Future Neuroscience and the Human Brain Project

Building a Neuroscience Community: community modelling and data repositories

Fondation Brocher, Hermance

11 – 13 June 2015

WORKSHOP REPORT

Table of Contents

1. Introduction

'Future Neuroscience and the Human Brain Project: Building a Neuroscience Community: community modelling and data repositories' was the title of a workshop held at the Foundation Brocher, Hermance, Switzerland. The workshop was organised by the Human Brain Project Foresight Lab at King's College London (Prof Nikolas Rose, Christine Aicardi, Michael Reinsborough, Paola Bello) in collaboration with Andrew Davison (UNIC Lab of CNRS, HBP) and Jeff Muller (Blue Brain Project, EPFL).

The aim of the workshop was to explore explore possibilities, issues and practicalities in **collaborative neuroscience with a specific focus on collaboration between diverse brain modelling communities and approaches.** At present there is considerable fragmentation of models and approaches to model building and simulation in neuroscience, and in particular in relation to modelling brains. The work of the Human Brain Project on modelling and simulation is one among a number of approaches, but it has a mandate from the European Commission to help to build a collaborative neuroscience community. In order to achieve this, the vision of the HBP must be integrated with the wider neuroscience community and hence it must integrate its ambition to provide a platform for models and simulation with existing initiatives in neuroscience. The challenge is to develop a mechanism to bring those various endeavours together, without implying that a particular platform should dominate: the task is to find whether it is possible to move from a competition to a consensus based decision-making strategy in this crucial area of neuroscience. Evidence from similar endeavors suggests that this is best done from the grassroots, starting from some practical steps to achieve minimal consensus and building out from that.

The purpose of this workshop was thus

- to develop a practical strategy for community building around brain models research;
- to build a roadmap for integrating the tools that this requires;
- to clarify the role of the HBP and the platforms that it is developing such that they can provide the best possible and most appropriate services for the neuroscience community.

The workshop consisted of short talks from selected participants addressing these issues from their own perspective and experience, together with workshop discussions on a number of key "collaboration challenges" with the aim of making plans as to how these might be achieved. The workshop was held under Chatham House Rule in order to facilitate open and productive discussion. [1]

Background

Nikolas Rose, of the Department of Social Science, Health and Medicine, King's College London, opened the workshop, followed by some opening remarks from co-organiser Andrew Davison.

Rose introduced the work of the Foresight Lab based at King's College London, part of the HBP Sub Project 12 (Ethics and Society). The aim of the Foresight Lab is to explore the future potential social and ethical issues raised by work of HBP, with a particular focus on medicine, neuroscience and robotics. The Foresight Lab is currently focussing on neuroscience and in particular model building and simulation of neural processes.

This workshop was originally planned to focus on the conceptual and epistemological questions raised by different approaches to model building in neuroscience, exploring their characteristics (top down, bottom up) and the different relations between data and models, experimenters and modelers. However, a different approach emerged during a workshop[2] held in London in April 2015. The HBP Hippocamp workshop brought together modellers and experimentalists from within and outside of the Human Brain Project. Co-organised by scientists in the HBPs 'Brain Simulation' (SP6) and 'Mouse Brain Data' (SP1) sub-projects, its aims were to engage the larger community of experimentalists and modellers working on hippocampus, highlighting existing modelling efforts and strategic datasets. In particular, it intended to define and bootstrap an inclusive community-driven model and data-integration process to provide well documented, open pre-competitive reference models of hippocampus.[3]

The Hippocamp workshop represented, in part, a way to respond to the European Commission's first technical review of the HBP.[4] Among the main recommendations outlined in the review by the European Commission were the following:

- Closer integration of the Data and Theory subprojects with the development of the ICT platforms. Specifically, this recommendation calls for a greater integration of simulation and modelling within HBP, and links both with data providing sub-projects (SP1 and SP2) and the cognitive and theoretical neuroscience subprojects (SP3 and SP4)[5]
- Building an HBP user community. The HBP consortium has to develop strategic plan which allows for multiple approaches to brain simulation, and to collaborate with international scientific communities.
- Integrated infrastructure development. Multi-level models of different brain regions require a digital data repository where experimentalists can upload data from experiments, and modellers can access experimental data. This represents a challenge in the effort to build a community: experimentalists collect data in multiple ways, and there are no standards – or incentives to create standards – for the annotation and storage procedures. The HBP could play a key role here, offering an infrastructure strategy to aggregate knowledge around brain atlases, even if they are not using the simulation tools. [6]

Challenges

Given the search for a community driven 'open science' model building approach, what role can the HBP play?

- Platforms: different communities could work on one model via the HBP simulation platform, or on their own separate models using the same data sets made available via the HBP simulation platform. Still, the simulation platform could allow for both processes to develop alongside one another.
- Collaboration. While a form of collaboration is necessary, for both modellers and experimentalists, it cannot be realised without the joint effort of all those involved, and it is not likely to happen without incentives. Could a data sharing facility, supported by the HBP, provide such an incentive? If so, what are the implications for the ways in which experimentalists and modellers collaborate?
- Technical issues. Technical issues emerged regarding data integrity, compatibility, storage, standards references. Can models of multiscales work together, and how? Who is entitled to

define standards and how? In the era of big data, metrology issues are undergoing significant changes.

- Diverse epistemologies. There are different 'epistemological cultures' among modellers and experimentalists – that is to say different conceptions as to what counts as knowledge, different beliefs as to what is known or unknown and what existing or new knowledge is relevant or irrelevant to the task at hand, different notions of what can count as evidence and as explanation - as well as different approaches to both data and biology – e.g. brain anatomy. The issue of 'interdisciplinary collaboration' - required to address salient problems - has been much discussed in studies of contemporary science and requires both recognition of such different knowledge cultures, and the fostering of bridging or translational procedures, practices and individuals.
- Ownership and credit. How to manage ownership, credit, responsibility? How can one give credit to those who curate data? What kind of relation is set between users of platforms and developers of platforms?

Nikolas concluded by hoping that this workshop would provide an opportunity to grapple with these issues: to help understand more clearly the issues at stake, to see the possible paths forward, and to learn from previous community building in science. The last part of the workshop, "The way forward", focuses on the tentative directions and roadmaps going discussed. And, he concluded, we hoped that this workshop could itself provide a model itself of collaborative open science.

Andrew Davison then gave an introduction from his own perspective on the rationale for the workshop and its central aims. He focussed on the need for collaboration, and why it was essential for neuroscience research in the near future. The main issue for him, at this point in time, is that it is impossible even for a large research group to envision developing, simulating and validating an accurate and complete model of an entire mammalian brain. Getting there will require the collaboration of individual scientists, and both small and large research groups. It will require community models, community databases and community tools. Over the past few years, there have been real efforts in these areas, but not so much in the development of community infrastructure projects. GitHub is one, for instance, but not tailored to the specific needs of the wider neuroscientific research. The Open Source Brain is another. A major problem of such infrastructures is their need for ample computer power and resources, as well as long-term support. Davison's position is that the HBP has a role to play in this respect. The challenge is how to get tool developers, infrastructure architects, diverse schools of modellers, and experimentalists from different traditions, to work together: this is the overall challenge that our wide ranging discussions tried to address during the workshop.

2. Summary of Discussions

The workshop took the form of five thematic plenary sessions over the first day and a half, each featuring either three or four short presentations followed by ample time for questions and answers. The main issues and challenges raised through these discussions then fed into the self-organization of four working groups, which reported back towards the end of the workshop. These reports formed the basis of the initial sketch of a roadmap with which the workshop concluded.

2.1 From experiments to data repositories (I and II)

The first afternoon was taken up by two plenary sessions focused on data, from production to curation.

The first presenter was Leslie Smith, Professor in the Department of Computing Science and Mathematics at the University of Stirling (UK) and also coordinator of the UK Neuroinformatics Node in the International Neuroinformatics Coordinating Facility (INCF). He talked of his experience with the project CARMEN (Code Analysis, Repository and Modelling for E-Neuroscience), an online collaborative Virtual Laboratory for neurophysiology. The concept of the project was to develop a portal-based system, where researchers would use the portal to store and share electrophysiological datasets, as well as their analysis and visualisation techniques. It involved 11 UK universities, and was funded from 2006 until 2015, initially by the Engineering and Physical Sciences Research Council (EPSRC) and then by the Biotechnology and Biological Sciences Research Council. Although the project has now run out of funding, the portal remains active. Yet it is at risk, because it is getting to its limits. It has almost reached full capacity and the computers on which it runs are becoming obsolete: it would require a lot more work to expand it and go on working. Leslie showed that the interest of CARMEN is multiple from the perspective of the Human Brain Project. First, it is a community e-infrastructure project, which over its ten year experience has faced many of the technological and sociological issues – although on a smaller scale – that the HBP is facing today: it has lessons to teach. Beyond, Leslie observed that good infrastructure projects require a long lifetime: sustainability is important. A partnership between CARMEN and the HBP could mean breathing new life into CARMEN while integrating valuable datasets into the HBP. More generally, Leslie expressed the wish that the HBP worked more closely with INCF.

Two members of the HBP Neuroinformatics team spoke next, both from Ecole Polytechnique Fédérale de Lausanne (Switzerland): Catherine Zwahlen, who is a data scientist, and Martin Telefont, who is part of the Blue Brain Project and leads the HBP data integration efforts.

Catherine gave an overview of the Knowledge Graph – the conceptual design of the knowledge base – that the HBP Neuroinformatics team would like to achieve, taking on board the "zoo of data out there" (of many different types, at many different resolution scales and timescales, produced through many different experimental techniques) that the HBP wishes to integrate. The core challenge is how to create the metadata in order to make the data discoverable, accessible, usable, publishable and citable. Six broad categories of metadata have been retained: observations and models; specimen;

contributors; location; methods and protocols; and disease. The Neuroinformatics team has chosen provenance, a form of structured metadata designed to record the origin and source of information, which is useful for evaluating whether data can be trusted, for integrating it with other heterogeneous data sources, and for crediting attribution to the data creators throughout the data life cycle. They are using PROV, the standard provenance model of the World Wide Web Consortium (W3C). Catherine showed how it could be used to develop a data model for implementing the Knowledge Graph, and how practically the data model could help manage the data life cycle.

Martin's talk complemented Catherine's theoretical presentation, by illustrating through three practical examples of how the data integration proposed by the HBP Neuroinformatics platform could work and of the benefits it could bring to laboratories. He made a few observations as preamble: data producers and data consumers are often the same people; the HBP involves dozens of different labs with different research traditions and practices; a consequence is the large variety of data that needs integrating – "we have one of everything but not much of anything." The first example was single cell morphology classification. Martin showed that two labs in the HBP Consortium, even when they were doing supposedly comparable work, could follow very different workflows leading to very different kinds of results, and how these differences became apparent and thus traceable through the Knowledge Graph. The second example was about cell distribution and volume estimations, and the Knowledge Graph was used to compare two labs following similar workflows and ending up with similar outputs – in this case aligned image stacks. The last example showed how the data model could help bringing together experimental work, analysis and modelling/simulation for end-to-end reconstruction and getting otherwise isolated research groups from different sub-projects to collaborate. Broadening his perspective to encompass "the rest of the planet", Martin explained that working with the three broad categories of actors – individual labs, data aggregators (e.g. the Open Connectome Project, NeuroMorpho) and large institutional initiatives (e.g. BRAIN, the Allen Institute for Brain Science) would require that the HBP Neuroinformatics platform be flexible and be prepared to play at times the role of 'harvester', by meta-indexing these different levels of initiatives, and at times the more ambitious role of 'curator', by building an archive of the best available data of a specific kind. Highlighting the social role played by Neuroinformatics, he told us in conclusion that "these days I spend about 80% of my time talking to people."

The last presenter in the first session was Shreejoy Tripathy, post-doctoral researcher in the Centre for High-Throughput biology at the University of British Columbia (Canada). He talked about the NeuroElectro website and framework that he has been developing and maintaining. The goal of NeuroElectro is to organize existing data on cellular neurophysiology, by extracting information about the electrophysiological properties of diverse neuron types from the existing literature and curate it into a centralized database. It uses semi-automated property extraction (text-mined data must be checked by human curators) for mining electrophysiological values and associated experimental metadata, on thousands of downloaded full-text articles from neuroscience journals. The neuron types used are those on the lists of mammalian neuron types generated in NeuroLex, the Neuroscience Lexicon, developed by Gordon Shepherd and colleagues. Shreejoy acknowledged that NeuroElectro tends to lump neuron types together: "I tend to be a 'lumper'. Other people in this room are 'splitter'." He observed that assessing the quality of data remained very much rule of thumb. Electrophysiological data – and nomenclature – are notoriously heterogeneous: accounting for differences in experimental conditions is paramount, to make corrections for methodology differences. The HBP could really contribute to the desperate need for standards. Further, he is starting to get requests to upload 'raw data' in NeuroElectro. He does not have the resources to do that, yet he thinks that raw data is really important. Actually, he would love it if he could stop mining text articles to get raw data, because "it is like turning a hamburger into a cow." In his view, the 'rawness' of data is not an issue, "the curators know what makes good data good and bad data bad."

Thomas Wachtler, Director of the INCF G-Node in the Department of Biology at Ludwig-Maximilians-Universität (Germany), the German Neuroinformatics Node, opened the second session with a talk focused on "Tools and methods for efficient data management in neuroscience", based on the experience of the INCF and G-Node. The INCF was established in 2005, to develop an international collaborative neuroinformatics infrastructure – data sharing and databases, analytical tools and computational models. It counts currently 17 member countries, each hosting a national node. The national nodes can be single institutions or networks, and play an active role in formulating and implementing INCF programs, which are long-term strategic undertakings to address issues of high importance to the neuroscience community. The 4 programs are digital brain atlasing, multi-scale modelling, ontologies of neural structures, and standards for data sharing. G-Node focuses on neuroinformatics solutions for electrophysiology: data conversion tools, methods for data and metadata management, data sharing platform, custom solutions for collaborative data exchange, hosting services, and teaching and training. Why put such emphasis on data management? The increase in complexity and volume of data poses a challenge for data organization. Collaborative efforts, re-use of data and reproducibility are hampered by the effort it takes to access and understand the data. The aim is to reduce this effort. Thomas made clear that the first step in sharing data, before sharing with collaborators and eventually the rest of the world, is to share data with oneself. Making sure that all useful information about a data set is and stays available – managing all the metadata – can be facilitated by an integrated organization of data and metadata. The main challenge in neurophysiology, as pointed earlier by Shreejoy, is the heterogeneity and complexity of data, and the general lack of standards (for data access, data annotation, stimulus description, etc). G-Node has been working hard on standardization and interoperability. They are involved in the development of NEO, a data model for neurophysiology that supports input from a wide range of neurophysiology file formats. Also, they have developed odML, a flexible and extensible metadata format, with a series of tools (libraries, editor, apps), including libraries to facilitate automated metadata collection in the laboratory. To facilitate data analysis and sharing, they are in the process of developing NIX, a file-based integrated organization of data and metadata, which combines a general data model derived from NEO and full odML metadata integration.

The next two speakers were again part of the HBP Neuroinformatics team: Sonja Grün, of Jülich Research Center (Germany), who leads the work package in charge of building tools for functional data analysis, and Paul Tiesinga, of Stichting Katholieke Universiteit (Netherlands), who heads the work package in charge of predictive neuroinformatics.

Sonja shared her experience of ensuring reproducibility – or failing to do so – when collaborating with experimentalists. She is a computational neuroscientist whose competence is in data analysis, and she is interested in the correlation between behaviour (function) and the dynamics of neural networks. It is work done specifically with monkeys, based on complex experiments that involve long training of the experimental subjects, and she has developed an increasingly close collaboration with experimentalists. She pointed that over the years, there has been a strong evolution of the activities associated to 'preparing' the work – to making it ready for machine learning and data mining – and this has resulted in an ever larger part of her time going into workflow definition, project management and software development. Working closely with experimentalists, she has come to realise that reproducibility of results was a major issue: data workflows are very involved, metadata are scattered and in different formats, data are complex and at multiple scales. A step towards resolving the issue would be to achieve complete provenance tracking (the structured metadata model of which Catherine had spoken earlier) across the 'experiment to results' workflow. The obstacles are the multiplicity of researchers and of manual steps involved, the scattering of documentation over different files and formats, the diversity of analyses software that are being used. Sonja's group has been working towards this goal, identifying required components for provenance tracking. One component is to collect complete metadata, provided through metadata annotation by G-Node's odML. Another component is the development of a common analysis tool: Elephant, which stands for Electrophysiology Analysis Toolkit – an open-source, community-centred library for the analysis of electrophysiological data in the Python programming language. Elephant incorporates the data object model NEO for flexibility. Thanks to these tools, the HBP Collaboratory can provide a provenance tracking framework. Among the problems to be solved, some are deeply cultural. For instance, not all experimentalists agree that there is a reproducibility problem. The experimentalists do not (yet) see the relevance of odML. They use MATLAB, and are not ready to use Python, the programming language privileged by analysts and modellers.

Paul presented the work of predictive neuroinformatics, which aims to help the HBP modelling efforts go from structure to function. The HBP model needs complete specification of all parameters, when only a fraction of the data is available: the idea is to predict the rest, building on the revival of anatomical studies that provide detailed structural information of neural circuits at cellular resolution. Different techniques are used for data prediction. One is data fusion, which consists in combining data from different modalities to improve reliability. Another targets missing data, and consists in inferring data based on existing data. The last consists in generating data according to general principles. The goal of predictive neuroinformatics in the HBP is to reconstruct long-range projections of neurons at cell-to-cell resolution, with their branching information. There are different approaches to reconstruction. A 'brain builder' approach would aim at 'growing' a brain using developmental principles. This is not the kind of approach that the HBP has adopted. The HBP has been using a method based on a simple geometric principle, Peter's rule, which can predict whether there is a potential synapse between two neurons: it generates the morphology of neurons first then constructs connectivity based on the morphology of the postsynaptic dendrite and the presynaptic axon. Another approach is now also used, which reverses the geometric principle: synapses are placed first then the morphology is regenerated. Paul went on to explain, in practice, how these methods were working. In conclusion, he argued that the kind of predictive neuroinformatics that is pioneered by the HBP results from profound changes in computational techniques – modelling as well as analytical tools – and molecular techniques over the past 15 years, which have revolutionized neuroscience. These changes are not reflected adequately in neuroscience training courses, and in particular, neuroscientists need more training in mathematics and informatics, if we are to overcome the key challenge of integration.

2.2 Lessons from elsewhere

The first plenary session of Day 2 was dedicated to perspectives coming from the humanities and social science.

Christine Aicardi, the first speaker, belonged to the Foresight Laboratory at King's College London (UK), which is part of the HBP Ethics and Society sub-project. Her talk was a survey of social science research identifying well-known challenges, good practices and issues in interdisciplinary collaborations – broadly understood as the combination of several approaches originating in different research traditions to address a question. She contrasted interdisciplinarity as it is engineered top-down by policy-makers for utilitarian purposes in knowledge economies, with the experience and practice of interdisciplinarity on the ground. These are two distinct realities, and growing grassroots collaborations within the constraints of a big science consortium like the HBP is a major challenge. She highlighted good practices that have been linked to positive experiences of interdisciplinary collaborations: developing the capacity to diversify perspectives in order to bridge across research fields; developing interlanguages between collaborating groups, which starts with specifying precisely what everyone means by the terms used; accepting that competition, conflicts and misunderstandings are part and parcel of interdisciplinary work – hard issues must be confronted, not dismissed on the grounds that they will disappear in time because they do not; and most importantly, building up trust between participants. Finally, she pointed at well-identified issues. First, bridging across specialized fields to gain breadth of perspective is time consuming and is often equated with superficiality and lack of expertise, which makes it a difficult career choice. Then there is an issue of temporality: interdisciplinary work requires long temporal cycles, incompatible with typical cycles of evaluation and assessment. Long tails of research, fertile periods of lesser productivity, non-standards outputs – all characteristic of interdisciplinary work – are not taken into account. Finally, there is the issue of evaluation and reward structures: interdisciplinary collaboration lacks recognition and material rewards, for lack of adequate evaluation criteria.

The next two speakers brought perspectives from outside the HBP. Niccolò Tempini has a dual background in philosophy and information systems. Based at EGENIS, the Centre for the Study of Life Sciences at the University of Exeter (UK), he is working on *The epistemology of Data-intensive Science*, an ERC-funded project. Dennis-Kenji Kipker, from the Institute for Information, Health and Medical Law in Bremen (Germany), is a member of the European Academy for Freedom of Information and Data Protection and of the German Association for Law and Informatics.

The starting point of Niccolò's presentation was that the HBP is going to develop information infrastructures to share across a very interdisciplinary community. With the design of such information infrastructures in mind, he reviewed some of their general characteristics and issues, based on extensive empirical research across the scientific and corporate domains by researchers in information systems, philosophy of science and sociology (see www.datastudies.eu). Information is context dependent: it is about what data and metadata are able to tell someone about something, and when experimental data are shared, especially in a large interdisciplinary environment, there are in addition different degrees of obstruction from the original phenomena. Findability, workflows, and the decontextualization/recontextualization of data (how data travel across contexts) are important design concepts. The handling of data is paramount – collection, curation, analysis, enhancement – and some argue that to curate and give particular metadata to datasets is a theoretical statement. Yet this is work that still struggles to be seen as part of scientific research. Thinking how this particular interdisciplinary community is going to make a new kind of science is maybe also about rethinking what counts as research work. Another important aspect to consider is that technology is a transformative force. ICT projects hardly ever achieve their planned goals, but they often achieve transformations that were unintended and may be undesirable. In the present case, the planned information infrastructures will undoubtedly change research practices. It is important to try and anticipate which part of everyone's work is going to be transformed. Similarly, technology can be powerful in making previously independent people interdependent. It is unlikely that one can create a new community without degrees of interdependencies and some could be seen as undesirable. ICT is also particularly powerful in giving people control over others' work, sometimes unintended. There have been cases where information infrastructures have failed because they afforded management control over the work of others in ways that were unplanned and unwelcome, igniting resistance among their users. In large interdisciplinary and distributed networks such as the HBP, this requires serious thinking. Niccolò also addressed questions of standards, temporality, and critical mass in relation to the design and management of information infrastructures. He concluded by proposing that a lot can be learned about information infrastructures by looking at lessons from both the scientific and the corporate worlds, and that, for instance, tools such as the 5 Ps mnemonic (problem, people, project, process, product) may seem to oversimplify, but can help avoid misleading assumptions, as each of the Ps brings attention to areas where potential problems can arise.

Dennis spoke of the processing of human personal data for research purposes, in the context of current and future regulations in the European data protection law. He explained that the foundational data protection articles in the Charter of Fundamental Rights of the European Union are concretized through two important regulations: the current Data Protection Directive 95/46/EC (DPD), which has been running since 1995, and the future General Data Protection Regulation (GDPR), which is still in draft. There are plans for its adoption late 2015, and it is expected to be applied from 2017. They are both completely general regulations and thus implicitly regulate the scientific use of data in the EU. The current DPD has to be implemented into national laws to gain validity, with national duty of compliance. The future GDPR is a regulation of the EU with immediate validity. National implementations are not needed any more, which simplifies things considerably as all EU countries will have the same data protection provisions. The DPD and the GDPR can be analysed together because they are similar in many ways, with only punctual dissimilarities. The starting point of every use of personal data in the EU, including scientific and health data, is the general prohibition principle: it is not possible to use personal data unless there are exemptions to do so. These exemptions can be granted under special legal exceptions, found in both the DPD and the GDPR, which are applicable to the processing of personal data in research contexts: consent, informed and specific; public interest (a case-by-case review is necessary, with application of the 'balancing test': legitimate public interest against individual fundamental rights and freedom interests). Still, the law grants certain privileges to scientific research that facilitate the work with personal data when legally permitted: (1) Controllers – natural or legal person who determine the purposes, conditions and means of the processing of personal data – have the right to store data for a longer period than in other contexts. (2) The principle of legitimate purpose is relaxed, to make the scientific processing of data that was collected in other contexts possible. (3) Controllers' obligation of information towards individuals whose data is concerned is relaxed. (4) Individual rights to access information concerning the processing of their personal data are restricted. (5) Researchers will not have to comply entirely with the new right to be forgotten and to erasure.

2.3 From data to model building and validation

This thematic session focused on the process of moving from data to model building and then validating models by testing them against empirical data.. Speakers from the Blue Brain Project, the Max Planck Institute in Tübingen, the HBP Neurorobotics Platform, and the Virtual Brain project based in Marseilles spoke successively.

Srikanth Ramaswamy presented data from more than a decade of work at the Blue Brain Project (BBP). The Blue Brain Project focuses on the reconstruction and simulation of a cortical microcircuit in a somatosensory area of the juvenile rat brain. The workflow is thus: they characterise and gather information on the morphological diversity of neurons in this specific part of the brain, then identify the boundaries of the microcircuit and the depth of each cortical layer. In the computer model this microcircuit volume is then populated with neurons. Within this volume the presynaptic axons and postsynaptic dendrites form several million appositions or touches. The project then models the various electrical types (which are cross-matched with morphologies to constitute morphoelectrical types) and synaptic types. Because so little of this vast amount of information is definitively given by experiment, the project needs to develop ways to project and reconstruct this data from estimation and calculation to fill the gaps in knowledge. These projections are then tested against whatever experimental data is available in an attempt to validate (or invalidate) the model. When complete the populated microcircuit forms a virtual tissue volume suitable for in silico experimentation. This virtual tissue volume consists of neuronal anatomy, neuronal physiology, synaptic anatomy and synaptic physiology all of which are reconstructed. While the beginning of Srikanth's presentation outlined the overall programme (work flow) of the Blue Brain Project, the bulk of the presentation was a detailed discussion of various methods for reconstructing the missing information (and the validation checks) which are necessary to populate and fill out the model. The final part of the presentation, which had to be curtailed for reasons of time, consisted of some brief remarks on simulations now being done to replicate in vivo experiments with known results to see how close the model validates to experimental data when run as a simulation.

Marcel Oberländer began by describing his presentation as an "alien opinion" amongst those in the room, but not so much so amongst the much broader neuroscience community (of which computational neuroscientists and modellers form only about 10%). This was because he felt the key problem to be solved for contemporary neuroscience lay not in the *sharing of data and analysing what is already out there*, but in *generating better new data that might remain consistent across* *multiple scales or levels of neuroscientific inquiry*, such that it might be combined, and analysis could then be able to tell us something new. He explained that at his lab they work almost entirely with in vivo data. Because this work was based on data collected from a living animal much of the inventory of cell types and classification structures from elsewhere was not useful, because it could only be acquired from non-living tissue samples. His lab, the Center for Integrative Neuroscience at the Max Planck Institute in Tübingen, works to generate a brain wide model of the rodent whisker system to help understand how the rodent undertakes a decision making process. Instead of looking at dissection data and trying to identify everything one can know about the cortical system and all possible functions, they attempt to understand how one function in the brain works and to describe this at all scales relevant to the function. Primarily they look at when a rat detects (by whisker) a gap across its path in a dark maze system, and seek to understand the neural processes involved in its decision as to whether to cross the gap. The challenge is to have a cross scale interpretation of decision making. How does the interplay of levels work to encode a sensory percept to the animal that is actually of behavioural relevance? Once that is understood further research can focus on how the percept implemented in the brain can trigger a decision making process. At Marcel's lab they begin by identifying a well-controlled perceptual decision task, and then identify the brain wide and task related neural circuitry that is involved in processing the information for this percept (the so called anatomical connectome or quantitative wiring diagram). This information is complemented with functional measurements in the living animal to examine how different stimuli are represented in the animal for this task. This is called a stimulus specific function connectome. Which of all paths that may be available for testing information are actually active for a certain stimulus? When are they active and so forth? The perceptual decision task they study is whether a rat in a darkened maze will jump over a gap if it can or cannot sense with a whisker the other side of the gap. All whiskers are trimmed away except one. If the whisker can reach and touch the other side of the gap the animal will reliably decide to cross the platform. However, if the whisker cannot, the animal will not cross. The animals actions are recorded in parallel with electrophysiological recordings from individual or pairs or a population of neurons. Afterwards the in vivo recorded neurons are labelled and the brain of the animal is sliced and analysed to learn the specific morphology of the recorded cells. Their group has focused over the years on developing better methods for recording, staining, labelling, imaging and reconstructing because the work is quite tedious. They attempt to collect data in a very consistent manner such that data across animals and levels might be combined for analysis. They have at the same time developed an average anatomical and functional model of the rat barrel cortex with synaptic connectivity and synapse distribution measured statistically. With this model they perform predictive simulation experiments and compare the results to their in vivo data. Their geometry of the barrel cortex connectome (determined statistically) compares well to recent experimental measurements. They were at this stage able to ask how certain morphological cell types respond to a particular stimulus in vivo and show particular cell types at particular cortical levels were more likely to be involved in certain types of information processing. They suggest this method as an example for future neuroscience. It can be used to map other functions in other species and gradually build knowledge. The lab also has released a neuroinformatics tool (and associated connectivity data) with which other experimenters can compare their own cell data to what the Max Plank lab model would predict of this cell data.

Unlike the previous presentations which were data driven microcircuit modelling (where we don't know microcircuit connectivity and other relevant information and therefore must reconstruct it.) the next two speakers, Mark–Oliver Gewaltig of the HBP Neurorobotics Platform, and Marmaduke Woodman of the Virtual Brain Project based in Marseilles, were both looking at modelling at much higher levels, in the first case by taking whole brain models and trying to connect them to the neuronal substrate level (so that cognitive scientists might test their high level theories on a suitable robotics platform), and in the second case by mapping brain imaging data to mass neuron models where the electrical relationship between regions of interest is modelled for individual patients. This helps plan clinical intervention in, for example, serious epilepsy.

Marc-Oliver began by suggesting that there was a certain link between trying to construct a whole brain model and trying to make the model do something. Thus he felt it was appropriate that part of his background was in robotics. The HBP Neurorobotics Platform is intended to allow behavioural and cognitive neuroscientists to replicate their (up until now untested) thought experiments in simulations run on this platform. To do a full simulation it is important to include a body. Whisker type experiments might look at whisking and part of the microcircuit in simulation but what is typically missing in these experiments is the mouse, the platform, and everything around it. During simulation one is not actually redoing the full gap experiment as performed in the lab. These simulations are only conceptualized versions of the original experiment where consciously or not we have removed many of the confounding factors. Instead, brain models for neurorobotics must look at synergies between brain and bodies. Often the hope by modellers is that if they are able to model the brain anatomically, bottom up, it will then (out of some magic) do something useful. Re-invoking a memorable phrase from the previous day Marc-Oliver pointed out that bottom up models are usually built statistically and therefore they are a bit like a hamburger compared to a steak. The constituent parts have been ground up and then clumped together again to roughly resemble the shape but important structural information has been lost, particularly the synaptic weights. One can only estimate this type of information up to a certain statistical average. This is where one has to reconstitute knowledge and the only way to do this is to bring in behaviour. Behaviour is not only the final outcome of simulation but it is an important constraint, explained Marc-Oliver. Although there are lots of existing cognitive models, there is (as of yet) no way to test these. But what if one could map these top down ideas to the neuronal substrate? If so then, it would be quite easy to test them. So the neurorobotics platform connects a body model (relatively easy to construct) to a whole brain model (of which there aren't many). Marc-Olivier described a workflow for the platform to rapidly turn data into models. There is a lot of data, many species. Therefore it is best to spit out models relatively quickly and when they fail (They will fail of course) one can iteratively use the failure to return and develop better models. Marc-Oliver then went on to explain some of the technical details of industrializing the workflow. Much of the HBP neurorobotics platform would take its data from the Allen Brain Institute dedicated to mapping the mouse brain. Some of the data needs to be corrected since the slices of mouse brain from which connection and other information have been taken have terrible alignment. The process of extrapolating data shares some similarity with other processes described earlier. First cell positions are taken, then numbers of excitatory and inhibitory cells, and the glia neuron distinction is verified, and here we stop (although in principle far more fine distinctions are possible). Mesoscale connectivity can be given by retrovirus introjection method (as per the work of the Allen Brain Institute) but microscale connectivity is more difficult. In principle a model like the Blue Brain Project or that presented by Marcel could be used until such time as a better model or real data comes along.

As the final speaker in this thematic session, Marmaduke Woodman described the Virtual Brain project. The Virtual Brain builds data driven models of electrical activity relationships between brain regions. Diffusion imaging of the brain and MRI scanning allow the team to reconstruct the cortical surface and major cartographic bundles in the brain. At each region of interest (ROIs) a dynamical mass neuron model is used to show the connection between the regions. 'Lead fields' translate neuronal activity to what is seen at the sensory level. The project then compares functional connectivity in the (model-based) simulation to empirically measured functional connectivity. Because of the constraints of visual imaging technology, the Virtual Brain cannot provide cellular models – instead it relies on mean field or neural mass models. For example a firing rate models or one of several different phenomenological models could be selected depending on what question you want to ask. For Marmaduke's team, the workflow begins with virtualizing a patient, for example an epilepsy patient. Many types of data are brought together sometimes with pre-processing or ad-hoc analysis. TI diffusion imaging, CT scans for areas, and some functional priors are brought together before obtaining sensory level data. When complete the patient model can be used to perform a simulation of the patient. This workflow involves ten different types of software. Unlike previously presented modelling projects, this process is not at all automated; it is all done by hand. Functional connectivity when measured by fMRI or EEG is non-stationary- it changes if you analyse it over time. 'Resting state' is a recent hot topic in the literature. Pathologies such as schizophrenia can be represented in such change patterns, so the Virtual Brain Project would like to be able to reproduce these features in their model (see their recently published paper). The relationship between regions cannot be reproduced in any model without taking account of the conductance velocity between regions. A bifurcation diagram shows a line of inflection when a patient goes from one state- for example a state of meditation, to another, for example paying attention to a noise. That diagram shows a unique relationship for each individual. Without the conductance values (for example if one tries to estimate them generically using statistical methods) you get a flat line. Marmaduke explained that the Virtual Brain Project often works with epileptic patients who are undergoing implantation and surgery, so whatever the criticism of their modelling approach one might have, if there is a real benefit to patients then that is valuable in and of itself. For example to model the propagation of a seizure in a patient they would begin by virtualizing the patient, making anatomy and connectome, then adding functional data to see how the seizure might generalizes from one hemisphere to the next. The question to be modelled is what information is necessary to prevent the seizure from propagating. What can be accomplished using the different possible clinical intervention strategies can also be modelled. Because of typical scale limitations in imaging technologies there are many limitations to this type of modelling. This is not an instrument that can comment on cell morphology. However, it often works for specific research and clinical questions. To manage complexity, Marmaduke explained, I throw more people at it. As a final comment, invoking the theme of privacy Marmaduke warned that he could directly recognize a person from their data, making it hard if not impossible to anonymize this type of personal data from the technician. "As more people get there connectome sequenced this will be as much a fingerprint as a thumbprint is now."

2.4 Pulling it all together: Community and platforms

For the final thematic session three speakers from Open Source Brain, OpenWorm, and the HBP Collaboratory respectively presented on the theme of community based platforms for cooperation within the neuroscience and modelling community.

Padraig Gleeson began his talk on the Open Brain Project by putting in context the normal development life cycle of coding and developing and updating models, often over a period of a couple of years where when the model shared with other labs and programing communities it might be fixed, updated and re-released by different parties. While work is happening to the model the repository always shows the most recent finished version of the model. To make computational neuroscience more scientific, models should allow reproducibility, accessibility, portability and transparency. There are many tools from the open source computing community developed over the past 30 years that can allow neuroscience an open community of modellers to track the history, provenance, changes and distribute this among users. The Open Source Brain Repository is a structured database of well tested spiking neuron and network models in standardised formats. It uses Github (a repositories sharing infrastructure for open source coding) to index and point to various neuroscience modelling projects so that updates are more carefully managed and shared. NeuroML is a standardized XML language for computational neuroscience to facilitate sharing and specification of neuronal morphologies, Ion channels, synapses, and three dimensional structure. pyNN is another python based coding language specifically for articulating models of neurons and running simulations. By using standardized representations with parameters, the parameters can be changed to test and update the model. The most recent project initiated in this manner is a collaborative and integrative modelling of the hippocampal area CA1 linked to a Github repository.

Stephen Larson spoke on behalf of the OpenWorm project, an open science project doing software engineering and developing a community with the common ends of mapping the C. Elegans neural system. The C. Elegans is microscopic worm studied by many scientific communities around the world as a model organism. It has recognizable behaviours in relation to its environment and other members of its species, a fully sequenced genome, 302 neurons, 95 muscle cells, and only a 1000 cells in total. The C. Elegans is the only organism to have its full connectome described. An important aspect of modelling the neural system of the C. Elegans is contextualizing the neural system within the body and the body within the environment. Just as researchers in robotics need to look at how the robot body relates to its environment so neuroscientists can benefit from thinking about this. For example, even though this microscopic worm is very simple we still don't know how the neurons make the worm crawl. Because motion may be coordinated with proprioceptive feedback it is therefore important to model the body along with the neural system. By placing all muscles in the context of the environment and then linking this to the neural framework a model can be built. By making gaps in knowledge clear to the user community these users can contribute to filling in these gaps. Model optimization can fill in gaps in knowledge. Software specific to this project like Geppetto and WormSim allow users to work with available data and simulate a C. Elegans attempting to fill in what gaps exist. The large community working on the OpenWorm project began through social networking, in this case a single message on twitter. All of the project meetings happen on line and then are posted on line allowing a large virtual community to participate actively in the project. For Stephen, this shows the enormous potential of open source, community based approaches to modelling, both building a committed modelling community and drawing in much expertise to work in a collaborative way on shared problems, in a way that would be impossible or very difficult for any small group of modellers on their own.

Next, Jeff Muller spoke about a very recent project beginning from within the HBP. The Collaboratory is intended to facilitate working together by scientists in the neuroscience and modelling community. Described as a 'dating site for scientific collaboration', the Collaboratory is an online platform that links users to various resources, data, and users to one another. The need for such a collectively organized hub is clear from the scale of the neuroscientific task. No single individual or lab group or even a well-funded large scale project such as the HBP can make noticeable progress on their own. A platform enabling users to extend its uses themselves, more like an ecosystem is what is needed. A hub for web based scientific collaboration, where sharing and collaboration happen around data, software and services. The HBP is developing half a dozen computing platform projects that are intended to support neuroscience with computing tools (for example supercomputing facilities for large scale simulations, a neurorobotics platform, etc.). Other neuroscience applications available from the greater research community can be linked into this. User extensibility is a key driver of such a networking platform. An early goal of the collaborator developers is to fix things that users need to make their extensions work. Users would be able to log-on to a common web interface with an HBP identity and access multiple application programs (based on the same standard as "Sign in with Google"). A provenance service would support technical data distribution and code attribution. In addition to applications the web interface would support various source-controlled python script with explicit package dependencies (known as Tasks). An overall software foundation (of necessary packages to run analysis applications) would be brought together through the Collaboratory, some developed or enhanced by the HBP and early contributing projects such as the Blue Brain Project but more and more coming from 3rd party sources as the user community extends the functionality of the Collaboratory (such as NEST, NEURON, Numpy, and other common tools for computer modelling in neuroscience). As the users begin to contribute validation tests for models, tools, and data then this will drive modellers to improve their models. The interaction around models between top-down and bottom-up modellers will eventually draw experimental validation data from the common HBP Neuroinformatics Platform.

3. Working groups

For the final part of the day, participants broke into four working groups on topics chosen as the most relevant after the above group discussions. These discussions included considerations of 1) how modelling could be relevant to clinical practice, 2) how modelling could support the bridging of different scales at which neuroscientific research is done (molecular, cellular, circuit-level, brain regions, etc.), 3) how community building within the modelling community could happen effectively, and 4) how to begin this practically by implementing collaboration across different platforms. The following morning when the groups reconvened there were four plenary presentations of the working group discussions.

3.1 Modelling relevant for clinical practice

In the first presentation on Saturday morning the working group to discuss clinical practice noted that they had given themselves a more precise title in the form of a question, "What kind of collaborative modelling might be required to model neuropsychiatric disorders in a way that was relevant to clinical practice?"

In the HBP the approach of the Medical Informatics Platform (MIP) is to federate clinical records and use data mining to recognize clinically relevant biomarkers. The intention was to model these in the simulation platform. There are several difficulties with this, for example the standardisation of imaging data, the mining of clinical records in diverse formats and so forth.. The working group did not feel they knew enough about the technicalities to assess the potential for this research although it was noted that a similar approach was being taken in USA by National Institute for Mental Health (NIMH) Research Domain Criteria project (RDoC). Some members of the group wondered whether or not data mined 'clusters' from a single time point would have either clinical relevance or give clues to aetiology or prognosis.

Obviously, clinical relevance means human data, so another challenge in obtaining such data for modelling is the concern to maintain privacy. For patients, they may be individually recognizable from scans – "my very own connectome", etc., even when image defacing occurs. But the overriding patient interest may not be in privacy, but in research into their condition. For persons participating as members of a research control group then in some jurisdictions, notably Germany, researchers' practice is to feed back to these 'controls' unanticipated findings of anomalies, even if benign, for example a benign tumour. These persons would then have to declare this information to insurance companies. There presently exists legislation to prevent discrimination based on genetic information. There was a discussion about whether it was desirable to press for extension of genetic non-discrimination legislation into the area of neurobiological data. If so who should do this? What role should the HBP take?

For collaborations between modellers and clinicians it is necessary to investigate what might be clinically useful. Clinicians might be induced to 'donate' their patient data to a repository (if de-identification and data security could be assured) that modellers could use to develop, in collaboration with tertiary medical services, models of disorders. These models could then be interrogated by clinicians in secondary or even primary care to directly benefit their particular patients, e.g. "Here is the fMRI and MEG data on Mrs X. Please run her on your modeller and tell me what her diagnosis is, and under what treatment might she be expected to show the best results… "

Modellers need to be able to specify what data on patients is required and in what format. This raises the problems of standardisation and how to ensure standardisation. Would this be something that could be done at European level – models need standards – or was there another level at which to do this? To build a model of a complex neuropsychiatric disorder one must take into account the considerable variation in phenotypic presentation as well as differences in brain structure and connectivity (structural and functional) between individuals. Perhaps a different, more developmental approach is needed. There was discussion of how a developmental approach could make use of life course data, perhaps from cohort studies or long term data follow ups of children, as already exist in some clinics. During this discussion one novel collaboration on this theme was developed between two participants in the discussion group, a brain modeller and a children's clinician using data from the clinic.

3.2 Bridging Scales

The 'bridging scales' working group reported that they had had a very lively discussion. However, in contrast to the other working groups, they had not reached much agreement.

However they did that *every level matters.* Many in the neuroscience community work at one scale of resolution in the brain so of course focus on that scale and are inclined to think other scales are less important. For instance, was cognitive science was important for all of neuroscience; did molecules matter for all neuroscience questions or just for particular problems. But the group did feel it was important to foster a sense that every level matters and encourage the different communities working on different types of problems at a particular scale to talk to each other, and they noted that there are already many groups who work across, different scales.

The group discussed whether (in order to promote cross scale science) one should identify individuals in the field who are already working across scales and support them or whether one should go to the rest of the neuroscience community and support them to do cross-scale modelling and cross-scale science. The suggestion for the latter was that one first describe what kind of levels exist – molecular, cellular, circuit level, behavioural level, etc., then ask what are the observables at these levels. What parameters (or models) at these scales will describe these observables? And what could be the interconnections between these different parameter observables, between one scale to the other? Finally, the important step was to look for some specific use cases where science was required to cross scales in order to answer some particular question or need for description.

A common argument is that neuroscience should benefit society, for example, by providing some answers to major diseases of the brain. But what is the proper level to study Alzheimer's disease? Well, probably the molecular level, particularly if we imagine some kind of drug is necessary. But this is also a good example of how neuroscience must bridge scales because one comes from a phenotype identification where one looks at the behavioural level. Neuroscientists might identify an animal model similar to the phenotype and then find pathology, for example changes in morphology or in the structure of the circuitry that allows the researchers to limit or narrow down what could be a potential molecule to study. For example, imagine finding that inhibitory cells from Alzheimer's patients have anomalies in relation to GABAergic release. This might indicate that a molecular process might be identified that influences GABAergic transmission and may have clinical relevance.

It was noted that collaborative work between labs takes a long time to develop agreement, even about the language used to talk to each other and what terminology means to each party (even when the scientific questions being asked were the same). Perhaps one role of the HBP could be to help develop lab cooperation.

Of course, every scale matters but there need not be a single way to integrate all models or data. The challenge is to integrate data across scales. And then one must be able to validate that cross-scale integration. What kinds of new experiments or technologies must be developed in order to validate cross scale integration of data or to determine the limits or particular questions for which the data integration is no longer valid.

So in summary, this working group agreed that every scale matters and hence that it was important to encourage and facilitate cross-scale science. The group felt there were two possible ways to encourage cross-scale science: One can provide exemplars of people already doing this whose approach can be emulated; or one could construct a system that allows scientists to contribute different levels and observables and models and which may then enable translation between levels on the IT infrastructure model (which represents the brain levels)- such a system can help modellers identify people that can develop experiments and validate (or invalidate) their model-based theory.

3.3 Community

This group had extensive discussions about the potential effects of a provenance system. For younger scientists an effective system of reporting data and code provenance potentially helps them with obtaining credit within their career for the work they have done, for example reputation points. For some in the group this raised the question: is there a risk that the collective spirit and free contribution system might be injured in an accounting system where every contribution was measured or priced? However members of the group felt less concern about this occurring in the scientific community than in other communities.

What were the challenges to ensure security of data within an open or partially open provenance system. Even if it is not clinical data from humans there is potentially concern about misuse. For example, contributing data to an open repository that is very labour intensive to collect (such as that obtained after intensive training of animals to undertake tasks) raises fears that one will be 'scooped,' that the first paper published analysing the data will come from someone outside the original data collection team. Some in the discussion wondered if the community is really so large that people can actually get away with this without suffering reputational damage. Others were concerned that between some (particularly competitive) labs there might even be worries about the provenance trail (without the data) being made public because this informs others know about what you are working on and how you are going about it. An open provenance trail might also make methods information on controversial animal studies publicly available. Are there processes that enable contributors to have some granularity of control over what they make public and to whom?

Because provenance data can also be thought of as data about those who have collected the data or written the code, the group asked whether the provenance tracking of the community contributions i could be used for employer surveillance? Does my rat data provenance record (when used by my employer to evaluate how hard I am working, what quality of data am I contributing, etc., for employment assessment) constitute me as a data subject (with rights under EU law)? To what extent is this restructuring of the work and work measurement system?

The group discussed whether a provenance system could be used to insure better quality control for data. While the first problem was to have metadata available so as to encourage collaboration and enable data sharing, the second problem is to improve data quality, experimental design, and generally to have good high quality science. At some level these are interlinked - transparency of process increases cooperation. Data has to be good data to be used by others. The desire to produce shareable data is likely to improve scientific work. However, the primary purpose of a provenance system is first and foremost to encourage data sharing, so it was generally agreed an emphasis on quality control was unnecessary, a distraction, or at least not a priority.

The group discussed pluses and minuses to data sharing. While transparency of information is generally good it must be remembered that lots of neuroscientists don't want to share data. They may have socially logical reasons for this. It is important to respect the differing culture of various sub-communities within neuroscience. While astronomers share data, biochemists share data, these historical examples demonstrate that it took time and work to build a culture of sharing data into the discipline.

The group also discussed how the design of a platform encourages (or can discourage) integration of communities. The platform does not engineer community relationships but rather it facilitates them. The immediate community of modellers is the first and most important community to facilitate. Face to face meetings help more than online meetings, although online knowledge of others builds expectation and creates success in a later face to face meeting. After face to face meetings happen collaboration at a distance is easier to make successful. Within the HBP it would be good to have more meetings between the sub projects, face to face. Often times simply working at another lab (cross-trains or visits) can be helpful but the best collaboration seems to happen in relationships built by working on a common project. For a platform it is important to have issue tracking so that the things that people need the platform to be able to provide for them can be addressed. Human support budgets are necessary for these services. Platforms need to be maintained. User questions need to be resolved. One platform builder suggested that if these additional costs in the planning weren't considered then we designers will have to service all our users ourselves! It was noted that the HBP subproject on neuromorphic computing (SP9) was building this into their budget.

The group discussed the differences between top-down and bottom-up community methods. Both bottom-up and top-down could work but they require thought about what is needed to make them successful. For example, bottom up communities requires participants to have salaries and travel budget - so most postdocs, etc. will have this (although not all). Being able to travel facilitates community projects and bottom up cooperation.

Past experience suggested it was helpful to consider the long-term sustainability of the network at the outset. While often highly charismatic leaders are needed for the success of a project, particularly during the initial period, it is also important to consider how to fill their shoes. Over a period of ten years the necessity of having particular key people should be replaced by infrastructure. Because there is good knowledge of how computing will change in the immediate future there is less risk of the products created for and by this platform becoming obsolete (irrelevant or not able to work with new IT systems in ten years). The use of NeuroML or other operating system/platform independent languages make any project less subject to fundamental changes in IT infrastructure provision.

The last part of the discussion began to flesh out who the users and first movers might be. It was thought that by prioritizing first movers and then adding other categories of users a basic road map for community building could be planned. When publishing occurs because of these tools, this will drive the adoption of the platform. Therefore the working group thought it would be useful to emphasize champions who might publish with the system and drive adoption.

3.4 Bi-directional translations between HBP and other platforms

This working group set itself the task of asking what it would really take to integrate the HBP, beyond what was presented earlier about the Collaboratory. They began their explanation with the simple definition of an app and went on to examine the technical requirements of an expanded more fully functional system.

An 'app' is a website. To get more functionality than embedding it in an iFrame with no connection, it can be embedded in JavaScript. That app will receive a token to do things within the HBP platform ecosystem on the users behalf. A software catalogue could list available HBP services and point to documentation. Some important services that need to be implemented are a monitoring API, a knowledge graph API, and a task service. A provenance system can point to data via URLs or something called UUIDs.

The group also discussed how this might integrate with the European Commission's ESFRI (European Strategy Forum on Research Infrastructures) roadmap. There was discussion about portals, comparing the Neocortical Microcircuit Collaboration Portal https://bbp.epfl.ch/nmc-portal/welcome associate with the Blue Brain Project with the microcircuits.epfl.ch data sharing website, how that will be updated and how data will be able to flow to the community once a specific Blue Brain paper is published. There were also constraints on use based on size of data flow. Since the dataset that represents the cortical microcircuit in a point neuron form is 500GB this is not so easy to download. There are trade-offs between what could be done locally with a smaller or summarized data set and what things would have to be done on the cloud.

Should the HBP be leading the way with making data public? The working group discussed the difference between the Allen Institute (which has a making data public team) and the HBP. One challenge for the HBP is that unlike the Allen Institute, they are not a single institution but many. However, data releases could be coordinated with the HBP communications team, particularly when multiple groups have data to release.

The working group spent some time considering the way that the OpenWorm works. Open uses 'agile' methodology. The group writes up tasks on Github, and this makes them open to any who want to participate by using an issue tracker. There are regular meetings every two or four weeks. In between those involved do what's known as a 'sprint'. They use a 'kanban' board to show tasks (laid out as cards on a board) that are 'backlogged', 'to-do', 'in progress', 'in review', 'done'. At every meeting they review what has been done, pull new tasks out of the back-logged pile and plan what will be next. All meetings are open, mostly on-line, recorded and then posted on-line. This makes the whole process open to volunteers. In between meetings there is also effort to recruit volunteers, often by using social media. There was discussion about how a volunteer system worked and did it have limitations for a big EU project. It was also noted that while such a community structure can create a lot of energy and dynamism it takes a lot to get it to happen. It is not free; it requires resources. This was the organizers' experience with Brainscales, a recent EU funded neuroscience project communities.

4. Conclusions

4.1 The way forward and next steps

The last day started with the four working groups presenting their ideas and propositions. All participants got a chance to ask questions and give their views, before splitting again in smaller groups that were tasked to establish tentative directions and roadmaps, starting from the working groups' proposals. The final plenary session was devoted to presenting these tentative directions and roadmaps, and discussing them.

(1) Modelling relevant to clinical practice

The working group presented what they thought needed to be done in several directions, within and outside the HBP, to move to a more collaborative relationship with European stakeholders and clinicians, and to establish trust in the HBP from everybody's point of view. The first direction they explored relates to the HBP Medical Informatics Platform (MIP), developed in sub-project 8. The HBP needs to find ways of demonstrating publicly how the MIP could contribute to the challenge of brain disorders. In order to achieve this, it should try and link the MIP more closely to 4 groups of stakeholders: European parliamentarians, so that they understand and trust what the MIP is doing; organized groups of scientists – it seems that to date, links have not really been established, yet need to be built, between the HBP's work in the area of Medical Informatics, that of the European Brain Council and that of the European Neuroscience Society; patients, and in particular the patients groups – the MIP should avoid over-claiming because patients would not believe it, but they could work together and that would encourage patients to share their data; and also seek to establish connections in the European data space especially around the use of big data for clinical research.

The second direction they considered would be for the HBP to look beyond the MIP, which has its own particular way of approaching the modelling of brain disorders based on data mining of big data, and to look for other approaches that would start from the challenge of particular disorders and how to model them. This would require involving a wider group of practitioners – clinicians, epidemiologists – and to be clear about what the clinicians might expect the HBP to provide. One key thing that clinicians would want is differential diagnosis of brain disorders. Another revolves around rare diseases, where often individual research groups and clinical practices only have a few cases: building up a big enough dataset to look at rare neurobiological or brain disorders, to enable analysis on powered up samples. In addition, there is a need for better drugs and therapy development, although that is a longer term issue, which raises difficulties with regard to commercialization and ownership.

The third direction that the group started mapping builds on the view that there are unexplored troves of data collected by teams who were already funded by the EU, in areas like traumatic brain injuries, stroke, or epilepsy (EPICURE project). The HBP could perhaps contribute to the analysis of that data and the standardization of data formats. Regarding standardization, the HBP has yet to decide which position they want to adopt – getting involved or not – but there are other organisations, like INCF, of which it is an explicit goal and that the HBP can collaborate with. But for partnerships to develop, data collecting groups must have an incentive to become involved with the

HBP. One suggestion was that there may be sources of funding within the EU where groups who were funded to collect data can apply for the curation of the data and for training purposes, in order to maximize data usage. Another incentive for the groups concerned would be the opportunity to get back bigger data and better algorithms. The group thought that these were three directions in which practical steps could be taken, that would not raise unrealistic expectations. Their last recommendation was that the HBP did not position itself as leading all these developments, but rather placed itself at their service, to help them and work together.

(2) Bridging scales

As a preliminary, the working group pointed that bridging scales has repeatedly been highlighted as one of the major problems – the most complex scientific problem – not just in the HBP but in general in neuroscience, which suggests that there is much more work to be done than can achieved by the HBP. In order to move the problem forward, the group has isolated what they think are some of the key issues. One is comparison between different species. It is a strategic objective. With the HBP moving towards an infrastructure focus on generic tools, the mouse and human limit that it had in the ramp-up phase are lifted. Groups working on all possible animal models must be able to use the platforms. This also raises the question of the relationship between the HBP and the other brain projects, which works on different species than the HBP. Possible complementary strategies should be explored. Another is the validity of extending datasets algorithmically. It is a key scientific approach used in the Blue Brain Project laboratory, and it is an important scientific question which discussion the HBP needs to be part of. Then, there is the issue of what theoretical approaches can be used to bridge scales. There is already a sub-project in the HBP dedicated to this, but it something the HBP probably needs to do more of and to get more communities working at different brain scales involved into. And finally, tools need to have specifications for the data which they can accept as valid. Regarding concrete actions, there are plans to build a workshop series into the agreement for the next phase of the HBP (SGA1) – the first two years of the operational phase. It should involve specific groups in the HBP (Theory, Simulation, also components from Neuroinformatics platform) and the respective communities outside of the HBP. The provisional title of the workshop series is "Systems Neuroscience: Scientific Integrity in Data integration", where integration is meant to span scales and close gaps. The hoped for outputs would be better scientific agreement on standards for these problems and a better sense of validity of the science being pursued. For concrete action regarding tool specification, it really became clear in the discussion that through the development a better validation protocol must be implemented with respect to the incoming data: specification of the data that is coming in, specification of some of the properties that they have, and validation before actual implementation in code. A concrete scenario envisioned is a format of workshop which picks a specific tool, brings the potential community to the tool, help them make their data work with the tool, and if it fails try in the same session to encode the validation check in code – the aim being to produce better tools.

(3) Communities

There was no further work done on the Communities roadmap after the working group presented their ideas and propositions from the previous day. The considerations that they had put forward to start thinking about a roadmap had primarily to do with users. For instance, who would be the first movers? What were the different categories of users that the HBP would target, and in which order of priority? Among the different categories inventoried with high priority were the internal users, then the wider circle of collaborators – like for instance the group of external modellers and experimentalists that was part of the Hippocampal modelling experiment initiated at the *HBP Hippocamp CA1: Collaborative and Integrative Modelling of Hippocampal Area CA1* workshop, held at UCL, March 31st – April 1st 2015. Another important category was that of the HBP partnering projects. The working group thought that on-boarding these different categories of users would eventually lead to overlapping collaborations and help reach critical mass. Further, broadening the scope of the tools and services offered by the HBP platforms, which largely relies on reaching critical mass, would allow for an increasingly broader user base. Different groups require different uses, and it is desirable to reach a stage where there is a positive feedback loop between the improvement – in breadth and strength – of the platforms' capacity to solve problems and have many uses, and the development of an ever larger community of users. Publishing the results brought about by using these tools should help drive the process of more users becoming involved – which means that there is a need for champions who will demonstrate the use of the tools and publish about it. In the longer term, another alley to explore would be how to open up the HBP platforms for use by a wider public or outside amateurs, to do citizen science. Lastly, the working group thought that in order to help pushing forward the recruitment of users and the development of a community, it would be important to attend to the credit system for all the intermediary roles involved: curators, coders, etc. Citing developers in methods papers is an example. Although publishing methods papers is not thought of as high profile, there are now places to publish such papers, and it should be noted that of the 100 most cited papers in science, nearly all are methods papers.

(4) Bi-directional translations between HBP and other platforms

One conclusion from the workshop discussions was that perhaps the most fruitful level to initiate collaborations between HBP researchers and those in other initiatives was at the medium to low hierarchical level – what was referred to as 'PI minus 2' – around practical use cases like a specific brain function or a specific brain area. One example is the collaborative work coming out of the *HBP Hippocamp CA1* meeting in London. Having more such initiatives could potentially be very successful in the short to medium term. These initiatives may not take place around the core activities of a lab, but can be initiated and managed by someone one or two steps below the top PIs in the form of quick and agile work programmes. Those can move forward with collaborations among those who want to be more open, and might produce results. So, a good way to engage the community and move forward might be to encourage such low level initiatives which do not depend on strategic decisions at the very top. As a complement and alternative to changing things top down with strategic decisions on collaborations, pushing things forward from the middle might be a way to go.

A second key topic concerned standards. In short, the conclusion was 'release, release, release' as openly and in the most user-friendly way possible. The first movers, if they are big enough, set their own standards. Sometimes there is a disconnect between technical people who want to be a bit more open, a bit more engaging with the community, and some PIs who may have other priorities. The example of the Allen Brain Institute shows the benefits of the open approach they are taking, and this can provide a powerful exemplar for PIs and Directors of the HBP, showing the value of openness.

A third topic focussed on the need to ensure that the design of the HBP platforms did not exacerbate the digital divide between the computational neuroscience community and the rest of the neuroscience community. The working group came up with the view that there should be close attention paid right now to all the different categories of potential users, and it was suggested that the HBP should, again, 'be more like the Brain Allen Institute', in the way that for instance their website is set up and makes it easy to find things. This relates to the challenges of community building and of bridging scales. To start collaborations on specific use cases that try to bridge scales, different schools of experimentalists and of modellers will need to come together, and the HBP infrastructure should play a positive mediation role: 'findability' is going to be key. There needs to be a focus on the user interface question – how to navigate the HBP platforms, how to find other groups, models, datasets, etc. Finally, the working group pointed that some of the elements of the HBP user access infrastructure that were presented as solutions would be worth opening to scrutiny and questioning, like for instance monitoring APIs. They involve complex flows of metadata that can potentially be used in all many different ways, some of which unforeseen, which are worth interrogating.

4.2 Take home points and future challenges

We left the workshop with a number of take home points and future challenges, which, we hope, will form the basis of future discussions, future work and future workshops.

4.2.1 How to create incentives for collaborative neuroscience for early career researchers?

- Open source software and good provenance tracking provide some technical solutions
- Creating a culture of open science and open data is vital
- Need to build resources for community building – financial, organizational – into the structure, budget and reward system of each research lab
- Need to build a durable infrastructure, including a data sharing facility, supported by the HBP, which would provide an incentive for collaboration between experimentalists and modellers so that networks of collaboration do not depend on specific individuals
- Need to reward and provide incentives for the curators of the data, the coders and those whose role is the crucial but less glamorous one of maintain the platform

4.2.2 How to build a collaborative neuroscience community?

- No aspiration for 'one model to rule them all' – all scales are important, but bridging scales is a challenge that needs more work: what brain scales are best for what problems, and what kind of data is best able to be combined across multiple scales?
- Need to open up the HBP platforms to multiple users from across the diverse constituents of the community.
- Need to develop a positive feedback loop between the internal and external users of the HBP platforms, so that the users feed back into the improvement of the tools as they use them
- Need to build links below the PI level in multiple local collaborative initiatives
- Need to find open source ways of developing standards, not forgetting the evidence that 'first movers' in open science often do shape the standards for others
- Need to develop multiple interfaces with cognate initiatives – for example between the HBP's Medical Informatics Platform and the European Brain Council, and the European Neuroscience Society, as well as with the multiple other European and international projects working on brain disorders (eg the EPICURE project) and with clinicians and patients, to

establish connections and standards in the European Data Space for the use of patient data in for big data clinical research.

4.2.3 How to tackle the challenge of engagement?

- Need to identify the different communities for engagement, and develop strategies to engage with each.
- Need to prioritize engagement with the potential user communities of the platforms, and this needs a different strategy from that of public engagement.
- Need to develop a clear and consistent public engagement strategy, to open up the communication between public expectations and the work of the HBP
- Need to open the HBP to the public, and invite them in as partners, not as data subjects: e.g collaboration between the Medical Informatics Platform and patients and patients groups.
- Need to ensure that the HBP does not become "a solution looking for problems" but begins with an engagement with potential end users and carries out this engagement throughout the project so that users are involved in shaping and evaluating the outcomes.

This workshop has been funded with support from the Fondation Brocher.
More information on the Brocher foundation programme is available at www.brocher.ch

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 604102, Human Brain Project.

Notes

[1] The meeting was recorded for the purposes of drafting this report: all participants consented to this, on the basis that, while the presenters would be identified, the contributions to the discussions would be anonymous, and that all participants would have the opportunity to review the reort before it was made public.

[2] 31st March and 1st April 2015, titled 'HBP Hippocamp CA1: Collaborative and Integrative Modelling of Hippocampal Area CA1,'

[3] http://neuralensemble.org/meetings/HippocampCA1/.

[4] HBP 1st Technical Review Report: http://ec.europa.eu/digital-agenda/en/news/1st-technical-review-human-brain-project-hbp-main-conclusions-recommendations

[5] "This will remain a significant scientific problem with this WP, and with the aims of the Brain Simulation Platform in general, and one which is not really satisfactorily solved by the algorithms and methods articulated in the progress report and the review presentation. This has highlighted the crucial role of the Neuroinformatics Platform [SP5], both with regard to ontologies and actual data, and the need for WP6.4 groups to interact closely with SP1 and SP5, to ensure that urgently needed data is given high priority."

[6] Indeed, this is also the diagnostic of the HBP technical reviewers, with the recommendation that Neuroinformatics should involve as many users as possible as soon as possible through "a concrete plan for attracting users to the platform, including the concrete incentives [it] offer[s] them to do so."

5 Workshop Programme

Programme

Wednesday 10th June

The Workshop will officially starts on Thursday at midday. However, those planning to arrive on Wednesday 10th are invited to an informal dinner at 19.30 at the Fondation Brocher.

Thursday 11th June	
10.00 – 12.00	*Arrival and informal discussions*
12.00-13.00	*Lunch at the Fondation Brocher*
13.00-13.30	**Welcome and rationale for the meeting** The challenge of building community approaches to modelling and atlasing in neuroscience and the role of the HBP *Nikolas Rose and Andrew Davison*
13.30-15.30	**From experiments to data repositories**
15 minute presentations followed by discussion	*Leslie Smith* - Experience with CARMEN *Martin Telefont* (BBP) - Data integration on the ground: what's coming and what is still needed *Catherine Zwahlen* (BBP) - HBPCORE, data and metadata lifecycles, Atlasing progress *Shreejoy Tripathy* - NeuroElectro.org: organizing the world's neurophysiology data and making it available for reuse
15.30-16.00	*Coffee Break*
16.00-18.00	**From experiments to data repositories II**
15 minute presentations followed by discussion	*Thomas Wachtler* -Tools and databases created by the G-Node *Sonja Grün* - Experience of ensuring reproducibility when collaborating with experimentalists *Paul Tiesinga* - Predictive neuroinformatics work in HBP

18.00-18.15	**Conclusions from Day One**
19.30	*Workshop Dinner*
Friday 12th June	
9.00-10.00	**Lessons from elsewhere**
15 minute presentations followed by discussion	*Christine Aicardi* – Interdisciplinarity and collaboration *Niccolò Tempini* – Big data and open science *Dennis Kipker* - Data Processing for Research Purposes – Current and Future Regulations in the European Data Protection Law
10.00-10.30	*Coffee Break*
10.30-12.30	**From data to model building and validation**
15 minute presentations followed by discussion	*Srikanth Ramaswamy (BBP)* - Data driven brain microcircuit modelling at the BBP *Marcel Oberländer* - Construction of data driven neuron models *Marc-Oliver Gewaltig* (BBP) - Data driven whole brain modelling at the BBP *Marmaduke Woodman* - Experience of data-driven modelling with The Virtual Brain
12.30-13.15	*Lunch at the Fondation Brocher*
13.15-14.45	**Pulling it all together: Community and platforms**
15 minute presentations followed by discussion	*Padraig Gleeson* -Experience of collaborative modelling with Open Source Brain *Stephen Larson* - OpenWorm *Jeff Muller* - Collaboratory Platforms, Provenance, Software Foundation
14.45-15.15	*Coffee Break*
15.15-15.30	Working Group kickoff
15.30-18.00	Working Groups
19.30	*Workshop Dinner*

Saturday 13th June	
9.00-10.30	**Conclusions and the Way Forward**
	Presentations by the working groups and discussion of their proposals
10.30-11.00	*Coffee Break*
11.00-12.30	**Practical Next Steps**
	Development of a Roadmap and Plan of Action
12.30-14.00	*Lunch at the Fondation Brocher*
14.00 onwards	Departure of delegates
14.00 – 15.30	Internal working meeting of Foresight Lab and Workshop Organisers

6 List of Participants

Attendees	Institution
From the HBP modeling and atlasing communities and ICT platforms architecture	
Jeff Muller	EPFL (Switzerland)
Andrew Davison	CNRS (France)
Paul Tiesinga	Stichting Katholieke Uni. (Netherlands)
Anrantxa Cedillo	EPFL (Switzerland)
Martin Telefont	EPFL (Switzerland)
Catherine Zwhalen	EPFL (Switzerland)
Sonja Gruen	Research Center Jülich (Germany)
Marc-Oliver Gewaltig	EPFL (Switzerland)
From non-HBP neuroscience platforms	
Padraig Gleeson (Open Source Brain)	Department of Neuroscience, Physiology and Pharmacology, UCL (UK)
Stephen Larson (OpenWorm)	MetaCell LLC (USA)
Shreejoy Tripathy	Centre for High-Throughput Biology (CHiBi) British Columbia University

Attendees	Institution
Thomas Wachtler	University of Munich (Germany)
Marcel Oberländer	Max Planck Institute for Biological Cybernetics (Germany)
Srikanth Ramaswamy	EPFL (Switzerland)
Leslie Smith	University of Stirling (United Kingdom)
Marmaduke Woodman	Institut de Neurosciences des systèmes (Inserm and Aix-Marseille University) (France)
Lia Domide	Codemart (Romania)
Other neuroscientists	
Matilde Leonardi	Fondazione I.R.C.C.S. Istituto Neurologico Carlo Besta (Italy)
Social Scientists	
Niccolò Tempini	Egenis, the Centre for the Study of Life Sciences, University of Exeter (United Kingdom)
Ethicists and regulators	
Jutta Lindert	European Public Health Association (EU)
Dennis-Kenji Kipker	European Academy for Freedom of Information and Data Protection (EAID)
Sabine Mueller	Charité – University Medicine Berlin (Germany)
From the Foresight Lab	
Nikolas Rose	King's College London (United Kingdom)
Christine Aicardi	King's College London (United Kingdom)
Michael Reinsborough	King's College London (United Kingdom)
Paola Bello	King's College London(United Kingdom)
Tara Mahfoud	King's College London(United Kingdom)
From the Ethical and Social Sub-Project of the HBP (SP12)	
Jean-Pierre Changeux	Institut Pasteur (France)
Kevin Grimes	Karolinska Institutet (Sweden)
Markus Christen	Universität Zürich (Switzerland)
Michele Farisco	Uppsala University (Sweden)
From HBP Co-ordination and Management	
Richard Walker	EPFL (Switzerland)
Kathleen Elsig	EPFL (Switzerland)