Switzerland*
[l] *Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany*
[m] *Department of Data-Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Belgium*
[n] *San Camillo IRCSS, Venice, Italy*
[o] *Université Paris-Saclay, CEA, CNRS, Neurospin, Baobab, Gif-sur-Yvette, France*
[p] *Department of Neurology, University Medicine Greifswald, Greifswald, Germany*
[q] *German Center for Neurodegenerative Diseases (DZNE) Standort Greifswald, Greifswald, Germany*
[r] *Leicester De Montfort Law School, De Montfort University, Leicester, United Kingdom*
[s] *Centre for Computing and Social Responsibility, De Montfort University, Leicester, United Kingdom*
[t] *Department of Innovation and Digitalisation in Law, Faculty of Law, University of Vienna, Austria*
[u] *Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain*
[v] *Institució Catalana de la Recerca i Estudis Avançats, Barcelona, Spain*
[w] *Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[x] *School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Melbourne, Clayton, Australia*
[y] *Baycrest Health Sciences, Rotman Research Institute, Toronto, ON, Canada*
[z] *Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[aa] *Department of Health Sciences, Boston University, 635 Commonwealth Ave., Boston, Massachusetts 02215, United States*
[bb] *CSCS Swiss National Supercomputing Centre, Lugano, Switzerland*
[cc] *Institute of Basic Medical Sciences, University of Oslo, Norway*
[dd] *C. and O. Vogt Institute for Brain Research, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Germany*
[ee] *CodeBox GmbH, Stuttgart, Germany*

[*] Corresponding authors.
E-mail addresses: michael.schirner@bih-charite.de (M. Schirner), petra.ritter@bih-charite.de (P. Ritter).

https://doi.org/10.1016/j.neuroimage.2022.118973.
Received 30 November 2021; Received in revised form 2 February 2022; Accepted 3 February 2022
Available online 4 February 2022.

ARTICLE INFO

Keywords:
Brain modelling
Cloud
Connectome
Neuroimaging
Network model
High performance computing
Reproducibility
Data protection

ABSTRACT

The Virtual Brain (TVB) is now available as open-source services on the cloud research platform EBRAINS (ebrains.eu). It offers software for constructing, simulating and analysing brain network models including the TVB simulator; magnetic resonance imaging (MRI) processing pipelines to extract structural and functional brain networks; combined simulation of large-scale brain networks with small-scale spiking networks; automatic conversion of user-specified model equations into fast simulation code; simulation-ready brain models of patients and healthy volunteers; Bayesian parameter optimization in epilepsy patient models; data and software for mouse brain simulation; and extensive educational material. TVB cloud services facilitate reproducible online collaboration and discovery of data assets, models, and software embedded in scalable and secure workflows, a precondition for research on large cohort data sets, better generalizability, and clinical translation.

1. Introduction

This paper introduces cloud services for brain simulation that are now being offered on the open brain research platform EBRAINS (european brain research infrastructures; ebrains.eu), which makes scientific data, tools, and results accessible to everyone within a protected environment that promotes reproducible work. Scientific studies depend on increasingly complex workflows that are often difficult to replicate and the produced findings are often not confirmed by additional data (Aarts et al., 2015; Ioannidis, 2005). The data and the computational steps that produced the findings as well as the explicit workflow describing how to generate the results were identified as the minimal components for independent reproduction of computational results (Stodden et al., 2016). EBRAINS addresses these challenges by offering modelling and simulation services for collaborative brain research, databases with annotated and curated data of many modalities, atlases of human and rodent brains, image processing workflows, supercomputing resources, neuromorphic systems, and virtual robots. EBRAINS was developed by the Human Brain Project, a research initiative funded by the European Commission with the mission to decode the human brain (Amunts et al., 2019, 2016). TVB cloud services (Tables 1, 2) were developed by the Human Brain Project subproject "The Virtual Brain" in collaboration with the two Human Brain Project partnering projects TVB-Cloud (virtualbraincloud-2020.eu) and TVB-CD (bit.ly/3ogLYtb). To provide supercomputing resources, the Human Brain Project offers as part of the Interactive Computing E-Infrastructure project access to compute and storage resources of the Fenix infrastructure (fenix-ri.eu), a network of six European supercomputing centres.

TVB cloud services are interlinked and make use of EBRAINS cloud services (Fig. 1), which we briefly introduce in the following before focussing on the TVB services. Please see Table 3 for a glossary of technical terms and abbreviations. The 'Collaboratory' (Supplementary Note: The EBRAINS Collaboratory) provides online workspaces, called 'collabs', where research teams can exchange data and work together on documents, secured with access control to restrict usage to authorized users. 'Lab' provides JupyterLab instances for developing applications and running code in a protected environment that cannot be accessed by other users. Jupyter notebooks provide a programmatic interface to EBRAINS services, allow to execute live code and to link processing steps with visualized results and documentation. Data can be found and accessed via the 'KnowledgeGraph', which provides a graphical user interface (GUI) and Application Programming Interface for searching, populating, and editing the data base. The KnowledgeGraph uses controlled vocabularies and ontologies that are mapped with existing neuroimaging and brain simulation ontologies to store data in a structured format, which enables to search the EBRAINS platform for data sets and to identify related information (Supplementary Methods: Data integration and TVB-ready data). In addition, EBRAINS offers services for professional curation of data sets including minting of persistent identifiers like Digital Object Identifiers (DOI; doi.org), licensing, versioning, and setting up of data sharing agreements. RESTful APIs are used for connecting different cloud components, as well as for authentication, data transfer and control of supercomputers. Atlases provide common spatial reference spaces including a multilevel atlas of the human brain as well as the Waxholm Space rat brain atlas (Osen et al., 2019; Papp et al., 2014). The Multilevel Human Brain Atlas uses the Julich-Brain probabilistic cytoarchitectonic maps (Amunts et al., 2020) to link with template spaces such as BigBrain (Amunts et al., 2013) at the micrometre scale and MNI (Das et al., 2016) at millimetre scale, and combines them with imaging-based maps of function (Evans et al., 2012) and connectivity (Guevara et al., 2017). Linking a growing set of multimodal features, the Human Brain Atlas captures brain organization in its different facets.

What are the benefits of a cloud-based research platform? One important advantage are on-demand scalable computing resources. Neuroimaging and brain modelling workflows that are used to analyse large data sets (like the UK Biobank or the Human Connectome Project data sets) require processing power and storage beyond what personal computers can offer. On EBRAINS a network of powerful supercomputers enables to scale computing resources to the needs of a project. Another key advantage of cloud-based research is the ability for interoperable and reusable sharing of data and software, which is an urgent need as there is typically not one individual researcher doing all the work from data acquisition, analysis, hypotheses generation, model building, validation, up to writing and publishing. Rather, it is getting increasingly common that multiple teams, with team members being potentially scattered all around the planet, work together in large projects that require ongoing interaction and synchronization of data and code. Instead of frequently transmitting data sets via the internet and maintaining intricate software environments at multiple computing sites it is more efficient and practical to have a shared platform where teams can work together on datasets and run software in a common computing space. Problematically, sharing of and collaborative work on personal data raises privacy concerns: highly personal and detailed health data like MRI can be misused for malicious intents and must therefore be thoroughly protected, which is reflected in legislation like the General Data Protection Regulation (GDPR) of the European Union. With TVB on EBRAINS we created a software environment that globally implements state-of-the-art security mechanisms like encryption, access control and sandboxing to protect personal data, while at the same time workflows can be flexibly and reproducibly modified using containerized applications. These globally implemented measures for data protection make it easier for individual researchers to protect confidential data and to comply with the law. An additional benefit of TVB on EBRAINS workflows is that mechanisms for data management, provenance tracking and reproducible research are directly embedded using DataLad (Halchenko et al., 2021), which enables explicit tracking of all inputs, codes and processing steps that produced a result in a manner similar to how GitHub (github.com) is used for source code management. Having reproducibility already "built-in" makes it not only easier for the scientist to understand and re-use their own complex workflows years later. More importantly, it makes it also easier for everyone else to understand and use a complex workflow or just individual steps thereof. With simple commands a reviewer, a student, or another researcher can start the entire process or just individual steps and verify the consistency and correctness of the research, or use

Table 1
TVB cloud software, source codes and URLs leading to their main entry points.

Service	URLs
The Virtual Brain	*Web-App*
Brain network simulation	thevirtualbrain.apps.hbp.eu
	Collab
	wiki.ebrains.eu/bin/view/Collabs/the-virtual-brain
	End-to-end use case
	wiki.ebrains.eu/bin/view/Collabs/user-story-tvb
	Source code
	github.com/the-virtual-brain/tvb-root
	Python library
	pypi.org/project/tvb-library
	Container image
	hub.docker.com/r/thevirtualbrain/tvb-run
	Demo brain network model data
	zenodo.org/record/4,263,723#.YYRPgL1Bzxg
TVB Image Processing Pipeline	*Web-App*
Connectome analysis	thevirtualbrain.apps.hbp.eu
	Collab
	wiki.ebrains.eu/bin/view/Collabs/tvb-pipeline
	Source code
	github.com/BrainModes/tvb-pipeline
	Container images
	hub.docker.com/r/thevirtualbrain/tvb_converter
Multiscale Co-Simulation	*Web-App (TVB-Multiscale)*
Two toolboxes for concurrent simulation of large-scale and spiking networks	tvb-nest.apps.hbp.eu
	Collab (TVB-Multiscale)
	wiki.ebrains.eu/bin/view/Collabs/the-virtual-brain-multiscale
	Collab (Parallel CoSimulation)
	wiki.ebrains.eu/bin/view/Collabs/co-simulation-tvb-and-nest-high-computer
	Source code (TVB-Multiscale)
	github.com/the-virtual-brain/tvb-multiscale
	Source code (Parallel CoSimulation)
	github.com/multiscale-cosim/TVB-NEST
	Container image (TVB-Multiscale)
	hub.docker.com/r/thevirtualbrain/tvb-nest
TVB-HPC	*Collab*
Automatic code generation	wiki.ebrains.eu/bin/view/Collabs/rateml-tvb/
	Source code
	github.com/the-virtual-brain/tvb-root
Fast_TVB	*Collab*
Parallelized simulation (multithreading)	wiki.ebrains.eu/bin/view/Collabs/fast-tvb
	Source code
	github.com/BrainModes/fast_tvb
	Container image
	hub.docker.com/r/thevirtualbrain/fast_tvb
Bayesian Virtual Epileptic Patient	*Collab*
Epilepsy modelling	wiki.ebrains.eu/bin/view/Collabs/bayesian-virtual-epileptic-patient
	Source code
	github.com/ins-amu/BVEP
TVB Mouse Brains	*Collabs*
Mouse brain simulation	wiki.ebrains.eu/bin/view/Collabs/tvb-mouse-brains
	wiki.ebrains.eu/bin/view/Collabs/mouse-stroke-brain-network-model/
TVB-ready dataset	kg.ebrains.eu/search/instances/Dataset/a696ccc7-e742–4301–8b43-d6814f3e5a44
SC, FC, and fMRI from tumour patients and controls	
openMINDS metadata for TVB-ready data	*Collab*
Metadata in JSON-LD format	wiki.ebrains.eu/bin/view/Collabs/openminds-metadata-for-tvb-ready-data
	openMINDS schema
	github.com/HumanBrainProject/openMINDS
TVB atlas adapter	*Collab*
Brain atlas	wiki.ebrains.eu/bin/view/Collabs/sga3-d1–1-showcase-1
	Source code
	github.com/FZJ-INM1-BDA/siibra-python; github.com/FZJ-INM1-BDA/siibra-api
	Visualizer
	brainsimulation.org/atlasweb_multiscale
INCF TVB training space	training.incf.org/collection/virtual-brain-simulation-platform
Education and training	

and adapt it for another problem, without necessarily needing domain knowledge about the used software, which helps to make workflows and results more robust and easier to review and reproduce.

In the following we guide readers through the main components of TVB on EBRAINS, highlighting their main features and the respective advantages of cloud-based operation. Subsequently, we demonstrate an end-to-end use case example including the implemented mechanisms for reproducibility and provenance tracking (please see additional use cases in the Supplementary Material). We conclude the main part with a description of data protection mechanisms, the TVB on EBRAINS shared responsibility model, and a discussion. Technical details about the services and their deployments can be found in the Methods section and

Table 2
Publications using software, workflows or data sets underlying different TVB cloud software.

Cloud service	Publications
The Virtual Brain	(Ritter et al., 2013; Sanz-Leon et al., 2013, 2015)
TVB Image Processing Pipeline	(Proix et al., 2016; Schirner et al., 2015a)
Fast_TVB	(Costa-Klein et al., 2020; Schirner et al., 2018; Shen et al., 2019; Zimmermann et al., 2018)
Bayesian Virtual Epileptic Patient	(Hashemi et al., 2020; Jirsa et al., 2017)
TVB Mouse Brain	(Melozzi et al., 2019, 2017)
TVB ready datasets	(Aerts et al., 2020, 2018)
INCF TVB training space	(Matzke et al., 2015)

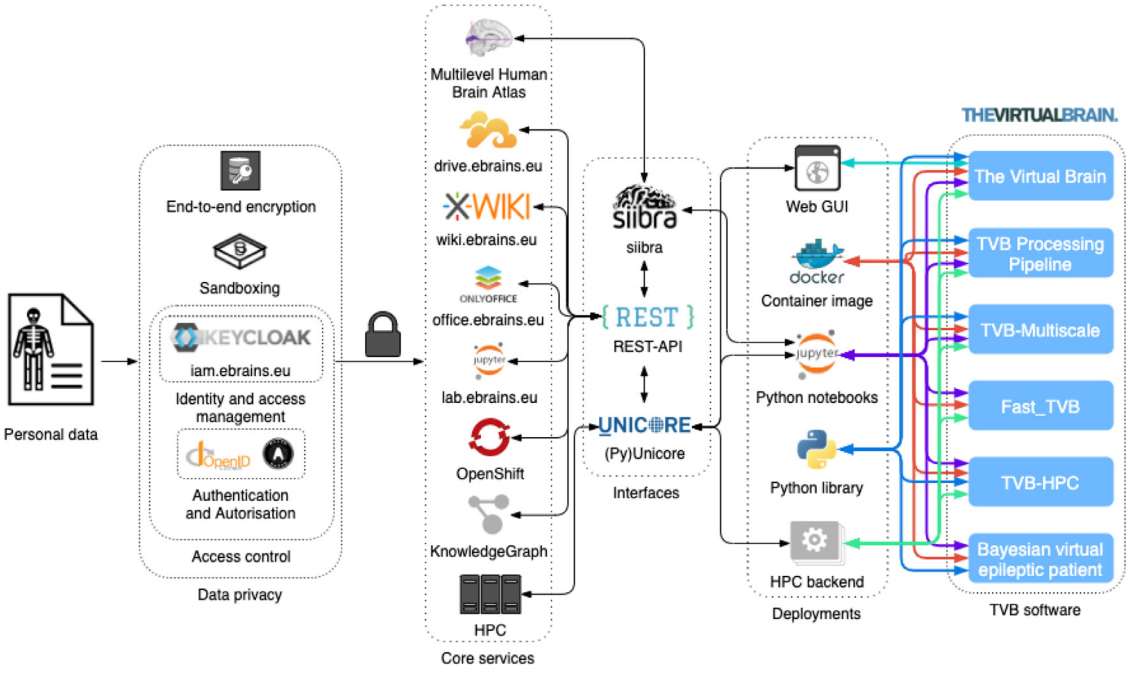

Fig. 1. TVB on EBRAINS cloud services. Human brain network modelling and neuroimaging require personal data applicable to data protection regulation. Encryption, sandboxing, and access control are used to protect personal data. EBRAINS provides core cloud services: the 'Multilevel Human Brain Atlas' provides maps of structure, function, and connectivity in multiple reference spaces; 'Drive' for storing and sharing files; 'Wiki' and 'Office' to create workspaces and documents for collaborative research; 'Lab' for running live code in sandboxed JupyterLab instances; 'OpenShift' for service and resource management; 'HPC' are supercomputers for resource-intensive computations. All software components interact via RESTful APIs and use UNICORE for communication with supercomputers. Software components exist in the form of web GUIs, container images, Python notebooks, Python libraries and high-performance machine codes. Curated scientific results, input and output data can be loaded from and stored into the EBRAINS KnowledgeGraph using openMINDS-compliant metadata annotations to enable efficient and robust sharing and reproducible re-use. The connectors show interactions between different components (colours group connectors according to different forms of software implementation).

exhaustive online documentation (Table 1). Supplementary material provides further information on the different components of TVB on EBRAINS.

2. Results

2.1. The Virtual Brain

TVB (thevirtualbrain.org) is an open-source software for simulating and analysing brain network models, which describe the brain as a graph that is composed of nodes that represent brain areas and edges that represent physical connections between these areas (Supplementary Note: Brain simulation with TVB) (Ritter et al., 2013; Sanz-Leon et al., 2013). TVB can be directly used on EBRAINS from a web GUI (Table 1), without the need to install further software or to have a specific operating system, computing environment or hardware. In addition, TVB can also be used as a Python library for programming in the EBRAINS Lab (Fig. 1). Via these interfaces users can upload brain network models, configure, and run simulations, as well as postprocess and export results. TVB usage is introduced through Jupyter notebooks, explanatory videos, and technical documentation (Table 1). TVB's main documentation is hosted at docs.thevirtualbrain.org.

Importantly, TVB interfaces with supercomputers to rapidly perform simulations that require extensive processing time and storage space. For example, parameter space explorations with hundreds of parameter sets can be simulated in parallel. The web GUI simplifies the process of running high-performance simulations as no further knowledge about supercomputer usage is required: the entire process of sending encrypted data to a supercomputer, decrypting, sandboxed processing, encrypting of results and transmission to the web GUI is handled by the software automatically without any intervention by the user.

2.2. TVB Image Processing Pipeline

Brain network modelling requires a description of the anatomical network that connects brain areas, called structural connectivity, which can be estimated from diffusion-weighted MRI data using the TVB Image Processing Pipeline. The pipeline takes anatomical, functional and diffusion MRI as input and provides as output structural connectivity, region-average functional MRI time series, functional connectivity, brain sur-

Table 3
Glossary of technical terms and abbreviations.

Notation	Description
access control (computer security)	selective restriction to consume, enter or use a resource
annotation of data	categorization and labelling of data
authentication (computer security)	verifying the identity of a computer system user
authorisation (computer security)	specifying access rights and privileges to resources; access control rules are used to decide whether access requests from (authenticated) users shall be granted or not
API; application programming interface	interface that connects computers or software
BIDS	Brain Imaging Data Structure; a standard for organizing neuroscience data
brain network model	system of coupled differential equations for simulating brain activity
checksum	a small block of data that contains information about the contents of another block of data for the purpose of detecting errors
cloud computing	on-demand availability of computing power and storage over the internet
cloud service	infrastructure, platforms, or software made available through the internet
container image, containerization (software)	(creating) executable packages of software that include all dependencies needed to run an application reliably in different computing environments
controlled vocabulary	carefully selected list of words and phrases for unambiguous tagging of units of information
curation	organization and integration of data collected from various sources
data sharing agreement	legal contracts that detail what data are being shared and the appropriate use for the data
differential equation	equation that relates functions and their derivatives (rate at which the value of a function changes with respect to a change of its argument)
EBRAINS	European Brain Research INfrastructureS
encryption	converting information into secret code that hides the information's true meaning
functional connectivity	statistical relationships between brain signals represented as a network; often a matrix of pairwise correlation coefficients between region-average fMRI signals
General Data Protection Regulation	a regulation in European Union law on data protection and privacy with the aim to increase individual's control and rights over their personal data
GUI	graphical user interface
Jupyter notebooks	open-source web application to create and share documents that contain live code, equations, visualizations and narrative text
JupyterLab	web-based interactive development environment for Jupyter notebooks
key (computer security)	a piece of information, which, when processed through a cryptographic algorithm, can encode or decode cryptographic data
knowledge graph	a data model and database for linking, integrating, and storing information in a graph structure
licensing (software)	providing a software product with a legal statement (license) that governs its use and redistribution
metadata	data that provides information (annotations) about other data
metadata schema	a definition how metadata is structured
MRI	magnetic resonance imaging
neuromorphic systems	electronic analogue circuits to mimic neuro-biological architectures
ontology (information science)	a way to organize data, information, knowledge by defining concepts, categories and their relationships
openMINDS	specifications for structuring metadata in neuroscience (github.com/HumanBrainProject/openMINDS)
persistent identifiers	a long-lasting reference to an (often digital) object (e.g., document, file, web page); one example are digital object identifiers (DOI, doi.org), which are widely used to identify publications and data sets
public-key cryptography	a system that uses a different key for decryption than for encryption; this has the advantage that the decryption key needs not to be communicated via insecure channels, while the key for encryption can be known by everyone ("public") without compromising safety
RESTful API	an architectural style for APIs where resources are provided in a textual representation that can be read and modified with a predefined set of operations
sandbox (computer security)	security mechanism for separating running programs in an effort to protect computing systems from failure or attacks, often used to run untrusted programs and code
structural connectivity	aggregated descriptions of the networks that couple neurons, neural populations and brain areas
supercomputer	a computer that is shared by many users and that provides a high level of performance regarding processor time, memory and storage space
TVB	The Virtual Brain, a software to simulate brain network models
UNICORE	interface for exchanging data and commands between different computers in a network (unicore.eu)
versioning (software)	assigning unique version names or unique version numbers to unique states of computer software
version control	tracking and managing changes to software code or data sets
virtual robots	computer simulation of a physical robot

face triangulations, projection matrices for predicting EEG, and brain parcellations. The outputs can be directly uploaded to TVB for brain simulation and analysis. Users can configure and control pipeline steps from the TVB web GUI (Table 1), without needing to directly operate a supercomputer. A workflow orchestrator coordinates the execution of the pipeline and deals with privacy and reproducibility aspects. GUI and orchestrator ensure that the highly personal human brain data can only be accessed by authorized users, that they are always encrypted while at rest or in transit, and that they are only decrypted and processed inside a sandbox that is inaccessible by users of the cloud environment. In addition, the pipeline orchestrator supports provenance tracking and actionable reproducibility: the entire code, data, and all computational steps necessary to reproduce results starting from the raw data can be stored and re-run with a small set of simple commands on a chosen level of granularity, which enables easy reproduction of research results. The pipeline supports flexible processing workflows as it consists of a sequence of container images that can be adapted, exchanged, added, or removed. Containerization makes the pipeline more platform-independent: it can be executed on all similar hardware platforms that support container runtimes like Docker or Singularity. Accordingly, the pipeline serves as a prototypical example for general-purpose protected and reproducible cloud workflows.

2.3. Multiscale Co-Simulation

Multiscale Co-Simulation are two new Python toolboxes for simulating large-scale brain networks with TVB that interact with spiking networks in NEST (Gewaltig and Diesmann, 2007). The toolboxes

provide interfaces to couple the two simulators by connecting the programmatic Python interface of TVB (Sanz-Leon et al., 2013) with PyNEST (Eppler et al., 2009), a Python wrapper for NEST. Multiscale Co-Simulation can be downloaded as standalone container image or used on EBRAINS from Jupyter notebooks (Table 1).

The need for a high-performance environment is for multiscale co-simulations even more important than for single-scale simulations: instead of one resource-demanding simulator there are two and they need to be executed in parallel. Critically, the two simulators need to synchronize to exchange their respective inputs, which is costly because the latency of network interaction is often orders of magnitude higher than the time needed to compute these inputs. To address the involved bottlenecks, the toolboxes implement routines that optimize communication and parallel execution. The Multiscale Co-Simulation project is under ongoing development currently focussing on postulating and validating coupling scenarios between the scales, optimizing the user interfaces as well as optimizing performance. See Supplementary Methods: Multiscale Co-Simulation for more information.

2.4. High-Performance implementations of TVB

Large software products like TVB are often designed with the goal to ease maintainability and long-term development, but that often comes at the cost of non-optimal execution speeds and resource consumption. Algorithms that are not optimized for speed can be orders of magnitude slower than optimized versions: instead of taking days or weeks, a simulation can be done in mere minutes, depending on how it is implemented. Problematically, optimizing computer code for speed is challenging and a task that is largely independent from scientific tasks like postulating and validating a new model: researchers must be put in a position where they can easily manipulate a given model in order to rapidly test hypotheses. To make it easier to simulate high-performance codes, two different strategies were realized. The first one, TVB-HPC (Table 1), automatically produces high-performance codes for CPUs and GPUs using an easy XML-based language called RateML for model specification. RateML is based on the domain-independent language 'LEMS' (Vella et al., 2014), which allows for the declarative description of computational models using a simple XML syntax. The already existing example implementations can be easily adapted to test different models, without requiring any knowledge about algorithmic optimization. The second one, Fast_TVB (Table 1), is a specialized high-performance implementation of the "Reduced Wong Wang" model (Deco et al., 2014; Sanz-Leon et al., 2015). Written in C it makes use of several optimization strategies and a sparse memory layout to efficiently use CPU resources, which makes it possible to simulate extremely large models with millions of nodes even on a standard computer in a reasonable time. Further information and benchmarks are provided in Supplementary Methods: High-performance implementations.

2.5. TVB atlas and data adapters

TVB on EBRAINS provides interfaces for interoperability with different components and services offered on EBRAINS, which enables researchers to plug in different analysis and modelling tools into their custom workflows. While the different TVB components are already interoperable by design, there is a need for 'adapters' that enable to interconnect with other EBRAINS services like the siibra toolbox, which connects TVB with the Human Brain Atlas (Table 1). The Human Brain Atlas characterizes brain regions with a growing set of multimodal features, including transmitter receptor densities (Palomero-Gallagher and Zilles, 2019), cell distributions, and physiological recordings, based on the Julich-Brain cytoarchitectonic maps (Amunts et al., 2020). Aligned with standard brain templates, the Human Atlas can be registered with individual brains to export multimodal microstructural "fingerprints" that can be used to set the parameters of brain models. The siibra adapter gives direct programmatic access to EBRAINS atlas services like selecting a parcellation, browsing and searching brain region hierarchies, and obtaining maps of atlas features like the distributions of cell densities, neurotransmitters, or gene expression data. Internally, siibra connects with repositories like the EBRAINS KnowledgeGraph or the Allen Brain Atlas to retrieve the requested data, hiding the complexity of interacting with different services and minimizing common risks like misinterpretation of coordinates from different reference spaces. Complementary to siibra a viewer was implemented to visualize different atlas maps on the cortical surface (Table 1).

Additional adapters are under development that connect TVB with the Knowledge Graph and the Human Intracerebral EEG Platform to inform brain network model parameterization and to compare simulation results with empirical data. For example, it is planned to link intracranial electrophysiology recordings with the respective Julich-Brain regions to set model parameters based on direct measurements of effective connectivity and transmission delays from stimulation experiments (Trebaul et al., 2018). See Supplementary Methods: TVB atlas and data adapters for more information.

2.6. Data integration and TVB-ready data

Another advantage of cloud-based operation is that research results from different groups can be directly integrated into a central data record where they can be found and re-used by others. This functionality is provided by the EBRAINS KnowledgeGraph, an ontology-based graph data base where data sets are richly annotated with openMINDS metadata in order to ensure their interpretability in the future (Table 1). The openMINDS metadata annotations define an exact classification of research inputs and outputs (for example, empirical recordings, software, articles, books, imaging coordinate systems, reference atlases, models, projects) against a scientific ontology or knowledge framework. To ensure data quality EBRAINS employs a team of expert curators who assist in creating and verifying that data format and metadata annotations fulfil state of the art practices for provenance tracking and data management with regard to long-term availability and interpretability of the results. Data in the KnowledgeGraph is protected by the 'Human Data Gateway', which controls access to human datasets through regulatory compliant data use agreements and access policies. A first example of modelling results that were integrated into the KnowledgeGraph are TVB-ready connectivity data sets in BIDS format from tumour patients and matched control participants. The data set contains region-average fMRI time series, FC, and SC from 31 brain tumour patients before and after surgery, and 11 healthy controls (Aerts et al., 2019). See Supplementary Methods: Data integration and TVB-ready data for more information.

2.7. End-to-end use case with reproducible brain model construction

Upon introducing the individual components of TVB on EBRAINS we now exemplify how they may be combined. Additional use cases are described in the Supplementary Material, especially in the section 'Advanced use cases and training'. To get acquainted with TVB one may start by performing a few test simulations with TVB's default structural connectivity to learn usage of the web GUI and the Python interface; documentation and tutorials explain the steps (Table 1). Visualizing the outputs for different parameter settings and fitting simulation results with empirical data (for example, using functional connectivity) helps to create an intuitive understanding of brain network model dynamics. Next, researchers may want to perform a more detailed analysis, for example, comparing individuals in patient versus control groups to study mechanisms of pathological versus healthy brain dynamics. Here, the researchers can use the TVB Image Processing Pipeline to compute individual structural and functional connectivity from human MRI data. Estimating connectomes from MRI data consists of many complex steps,

making it hard to explicitly track all the necessary provenance data to robustly reproduce a particular configuration of processing steps. Just a minor update of a dependency or an untracked renaming of a file can break the entire workflow and make a result not reproducible. The pipeline uses DataLad (datalad.org) to make its workflow reproducible in an actionable manner: all software and data are tracked in a way that the entire workflow or just individual steps can be easily re-run, archived, published and shared. With DataLad all data and code files are version-controlled and managed in a manner that is comparable to how software is managed with GitHub (github.com), allowing to capture complex hierarchical project structures and all computational steps from raw data to final figures.

When large cohorts are modelled users may find the speed of standard brain model implementations insufficient and switch to TVB's high-performance implementations, which allow fast execution and easy generation of high-performance codes for custom models with TVB's XML-based modelling language RateML. To inform model parameters researchers may decide to include microstructural information from the EBRAINS Human Brain Atlas using the siibra interface (Wang, 2020). Or they may extend large-scale models to encompass finer scales using TVB Multiscale to study hypotheses about brain function that span spatial scales from individual point neurons over populations to whole brain models. In a recent preprint this novel approach was used to study the effect of deep brain stimulation on a spiking basal ganglia model (Meier et al., 2021). Finally, the resulting data outputs can be annotated with metadata, curated, and integrated into the KnowledgeGraph for future reuse by the community.

2.8. Advanced use cases and training

In addition to the introductory use cases described above, EBRAINS provides tutorials for several advanced use cases (Table 1). The Bayesian Virtual Epileptic Patient tutorials showcase how Bayesian inference can be used to compute posterior probability distributions for region-wise parameter settings of TVB's Epileptor model in order to study the spread of epileptic seizures (Jirsa et al., 2017, 2014). The approach makes use of prior distributions obtained from empirical data (for example, a patient's structural connectivity, or lesions detected in MRI) and model simulations to take into account the likelihood for these observations. For example, estimating excitability parameters of an Epileptor brain network model yields a map of region-wise epileptogenicity to guide clinical decision-making. The Virtual Mouse Brain extends TVB with tractography-based as well as tracer-based mouse SC (Melozzi et al., 2017), which was estimated from the Allen Mouse Brain Connectivity Atlas (Oh et al., 2014). Tutorials demonstrate how to export mouse connectivity at different resolutions and how to simulate strokes in mice (Allegra Mascaro et al., 2020). In addition to these notebook tutorials the INCF (International Neuroinformatics Coordination Facility) training space holds a dedicated collection for TVB with didactic use cases, video tutorials, Jupyter notebooks and example data sets (Table 1). See Supplementary Methods: Advanced use cases and training for more information.

2.9. What can go wrong? Common pitfalls of brain network modelling.

Although cloud services make it easier to run scalable modelling workflows there are several limitations to consider. Already one of the first steps, creating a brain network model from MRI data, involves several caveats. One major limitation of MRI tractography is that coupling strengths and time delays of nerve fibre tracts cannot be directly measured (Sotiropoulos and Zalesky, 2019). Identifying and quantifying fibre tracts is based on a mapping from water diffusion to fibre orientations, which is in general an ill-posed problem as MRI voxels are too large to resolve individual fibers. Neither the orientation of fibers in a voxel can be resolved, nor can different arrangements like bending, fanning, crossing or kissing be distinguished. As a result, tractography provides only a model-based approximation of interregional coupling strengths and time delays. Problematically, these approximations are biased by factors like the distance of the regions, algorithmic choices, and individual anatomical properties (Jeurissen et al., 2019; Yeh et al., 2021). Furthermore, even if fibers could be reliably counted, there are several microstructural properties known to influence the strength of coupling that also cannot be directly measured like myelination, axon diameter and synaptic properties, which implicates that tractography results must be interpreted with caution (Jeurissen et al., 2019; Yeh et al., 2021). A related problem is node delineation and the question what is a meaningful parcellation of the brain to form the nodes of a network model? Unlike the microscale, where the mapping between nodes and neurons is obvious, defining nodes at the macroscale is less clear. An intuitive criterion would be functional homogeneity: voxels get grouped based on how similar their activity is, which is plausible, because one model node is usually governed by one type of dynamics. However, matters are complicated by individual structure-function variability. For example, the size of a well-characterized area like V1 can vary twofold in size across subjects (Amunts et al., 2000; Van Essen, 2013), which would be missed by group-level parcellations. Similarly, the scale and the number of nodes heavily impacts the resulting model and they must therefore be aligned with the goals of the research (Proix et al., 2016). For example, the parcellation must be fine enough to be able to represent and differentiate between the specific features of the system that are related to the aims of the research.

Probably one of the biggest challenges is to identify whether a given model can or cannot reproduce a set of observations, which is done in a process called 'inference' that works by comparing modelling outputs with the actual data and selecting the model that explains the observed phenomenon in a way that is deemed optimal. Problematically, already the related task of finding optimal parameter values for a given set of model equations suffers from the so-called 'curse of dimensionality': with each added dimension the space of possible model parameterizations increases exponentially (there is a combinatorial explosion in the possible values that the parameters can jointly take), making it harder to find models that generalize to the typically high-dimensional real-world scenarios in digital medicine (Berisha et al., 2021). Complex mechanistic models are poorly suited for inference, because computing the likelihood for a given observation is typically intractable (Cranmer et al., 2020), as this would require integrating over all potential outcomes of a simulation, the number of which increases exponentially with each model dimension. Likewise, complex systems are often degenerate, producing indistinguishable observations by infinitely many realizations of the same process. While new approaches for "likelihood-free" simulation-based inference are under development (Cranmer et al., 2020), in practical cases often recourse is made to traditional approaches like relying on the insights of scientists into the system to construct powerful summary statistics to effectively compare observed with simulated data. A related problem, especially regarding clinical application, is that models always involve (per definition) enormous simplifications and are often based on assumptions that are only weakly justified and might be very restrictive. Consequently, the conclusions that can be drawn are a function of the validity of the knowledge that was used to build the model and the efficiency with which the verbal knowledge was translated into mathematical equations and then into computer code. Especially in clinical applications false expectations, misinterpretations and overconfidence in simulated results can lead to significant real-life problems. Consequently, these workflows may not be used in a "turn-key" manner and with the expectation that they will automatically produce meaningful results. To produce meaningful results and to adequately interpret them knowledge about modelling and numerical methods as well as neuroscience domain knowledge are fundamentally necessary.

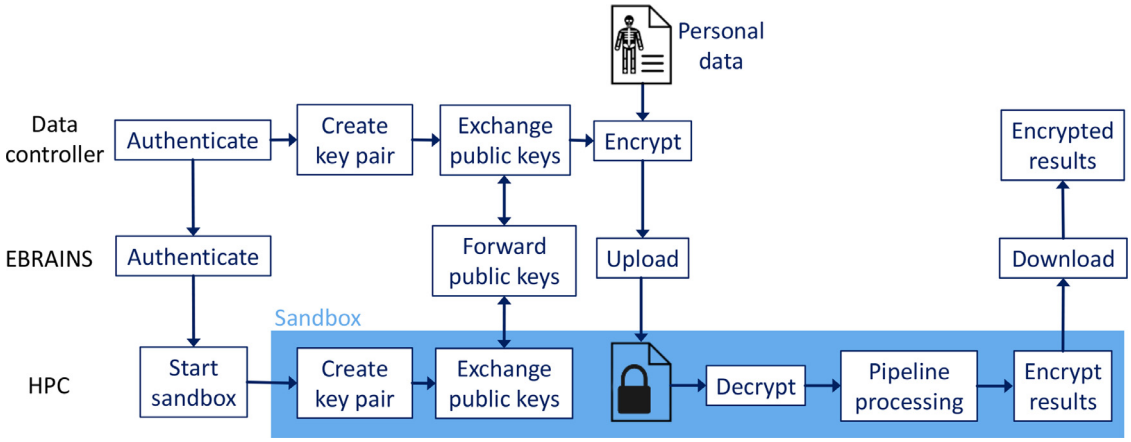

Fig. 2. Securing personal data processing workflows in shared environments. Access control ensures that only authorized users can access sensitive data. Sensitive data is encrypted with public-key cryptography on the data controller's computer before upload to the cloud. The key pair for upload is generated within a sandboxed process at the final processing site and the private key never leaves the sandbox. This ensures that the data can only be decrypted at the final processing site and that no human gets into possession of the key for decryption. All processing is performed in the sandbox and personal data is never written outside the sandbox in unencrypted form. A public key generated by the data controller is used for returning encrypted results, which ensures that only the data controller can decrypt the data.

2.10. Data protection in the TVB on EBRAINS cloud

Biomedical research is facing challenges because many methods lack technical infrastructure to protect the privacy of sensitive data. Research often involves that teams exchange and process sensitive data on shared infrastructure like the internet and high-performance computers, which poses risks for illegitimate access. Consequently, an important requirement for privacy protection is to enable secure processing of sensitive data in shared infrastructures, as the involved networks and computers can be accessed by many human and non-human users with only logical separation between them. Cloud platforms have the advantage that privacy technology and legal compliance measures can be globally implemented and offered as a standardized and certified service, which makes it easier for the individual researchers to overcome technical and organizational hurdles for demonstrating compliance with data protection law. The European Union's General Data Protection Regulation (GDPR) and similar international and national laws impose restrictions on the processing of personal data including storage and sharing. Problematically, biomedical data cannot be easily anonymized or pseudonymized such that all potentially identifiable information are removed, and potential re-identification is excluded (Byrge and Kennedy, 2018; Gymrek et al., 2013; Rocher et al., 2019). A principle means of ensuring GDPR-compliant data processing is the implementation of technical and organizational measures to ensure a level of security appropriate to the risk of the processing (Article 32 GDPR). To protect data by design and default (Article 25 GDPR), TVB on EBRAINS implements access control, public-key cryptography, and sandboxing (Fig. 2).

Access control mechanisms, like the TVB web GUI, hide direct access to systems where sensitive data are actively processed: users need to log into the GUI with their password and can only access data that they uploaded or created themselves or that was made available to them through the role-based access control and permission management functionalities of the EBRAINS Collaboratory (see Supplementary Note: The EBRAINS Collaboratory). Sensitive data is encrypted before upload to EBRAINS and remains encrypted at all times with the only exception being the time when a processing job is actively executed. Cryptographic keys are created ad-hoc and independently for each processing job and the system is designed such that no human gets into possession of the decryption key while the data is in the cloud: the sensitive data can only be decrypted at their final processing site by an automatic procedure. During the actual processing sensitive data may exist in unencrypted form, but only within isolated temporary memory locations that cannot be accessed by other users of the system (sandboxes). See Supplementary Methods: Data protection in the TVB on EBRAINS cloud for more information.

2.11. Shared responsibility & compliance

In addition to technical measures also organizational aspects must be considered for processing to be lawful. The GDPR describes two roles for lawful processing of personal data: data controllers and data processors. Data controllers are responsible for, and required to be able to demonstrate, compliance with GDPR (Art. 5, GDPR), by implementing technical and organisational measures that ensure appropriate security of the personal data (Art. 24, GDPR). In contrast, data processors process personal data only on behalf of data controllers, acting under the authority of the controller to carry out the processing (Art. 28/29, GDPR). When a user uses TVB on EBRAINS services to process personal data the user is always the data controller, while EBRAINS as a service provider is always the data processor, because the user is directing the processing through its interaction with the offered services, while EBRAINS is only executing the provided instructions. As data processor EBRAINS is responsible for protecting the global infrastructure with documented procedures and services on behalf of the user. As data controller a user maintains control over the data that it hosts or processes with TVB on EBRAINS, as mechanisms were put in place to prevent unauthorized access and to enable that data controllers can independently or jointly determine the means of the data processing. To use TVB cloud services a user must therefore agree to terms that clarify its personal responsibility regarding compliance with GDPR with respect to security precautions, access permissions, contact persons, personal responsibilities, monitoring, logging, and passing of information to third parties (ebrains.eu/terms).

3. Discussion

TVB cloud services were developed to lower the barriers to brain simulation and connectome analysis. They offer reproducible and protected workflows for collaborative computational neuroscience research. All codes are open source and available for download from EBRAINS and GitHub (Table 1). Software is packaged in container images that can be directly used without the need to install dependencies. Several software and data components have been peer-reviewed, and results were

published in academic journals (Table 2). To enable actionable reproducibility the image processing workflow is equipped with tools for data management and provenance tracking. All computational steps, inputs and software are tracked, and each step can be easily rerun and verified with a simple set of commands. Technical and organisational measures for protecting the privacy of personal data are globally implemented into the services offerings of the platform, making it easier for researchers to demonstrate compliance with data protection regulation. Access control, encryption and sandboxing ensure that sensitive data stays confidential. Comprehensive documentation in the form of manuals, tutorials, lectures, Jupyter notebooks, demo data, workshops, videos, use cases, mailing lists and support contacts provide efficient and didactic dissemination of knowledge and support. EBRAINS core services enable to map and organize complex projects by large remote teams into a persistent and replicable structure at a central and secure place, which makes it easier to pick up projects at a later time. The flexibility of the platform and its focus on community-driven research enable rapid adoption of advances in brain simulation and connectomics, as well as correction of errors. Technical and organisational security mechanisms are designed to provide highest data protection standards, while at the same time providing the required flexibility to enable state-of-the-art research. To keep the high quality of the cloud services, ongoing and future efforts are directed towards the continuous integration of improved community standards and best practices. The TVB on EBRAINS ecosystem can be transferred to other cloud environments within the European Open Science Cloud or beyond. Thus, it serves as a reference architecture for secure processing and simulation of neuroscience data in the cloud (Fig. 1 and Supplementary Discussion).

4. Methods

4.1. The Virtual Brain

The methods behind the main TVB neuroinformatics simulator are extensively described in several publications (Ritter et al., 2013; Sanz-Leon et al., 2015, 2013) and in online documentation (Table 1; docs.thevirtualbrain.org). To deploy TVB as cloud service it was implemented as container image executed on OpenShift, an open source container orchestration platform. This deployment serves TVB's GUI via the web and automatically scales the number of running instances of the TVB container depending on demand. The GUI is connected with the EBRAINS identity and access management system to perform access control: only registered EBRAINS user can access the GUI and they can only access the data for which they were given role-based permission. Depending on their complexity, simulation jobs are either directly computed in the running OpenShift instance that serves the web GUI or on a supercomputer. Currently users still have the responsibility to manually encrypt their data with a public key before upload, but in a next release it is planned that this will be automatically performed by the upload function. After upload every project is individually re-encrypted with a dedicated key. Decryption only happens when a user opens a project in the web GUI and the decrypted data is immediately deleted when the project is closed or the user logs out. The decrypted project is not directly written to a file system, but only stored inside the running container. For high-demand operations that run on the supercomputer data is only decrypted after the job gets started by the job scheduler and only inside the running TVB container. See Supplementary Methods: Brain simulation with TVB for more information.

4.2. TVB Image Processing Pipeline

The TVB Image Processing Pipeline (Schirner et al., 2015b) allows users to select and combine dedicated neuroimaging workflow containers, like BIDS Apps (see Supplementary Note: BIDS Apps), into reproducible workflows that process MRI data on supercomputers while protecting the privacy of personal data in compliance with data protection regulation. Containerization makes it easier to deploy neuroimaging workflows, as they often rely on a high number of dependencies and computational steps. Users can select amongst different neuroimaging containers like fmriprep for functional MRI processing (Esteban et al., 2019), Mrtrix3_connectome for diffusion MRI tractography (Smith and Connelly, 2019; Tournier et al., 2019), or the Human Connectome Project pipelines for both (Glasser et al., 2013). Like main TVB, the pipeline execution on the supercomputer can be controlled from the TVB web GUI without giving users direct access to the supercomputer. An orchestrator program on the supercomputer coordinates the execution of the container images and ensures that personal data is encrypted at all times, except for the duration of the processing and then only in the main memory of a sandboxed process (Fig. 2 and Supplementary Note: TVB Image Processing Pipeline for more details). To make workflow processing reproducible the open source distributed data management solution DataLad (datalad.org; (Halchenko et al., 2021)) was used for version control and provenance tracking: all files involved in a workflow (such as data, code and computational environment) are stored within nested directory trees, which allows to explicitly store the evolution of a data set from its raw state to the final result. Checksums allow the user to uniquely identify the contents of every file, which in turn allows to verify the correct execution of every computational step and thereby full computational reproducibility of the entire workflow. See Supplementary Methods: TVB Image Processing Pipeline for more information.

Declaration of competing interests

The authors declare no competing interests.

Credit authorship contribution statement

Michael Schirner: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing, Software. **Lia Domide:** Resources, Software, Writing – review & editing. **Dionysios Perdikis:** Methodology, Resources, Software, Writing – original draft, Writing – review & editing. **Paul Triebkorn:** Formal analysis, Methodology, Software, Writing – review & editing. **Leon Stefanovski:** Data curation, Formal analysis, Writing – review & editing. **Roopa Pai:** Conceptualization, Data curation, Formal analysis, Writing – review & editing. **Paula Prodan:** Software, Writing – review & editing. **Bogdan Valean:** Software, Writing – review & editing. **Jessica Palmer:** Visualization, Software, Writing – review & editing. **Chloê Langford:** Visualization, Software, Writing – review & editing. **André Blickensdörfer:** Visualization, Software, Writing – review & editing. **Michiel van der Vlag:** Methodology, Software, Writing – original draft, Writing – review & editing. **Sandra Diaz-Pier:** Methodology, Software, Writing – review & editing. **Alexander Peyser:** Methodology, Software, Writing – review & editing. **Wouter Klijn:** Methodology, Software, Writing – review & editing. **Anne Nahm:** Resources, Writing – review & editing. **Oliver Schmid:** Data curation, Methodology, Software, Writing – review & editing. **Marmaduke Woodman:** Methodology, Software, Writing – original draft, Writing – review & editing. **Lyuba Zehl:** Data curation, Methodology, Software, Writing – review & editing. **Jan Fousek:** Methodology, Software, Writing – review & editing. **Spase Petkoski:** Methodology, Software, Writing – original draft, Writing – review & editing. **Lionel Kusch:** Methodology, Software, Writing – original draft, Writing – review & editing. **Meysam Hashemi:** Methodology, Software, Writing – review & editing. **Daniele Marinazzo:** Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Jean-François Mangin:** Methodology, Resources, Writing – original draft, Writing – review & editing. **Agnes Flöel:** Resources, Writing – review & editing. **Simisola Akintoye:** Supervision, Writing – review & editing. **Bernd Carsten Stahl:** Supervision, Writing – review & editing. **Michael Cepic:** Supervision, Writing – original draft, Writing – review & editing. **Emily Johnson:** Supervision, Writing – orig-

inal draft, Writing – review & editing. **Gustavo Deco:** Conceptualization, Software, Supervision, Writing – review & editing. **Anthony R. McIntosh:** Conceptualization, Supervision, Writing – review & editing. **Claus C. Hilgetag:** Conceptualization, Resources, Supervision, Writing – review & editing. **Marc Morgan:** Resources, Supervision, Writing – review & editing. **Bernd Schuller:** Methodology, Software, Writing – review & editing. **Alex Upton:** Resources, Writing – review & editing. **Colin McMurtrie:** Resources, Supervision, Writing – review & editing. **Timo Dickscheid:** Conceptualization, Methodology, Resources, Software, Supervision, Writing – review & editing. **Jan G. Bjaalie:** Conceptualization, Resources, Supervision, Writing – review & editing. **Katrin Amunts:** Conceptualization, Resources, Supervision, Writing – review & editing. **Jochen Mersmann:** Conceptualization, Methodology, Resources, Software, Supervision, Writing – review & editing. **Viktor Jirsa:** Conceptualization, Resources, Supervision, Writing – review & editing. **Petra Ritter:** Funding acquisition, Methodology, Project administration, Resources, Visualization, Supervision, Writing – original draft, Writing – review & editing.

Data availability

The datasets generated and analysed during the current study are available in the EBRAINS KnowledgeGraph repository, search.kg.ebrains.eu. KnowledgeGraph is a DOI-minting repository where EBRAINS data and software services are indexed and referenced. To share personal data amongst researchers subject to national or international data protection laws (e.g., the General Data Protection Regulation of the European Union) protected environments and workflows as well as a shared responsibility model were created (Supplementary Methods: Data protection in the TVB on EBRAINS cloud).

Code availability

All software codes presented in this article have open-source licenses and can be downloaded from GitHub (Table 1). The EBRAINS database service KnowledgeGraph (search.kg.ebrains.eu) is a DOI-minting repository where all data and software services are indexed and archived. Source codes are deployed as cloud services that can be used on ebrains.eu and as standalone download versions that can be pulled as container images from Docker Hub (Table 1).

Acknowledgments

We gratefully acknowledge the Swiss National Supercomputing center CSCS for supporting this project by providing computing time through the Interactive Computing E-Infrastructure (ICEI) on the Supercomputer Piz Daint of the Fenix Infrastructure (projects ich10, ich12) and the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for supporting this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). We acknowledge support by H2020 Research and Innovation Action grants Human Brain Project SGA2 785907, SGA3 945539, ICEI 800858, VirtualBrainCloud 826421 and ERC 683049; Berlin Institute of Health & Foundation Charité, Johanna Quandt Excellence Initiative. Several computations have also been performed on the HPC for Research cluster of the Berlin Institute of Health. We acknowledge the use of Fenix Infrastructure resources, which are partially funded from the European Union's Horizon 2020 research and innovation programme through the ICEI project under the grant agreement No. 800858. German Research Foundation SFB 1436 (project ID 425899996); SFB 1315 (project ID 327654276); SFB 936 (project ID 178316478); SFB-TRR 295 (project ID 424778381); SPP Computational Connectomics RI 2073/6–1, RI 2073/10–2, RI 2073/9–1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.118973.

References

Aarts, A.A., Anderson, J.E., Anderson, C.J., Attridge, P.R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F.A., Bowman, S.D., Brandt, M.J., Braswell, E., Brohmer, H., Brown, B.T., Brown, K., Brüning, J., Calhoun-Sauls, A., Callahan, S.P., Chagnon, E., Chandler, J., Chartier, C.R., Cheung, F., Christopherson, C.D., Cillessen, L., Clay, R., Cleary, H., Cloud, M.D., Conn, M., Cohoon, J., Columbus, S., Cordes, A., Costantini, G., Alvarez, L.D.C., Cremata, E., Crusius, J., DeCoster, J., DeGaetano, M.A., Penna, N.D., 2015. Estimating the reproducibility of psychological science. Science (80–) doi:10.1126/science.aac4716.

Aerts, H., Schirner, M., Dhollander, T., Jeurissen, B., Achten, E., Van Roost, D., Ritter, P., Marinazzo, D., 2020. Modeling brain dynamics after tumor resection using The Virtual Brain. Neuroimage doi:10.1016/j.neuroimage.2020.116738.

Aerts, H., Schirner, M., Jeurissen, B., Dhollander, T., Van Roost, D., Achten, E., Ritter, P., Marinazzo, D., 2019. TVB time series and connectomes for personalized brain modeling in brain tumor patients [Data set] doi:10.25493/1ECN-6SM.

Aerts, H., Schirner, M., Jeurissen, B., Van Roost, D., Achten, E., Ritter, P., Marinazzo, D., 2018. Modeling brain dynamics in brain tumor patients using The Virtual Brain. eNeuro doi:10.1523/ENEURO.0083-18.2018.

Allegra Mascaro, A.L., Falotico, E., Petkoski, S., Pasquini, M., Vannucci, L., Tort-Colet, N., Conti, E., Resta, F., Spalletti, C., Ramalingasetty, S.T., von Arnim, A., Formento, E., Angelidis, E., Blixhavn, C.H., Leergaard, T.B., Caleo, M., Destexhe, A., Ijspeert, A., Micera, S., Laschi, C., Jirsa, V.K., Gewaltig, M.O., Pavone, F.S., 2020. Experimental and computational study on motor control and recovery after stroke: toward a constructive loop between experimental and virtual embodied neuroscience. Front. Syst. Neurosci. 14, 1–34. doi:10.3389/fnsys.2020.00031.

Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., Lippert, T., 2016. The human brain project: creating a european research infrastructure to decode the human brain. Neuron doi:10.1016/j.neuron.2016.10.046.

Amunts, K., Knoll, A.C., Lippert, T., Pennartz, C.M.A., Ryvlin, P., Destexhe, A., Jirsa, V.K., D'Angelo, E., Bjaalie, J.G., 2019. The human brain project—synergy between neuroscience, computing, informatics, and brain-inspired technologies. PLoS Biol doi:10.1371/journal.pbio.3000344.

Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.É., Bludau, S., Bazin, P.L., Lewis, L.B., Oros-Peusquens, A.M., Shah, N.J., Lippert, T., Zilles, K., Evans, A.C., 2013. BigBrain: an ultrahigh-resolution 3D human brain model. Science (80) doi:10.1126/science.1235381.

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K., 2000. Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? Neuroimage 11, 66–84.

Amunts, K., Mohlberg, H., Bludau, S., Zilles, K., 2020. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. Science (80).

Berisha, V., Krantsevich, C., Hahn, P.R., Hahn, S., Dasarathy, G., Turaga, P., Liss, J., 2021. Digital medicine and the curse of dimensionality. NPJ digital medicine 4 (1), 1–8.

Byrge, L., Kennedy, D.P., 2018. High-accuracy individual identification using a "thin slice" of the functional connectome. Netw. Neurosci. doi:10.1162/netn_a_00068.

Costa-Klein, P., Ettinger, U., Schirner, M., Ritter, P., Falkai, P., Koutsouleris, N., Kambeitz, J., 2020. Investigating the Effect of the Neuregulin-1 Genotype on brain function using brain network simulations. Biol. Psychiatry. doi:10.1016/j.biopsych.2020.02.118.

Cranmer, K., Brehmer, J., Louppe, G., 2020. The frontier of simulation-based inference. Proc. Natl. Acad. Sci. 117, 30055–30062.

Das, S., Glatard, T., MacIntyre, L.C., Madjar, C., Rogers, C., Rousseau, M.E., Rioux, P., MacFarlane, D., Mohades, Z., Gnanasekaran, R., Makowski, C., Kostopoulos, P., Adalat, R., Khalili-Mahani, N., Niso, G., Moreau, J.T., Evans, A.C., 2016. The MNI data-sharing and processing ecosystem. Neuroimage doi:10.1016/j.neuroimage.2015.08.076.

Deco, G., Ponce-Alvarez, A., Hagmann, P., Romani, G.L., Mantini, D., Corbetta, M., 2014. How local excitation–inhibition ratio impacts the whole brain dynamics. J. Neurosci. 34, 7886–7898.

Eppler, J.M., Helias, M., Muller, E., Diesmann, M., Gewaltig, M.O., 2009. PyNEST: a convenient interface to the NEST simulator. Front. Neuroinform. doi:10.3389/neuro.11.012.2008.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods. doi:10.1038/s41592-018-0235-4.

Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S., 2012. Brain templates and atlases. Neuroimage doi:10.1016/j.neuroimage.2012.01.024.

Gewaltig, M.-O., Diesmann, M., 2007. NEST (NEural Simulation Tool). Scholarpedia doi:10.4249/scholarpedia.1430.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage doi:10.1016/j.neuroimage.2013.04.127.

Guevara, M., Román, C., Houenou, J., Duclap, D., Poupon, C., Mangin, J.F., Guevara, P., 2017. Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. Neuroimage 147, 703–725.

Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y., 2013. Identifying personal genomes by surname inference. Science (80–) doi:10.1126/science.1229566.

Halchenko, Y.O., Meyer, K., Poldrack, B., Solanky, D.S., Wagner, A.S., Gors, J., MacFarlane, D., Pustina, D., Sochat, V., Ghosh, S.S., 2021. DataLad: distributed system for joint management of code, data, and their relationship. J. Open Source Softw. 6, 3262.

Hashemi, M., Vattikonda, A.N., Sip, V., Guye, M., Bartolomei, F., Woodman, M.M., Jirsa, V.K., 2020. The bayesian virtual epileptic patient: a probabilistic framework designed to infer the spatial map of epileptogenicity in a personalized large-scale brain model of epilepsy spread. Neuroimage doi:10.1016/j.neuroimage.2020.116839.

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med 2, e124.

Jeurissen, B., Descoteaux, M., Mori, S., Leemans, A., 2019. Diffusion MRI fiber tractography of the brain. NMR Biomed 32, e3785. doi:10.1002/nbm.3785.

Jirsa, V.K., Proix, T., Perdikis, D., Woodman, M.M., Wang, H., Bernard, C., Bénar, C., Chauvel, P., Bartolomei, F., Guye, M., Gonzalez-Martinez, J., 2017. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. Neuroimage doi:10.1016/j.neuroimage.2016.04.049.

Jirsa, V.K., Stacey, W.C., Quilichini, P.P., Ivanov, A.I., Bernard, C., 2014. On the nature of seizure dynamics. Brain 137, 2210–2230. doi:10.1093/brain/awu133.

Matzke, H., Schirner, M., Vollbrecht, D., Rothmeier, S., Llarena, A., Rojas, R., Triebkorn, P., Domide, L., Mersmann, J., Solodkin, A., 2015. TVB-edupack—an interactive learning and scripting platform for The Virtual Brain. Front. Neuroinform 9.

Meier, J.M., Perdikis, D., Blickensdörfer, A., Stefanovski, L., Liu, Q., Maith, O.Ü., Dinkelbach, H., Baladron, J., Hamker, F.H., Ritter, P., 2021. Virtual deep brain stimulation: multiscale co-simulation of a spiking basal ganglia model and a whole-brain mean-field model with The Virtual Brain. bioRxiv doi:10.1101/2021.05.05.442704.

Melozzi, F., Bergmann, E., Harris, J.A., Kahn, I., Jirsa, V.K., Bernard, C., 2019. Individual structural features constrain the mouse functional connectome. Proc. Natl. Acad. Sci. U. S. A. 116, 26961–26969. doi:10.1073/pnas.1906694116.

Melozzi, F., Woodman, M.M., Jirsa, V.K., Bernard, C., 2017. The virtual mouse brain: a computational neuroinformatics platform to study whole mouse brain dynamics. eNeuro doi:10.1523/ENEURO.0111-17.2017.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., Mortrud, M.T., Ouellette, B., Nguyen, T.N., Sorensen, S.A., Slaughterbeck, C.R., Wakeman, W., Li, Y., Feng, D., Ho, A., Nicholas, E., Hirokawa, K.E., Bohn, P., Joines, K.M., Peng, H., Hawrylycz, M.J., Phillips, J.W., Hohmann, J.G., Wohnoutka, P., Gerfen, C.R., Koch, C., Bernard, A., Dang, C., Jones, A.R., Zeng, H., 2014. A mesoscale connectome of the mouse brain. Nature 508, 207–214. doi:10.1038/nature13186.

Osen, K.K., Imad, J., Wennberg, A.E., Papp, E.A., Leergaard, T.B., 2019. Waxholm Space atlas of the rat brain auditory system: three-dimensional delineations based on structural and diffusion tensor magnetic resonance imaging. Neuroimage doi:10.1016/j.neuroimage.2019.05.016.

Palomero-Gallagher, N., Zilles, K., 2019. Cortical layers: cyto-, myelo-, receptor- and synaptic architecture in human cortical areas. Neuroimage doi:10.1016/j.neuroimage.2017.08.035.

Papp, E.A., Leergaard, T.B., Calabrese, E., Johnson, G.A., Bjaalie, J.G., 2014. Waxholm Space atlas of the sprague dawley rat brain. Neuroimage doi:10.1016/j.neuroimage.2014.04.001.

Proix, T., Spiegler, A., Schirner, M., Rothmeier, S., Ritter, P., Jirsa, V.K., 2016. How Do Parcellation Size and Short-Range Connectivity Affect Dynamics in Large-Scale Brain Network models?. Neuroimage.

Ritter, P., Schirner, M., McIntosh, A.R., Jirsa, V.K., 2013. The Virtual Brain integrates computational modeling and multimodal neuroimaging. Brain Connect 3, 121–145. doi:10.1089/brain.2012.0120.

Rocher, L., Hendrickx, J.M., de Montjoye, Y.A., 2019. Estimating the success of re-identifications in incomplete datasets using generative models. Nat. Commun. doi:10.1038/s41467-019-10933-3.

Sanz-Leon, P., Knock, S.A., Spiegler, A., Jirsa, V.K., 2015. Mathematical framework for large-scale brain network modeling in The Virtual Brain. Neuroimage 111, 385–430.

Sanz-Leon, P., Knock, S.A., Woodman, M.M., Domide, L., Mersmann, J., McIntosh, A.R., Jirsa, V.K., 2013. The Virtual Brain: a simulator of primate brain network dynamics. Front. Neuroinform 7.

Schirner, M., McIntosh, A.R., Jirsa, V.K., Deco, G., Ritter, P., 2018. Inferring multi-scale neural mechanisms with brain network modelling. Elife doi:10.7554/eLife.28927.

Schirner, M., Rothmeier, S., Jirsa, V.K., McIntosh, A.R., Ritter, P., 2015a. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. Neuroimage.

Schirner, M., Rothmeier, S., Jirsa, V.K., McIntosh, A.R., Ritter, P., 2015b. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. Neuroimage 117, 343–357.

Shen, K., Bezgin, G., Schirner, M., Ritter, P., Everling, S., McIntosh, A.R., 2019. A macaque connectome for large-scale network simulations in TheVirtualBrain. Sci. data. doi:10.1038/s41597-019-0129-z.

Smith, R., Connelly, A., 2019. MRtrix3_connectome: A BIDS Application for Quantitative Structural Connectome Construction. OHBM, p. W610.

Sotiropoulos, S.N., Zalesky, A., 2019. Building connectomes using diffusion MRI: why, how and but. NMR Biomed 32, e3752.

Stodden, V., McNutt, M., Bailey, D.H., Deelman, E., Gil, Y., Hanson, B., Heroux, M.A., Ioannidis, J.P.A., Taufer, M., 2016. Enhancing reproducibility for computational methods. Science (80) doi:10.1126/science.aah6168.

Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A., 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. Neuroimage doi:10.1016/j.neuroimage.2019.116137.

Trebaul, L., Deman, P., Tuyisenge, V., Jedynak, M., Hugues, E., Rudrauf, D., Bhattacharjee, M., Tadel, F., Chanteloup-Foret, B., Saubat, C., Reyes Mejia, G.C., Adam, C., Nica, A., Pail, M., Dubeau, F., Rheims, S., Trébuchon, A., Wang, H., Liu, S., Blauwblomme, T., Garcés, M., De Palma, L., Valentin, A., Metsähonkala, E.L., Petrescu, A.M., Landré, E., Szurhaj, W., Hirsch, E., Valton, L., Rocamora, R., Schulze-Bonhage, A., Mindruta, I., Francione, S., Maillard, L., Taussig, D., Kahane, P., David, O., 2018. Probabilistic functional tractography of the human cortex revisited. Neuroimage doi:10.1016/j.neuroimage.2018.07.039.

Van Essen, D.C., 2013. Cartography and connectomes. Neuron 80, 775–790.

Vella, M., Cannon, R.C., Crook, S., Davison, A.P., Ganapathy, G., Robinson, H.P.C., Angus Silver, R., Gleeson, P., 2014. libNeuroML and PyLEMS: using python to combine procedural and declarative modeling approaches in computational neuroscience. Front. Neuroinform. doi:10.3389/fninf.2014.00038.

Wang, X.-J., 2020. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. 21, 169–178.

Yeh, C.-H., Jones, D.K., Liang, X., Descoteaux, M., Connelly, A., 2021. Mapping structural connectivity using diffusion MRI: challenges and opportunities. J. Magn. Reson. Imaging 53, 1666–1682.

Zimmermann, J., Perry, A., Breakspear, M., Schirner, M., Sachdev, P., Wen, W., Kochan, N.A., Mapstone, M., Ritter, P., McIntosh, A.R., Solodkin, A., 2018. Differentiation of Alzheimer's disease based on local and global parameters in personalized virtual brain models. NeuroImage Clin doi:10.1016/j.nicl.2018.04.017.