Part 1A - Previous Research Track Record

The applicants are, respectively, PI (**Ingram**) and Co-I (**Austin**) on the current EPSRC-funded e-Science pilot project which is supporting the development of the CARMEN portal. As such, they have an established track record in developing this neuroinformatics resource and have delivered a functional platform within three years. Their complementary areas of expertise in neuroscience (Ingram) and pattern recognition and e-Science (Austin) provide the core elements of the project and the two host centres (Institute of Neuroscience, Newcastle and Department of Computer Science, York) have considerable expertise relevant to the project. Many of the existing consortium members are based in these centres, and the project will enable many of these, and consortium members in the nine universities outside Newcastle/York, to continue to contribute to the development effort.

Professor Colin Ingram holds the Chair of Psychobiology and is Director of the Institute of Neuroscience at Newcastle. He has published over 100 refereed papers and is internationally recognised for his work on control of the hypothalamo-pituitary-adrenal (HPA) axis and the interactions of HPA hormones and other neuroendocrine factors in modulating neural systems relevant to stress, anxiety and depression. Since 2000 his work has principally focused on the interactions regulating the activity of serotonergic systems and collaborative work with Organon / Schering Plough on synergistic interactions between SSRI antidepressants and glucocorticoid receptor antagonists has led to a worldwide patent for novel combination therapy for affective disorders.

Ingram is founding Director of the Institute of Neuroscience at Newcastle (http://www.ncl.ac.uk/ion/) which is one of the UK's largest university departments focussed on the field of neuroscience, with 52 university academic staff (15 clinical) and more than 30 honorary clinical academics. He has led the Institute through a period of major expansion (22 university academic appointments since 2005) and has been instrumental in establishing several major technological and organizational initiatives. As well as leading the CARMEN project, he initiated funding for a programme of work in 'neurotechnology' which is exploiting small scale engineering and electronics for neuroscience applications, such as prosthetic implants. He has also led a strategy of internationalisation which has resulted in formal partnerships with two leading research institutions – RIKEN Brain Sciences Institute in Japan and Monash University in Australia.

The Institute has a strong translational research portfolio covering a broad range of basic neurosciences, from studies of synaptic transmission in invertebrates through to human cognitive systems, and research programmes in clinical sciences (neurology, psychiatry, neurosurgery). The recognised excellence in research has led to the award of £22.3M of current research grants. Major areas of research relevant to the CARMEN project include neural network dynamics, the neurobiological basis of epilepsy, studies of visual and auditory processing, motor control pathways, and informatics approaches to cortical connectivity. The Institute is one of the largest centres in the UK conducting both *in vitro* and *in vivo* electrophysiological studies in animal models, and is one of the few centres in the world conducting *in vitro* recordings from human cortical tissue. Research facilities include EEG for adults and babies, and a UK-unique MRI capability that supports work on human (3T), non-human primate (4.7T) and small animal (7T) subjects. In his roles as Institute Director and Honorary Secretary of the British Neuroscience Association, Ingram is well positioned to identify the potential user base for the CARMEN portal and to be able to leverage expertise to assist in its development.

Professor Jim Austin holds the Chair of Neural Computation and has led the Advanced Computer Architectures Group (ACA) for over 20 years within the Department of Computer Science at York. He has published over 250 papers and reports and holds 6 international patents. His research centres on artificial neural networks and is most well known for the AURA binary neural network technology that is used for searching signal data within the proposed project. Since 2000 he has applied AURA to e-Science problems starting with the £3.5M EPSRC e-Science Pilot project, DAME, that he led in collaboration with Rolls-Royce Aeroengines. This project developed a distributed prognostics and diagnostics platform, which was further developed in the £3M BROADEN TSB funded project and then taken to market by Cybula Ltd. The technology and methods from these projects allowed his team to develop the CARMEN platform within the current project. Five core e-Science research staff at York have remained throughout DAME, BROADEN

and CARMEN and this has enabled the team to build up considerable skills and reputations within the international e-Science community. Two of these staff (Weeks and Jessop) will remain on the proposed project.

The Computer Science Department at York is one of the UK's leading Computer Science Departments with over 40 staff (www.cs.york.ac.uk) and will move into a new £15M building in October 2010. The Department has extensive computer facilities including hosting the York node of the White Rose Grid (run by the ACA group), one of the world's first metropolitan grid computing platforms hosted between Leeds, Sheffield and York Universities (www.wrgrid.org.uk). The WRG is used to host the CARMEN portal and software and in total has over 650 processing cores across the three sites. The WRG is part of the national grid service, allowing future extension of the CARMEN platform. The WRG is a regional e-Science centre, one of only three to be supported by EPSRC. It is currently funded until 2014 to support e-Science projects within the three universities.

The ACA has over 30 members including three lecturers, one professor and two support staff. Current projects include the £7M TSB, DfT, EPSRC Freeflow project aimed at developing intelligent traffic management systems and is being trialled in York, London (Hyde Park corner) and Maidstone, and uses AURA to detect traffic incidents in the transport network.

In 2000 Austin founded Cybula Ltd (www.cybula.com) to support exploitation of the research from the group within industry. The profitable company sells advanced pattern recognition system software and hardware, and has developed the Signal Data Explorer (SDE) software in collaboration within the DAME, BROADEN and CARMEN projects. SDE allows distributed signal data to be viewed, managed and searched, and is mainly used commercially in diagnostic and prognostic applications to predict potential failures in complex engineering assets. Cybula's customers for SDE include Bombardier Transportation (trains), Rolls-Royce Aeroengines (aeroengines), power utilities, rail infrastructure companies, and the oil and gas industry. Cybula will be a partner in the proposed project supporting the SDE software used in CARMEN.

The CARMEN and DAME projects were a core part of the UK's aim to develop an internationally strong capability in e-Science, the use of computer systems to support collaborative computer assisted research, leveraging Grids and Cloud Computing (distributed computing platforms). Both projects were major showcase 'pilot' projects, part of nine projects funded by the e-Science initiative in the UK.

Other organizations relevant to the project include the International Neuroinformatics Coordinating Facility (INCF; http://www.incf.org). The development of CARMEN has been timely as it has coincided with establishment of INCF as a major international bioinformatics initiative. Along with thirteen other nations, the MRC provides the UK financial contribution to support INCF in conducting its objectives of fostering partnerships and undertaking work programmes to support and establish neuroinformatics resources. Our close association with INCF provides a route to access world leaders in neuroinformatics as well as knowledge of the latest developments in the field. Ingram serves on the INCF oversight committee for the programme on 'Ontologies of Neural Structures' and has been appointed to chair the programme to define minimal metadata standards for neuroscience. This programme has taken its lead from the work conducted in CARMEN on defining a metadata schema that underpins the data repository (Minimal Information for a Neuroscience Investigation, Gibson *et al* (2008) *Nature Precedings* : hdl:10101/npre.2008.1720.1). Ingram also serves on the Steering Group for the UK Node of the INCF which chaired by Professor David Willshaw. The project will work closely with both INCF and the UK Node to both leverage the global expertise in neuroinformatics, develop strategic links with other appropriate e-Science platforms, and to publicise CARMEN to the wider community (see letters of support from Willshaw and INCF Director Professor Sten Grillner).Two further research groupings relevant to the project are the UK Mathematical Neuroscience Network and UK Spike Train Analysis Network. These EPSRC-funded networks bring together researchers with interests in using mathematics to describe experimentally observed neural behaviour and in developing methods to analyse neural activity patterns. Members of these networks have contributed services to the CARMEN platform and can use data to explore and test hypotheses. Letters of support from the two coordinators (Professor Stephen Coombes and Professor Stuart Baker) are attached.

Part 1B - Statement on Data Sharing

In line with BBSRC Data Sharing Policy, CARMEN provides a mechanism to facilitate and encourage data sharing in the neuroscience community. The CARMEN repository provides a platform for researchers conducting neurophysiological studies to comply with this policy (and similar policies of other funders) and it is our expectation that re-use of data held in the repository will lead to new scientific understanding as has been demonstrated in other fields of neuroscience, such as the fMRI Data Center.

A strong motivation behind providing the infrastructure for sharing neurophysiological (time-series) data is the high value of datasets arising from complex or rare experimental paradigms. Examples of this are recordings from task-trained non-human primates, recordings from human cortical tissue obtained from rare neurological cases, or recordings from transgenic animals which are no longer available. In each instance the opportunities for repeating a study are low and the primary data have great value to the neuroscience community. However, even for studies which are simple to repeat, economic and ethical considerations may motivate sharing, where appropriate.

It should be remembered that CARMEN fulfils two roles, a primary role as a virtual research environment in which data are kept private or are shared within collaborative groups whilst analysis is undertaken prior to publication. The security structure implemented in CARMEN enables collaborating groups to have exclusive access to the data during this pre-publication period. In this context it should be noted that sharing also extends to the community-derived analysis software services which are uploaded to the system. The secondary role of CARMEN is as a publicly accessible repository where analysed and published data may be made available for secondary analysis. The timescale for transferring data from private to public will depend on a number of factors and we have not yet enforced a policy on when the data should be made public since the system has only been live since summer 2009 and most data are still undergoing analysis. Note that this approach to acquisitions is different from those resources which provide a database of legacy data from published studies, e.g. CRCNS.org and neurodatabase.org. These are static systems that do not allow any interrogation or analysis of the data and there is little incentive for researchers to upload data which are already published. In addition, sharing of neural activity data from such static systems is limited by the wide range of proprietary data formats generated by the various acquisition systems used for recording. In CARMEN we are extending the opportunity to share by using a common internal data format (NDF) and translators that can read most data formats. This means all users have access to view and analyse the stored data irrespective of the data file's original format. Furthermore, in order to ensure the value of the data to any secondary user we have implemented a metadata schema (Minimal Information about a Neuroscience Investigation (MINI): Electrophysiology, Gibson *et al* (2008) *Nature Precedings*: hdl:10101/npre.2008.1720.1).). This includes details of the data originator and we strongly encourage contact in order to avoid any misinterpretation.

Whilst funding arrangements mean that we cannot currently guarantee retention of the data for a period of ten years, there is an expectation that any data contributed to the platform will remain available for the foreseeable future.

Finally, as well as sharing with users who access the repository through the CARMEN web portal, we propose to work with other neuroinformatics resources to make data available for use within a wider federated database.

Part 2: Description of Proposed Research

1. Background

The study of the brain and nervous systems is often conducted using neurophysiological techniques, such as intra- / extra-cellular recording, multielectrode arrays, and EEG, in order to understand how individual neurons and networks are controlled. These techniques generate large volumes of time-series data which need to be subject to detailed analysis in order to extract important information about regulatory mechanisms and patterns which contribute to overall activity. Increasingly this research is conducted by multidisciplinary teams comprising experimentalists who make the recordings and analysts who provide the methods to process the data, and these are not always co-located. Whilst a number of areas of neuroscience have been able to leverage the developments in e-Science to facilitate cooperative working and sharing of data resources (e.g. structural imaging data or genetic data), there is currently no system capable of supporting the type of time-series data generated by neurophysiology.

CARMEN (an acronym for *Code Analysis, Repository and Modelling for e-Neuroscience*; http://www.carmen.org.uk/) is an e-Science resource that enables cutting edge research on time-series data (signal processing) enabled through data and software sharing. Initially aimed at supporting neurophysiology, the core technology of the e-science platform also has potential application to other areas of bioinformatics that generate time-series outputs. CARMEN provides the infrastructure to support data sharing (both private and public file sharing), analysis of complex datasets (through access to user-defined analysis services), and long term curation of data which are generated by neurophysiological experiments. The resource aims at optimal reuse and integration of data, particularly where this applies to experiments that are complex to perform or which employ rare tissue. CARMEN was established by the EPSRC as a community-based e-science pilot project for a fixed term (2006-2010). A very considerable amount of development work has taken place and the pilot project has already delivered an extensible and distributed computing infrastructure with integrated metadata and analysis services accessed through a web portal. The initial release within the development consortium has demonstrated the utility of this resource through a number of new research projects and by the end of the existing funding will be fully operational, providing data archiving and workflows of services. However the project now requires more sustainable funding to maintain the resource and enable continued development.

CARMEN provides a highly sophisticated open access resource primarily for experimental and computational neurophysiologists. The CARMEN platform fulfils two major functions: firstly, it provides a web-based work environment (the portal) in which groups can collaborate by sharing their data and undertaking analysis and, secondly, it provides a repository that allows data and services to be shared widely within the neuroscience community, thereby achieving best value from research. The resource itself comprises both data and services for these data (analysis tools), and to the user it takes the form of a web portal that allows access to the resource. It currently resides on parallel cluster machines in York and Newcastle e-Science Centres, but because it consists of a set of web services it can be federated across multiple machines at other sites.

The following functions are currently available, or will be delivered by October 2010:

i) **Data upload and download:** We have developed a highly assistive web portal for sharing of data files. To cope with the multiple variants of proprietary (raw) electrophysiology data formats, we have developed a common internal file format, Neuroscience Data Format (NDF), which enables execution of services within the platform independent of the original format.

ii) **Metadata collection:** Integral to the data import function we have implemented a domain-specific extensible metadata schema developed within the project (MINI – Minimal Information about a Neuroscience Investigation: Electrophysiology) and a template-based architecture that facilitates rapid input of experiments of similar design. The design of this import function allows us to quickly build new metadata formats for other application domains.

iii) **Visualisation:** To enable display of the time-series data files CARMEN incorporates a powerful client-based signal analysis tool, Signal Data Explorer (SDE) which also has integral pattern search and matching technology (see figure). SDE is being developed in conjunction with our project partner Cybula who will continue to provide support for its development within CARMEN.

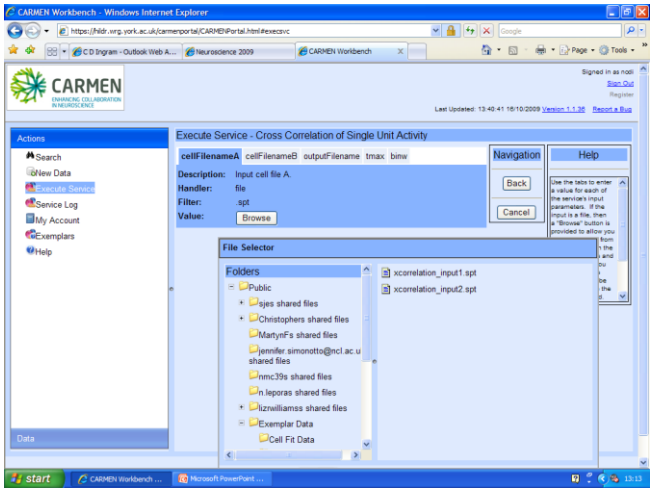
iv) Analysis services: The portal supports deployment and execution of analysis services (see figure). These are wrapped as web services allowing easy deployment and can be linked to run as client-defined workflows. The system currently has services for filtering, spike detection and sorting, as well as high level statistical and pattern analysis, tools for network analysis, and tools for analysing spike trains using information theory and Grainger causality. Workflows may be set up, stored, modified and recalled, enabling the automation of analysis of datasets.

v) Workflows: The portal will have the capability for running a number of services within a workflow and we are currently integrating technology to enable this. In this process the user selects the services to run together using a graphical editor to make frequently used scripts. These stored and shared in the same way as data and services.

vi) Security: In contrast to simple databases, CARMEN provides a working environment to support collaboration from the point of initial data collection through to publication. A security function controls access to data and services with authentication and authorization of users controlled through a simple interface in the portal. Once published, the data can be made public as part of a growing resource available to the neuroscience community for re-analysis.

vii) Search and provenance functions: The portal includes functionality for indexing and querying of the data and services, and we are working on the linkage between data and service descriptions so that users can intelligently discover appropriate analysis services for primary and derived data. Provenance traces will enable users to search prior analyses.

viii) Real time collaboration: Packaging and export of data directly from the experimental recording system will enable researchers share and analyse data in near real time in order to benefit from a web-based working environment.



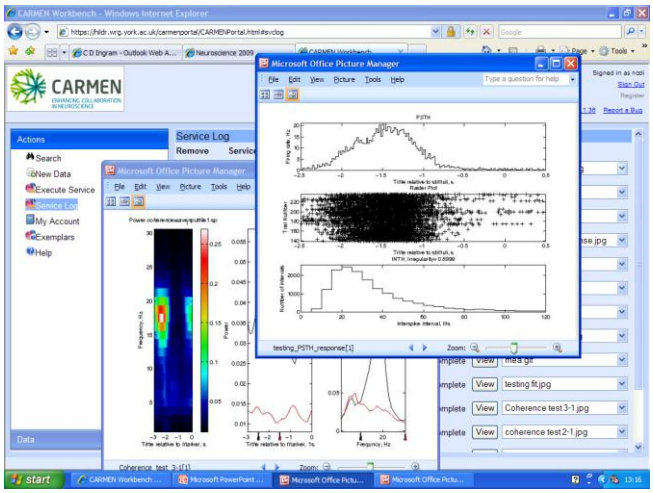


Figure. Screen shots of version 1 of the CARMEN portal illustrating (left) how during execution of a service parameters are entered via a GUI which is automatically built from the service metadata / WSDL; (lower left) the output from a service which can be either graphical or a results file that can be exported; and (below) the interactive data viewer (Signal Data Explorer) which can be used to display whole multielectrode arrays (inset) or selected data channels (main window), and to select events for rapid pattern matching.

2. Objectives and Programme of Work

(i) Objectives

Our overall objective is the continued deployment and further development of the CARMEN neurophysiology resource beyond the end of the current EPSRC e-Science pilot project funding. By October 2010 the CARMEN resource will have achieved a high level of maturity in respect of the user base and level of deployed data and services. If the neuroscience community is to profit from this development effort, firstly, it will be necessary to maintain the CARMEN portal, involving routine management of hardware and software, and support for user groups and, secondly, revisions and further technical developments are required in order to extend its functionality. The Bioinformatics and Biological Resources (BBR) funding will enable the transition from the current project-based development with focussed user group into a community-wide resource, and enable increased engagement both from BBSRC-funded project areas and from the wider neuroscience community. The BBR funding will enable the project to: (i) extend the user base into a true community facility and provide support for users; (ii) enable user-defined developments (for example developing novel workflows or services); (iii) support technically-oriented development work, particularly around support for EEG and optical recording methodology; and (iv) support development work on interoperability and interfacing to other complementary resources. No research activity will be directly funded, although a number of projects and PhD studentships associated with the CARMEN programme will continue to generate new data and services that can be deployed within the resource. By developing its full potential CARMEN is expected to underpin international quality UK neuroscience and support the Council-wide strategic priorities in neuroscience.

(ii) Programme of Work (see accompanying work plan)

In addition to basic system maintenance to enable continuity of the existing platform, the planned programme of work will involve the following main activities:

i) **Engagement:** To ensure maximum opportunities for extending the user base across the neuroscience community the project will deliver a comprehensive programme of promotional and support activities involving a combination of targeted user recruitment, displays and demonstrations at key conferences, and good support through user group meetings and training workshops, with the aim to increase uptake of the resource both for collaboration and curation of data. Neuroscience conferences targeted for promotional work will be the major conferences in the UK (BNA), Europe (FENS) and USA (SfN), as well as the annual congress of the INCF which has an exclusive focus on neuroinformatics. We will extend our work on developing special interest groups who are able to share common analysis methods, and will target individuals with high value data that are expected to generate high demand for access. We will also work with individual funding agencies (BBSRC, MRC, Wellcome Trust) to develop a framework for support of funded projects. Publicity will include regular updates of the project web site, and the distribution of an electronic newsletter and e-mail shots. Measurable targets for this engagement activity is to at least quadruple the number of active current users (currently 75) and to identify at least 15 published outputs that have directly arisen from the use of CARMEN.

ii) **User Support and Training:** In order to extend the functionality of CARMEN a key activity will be providing technical support for the deployment of new client-generated services, including the wrapping and testing of software as web services, and support for any problems which users may encounter during upload or access to the system. This will be provided on an on-going basis, while Training Workshops aimed at specific development efforts and User Group Meetings aimed at achieving more general support and feedback will be held at 6 month intervals for more direct support.

iii) **Revisions to the Portal and Internal File Format:** Version 1 of the CARMEN portal was released to the development consortium in March 2009 and, following revisions based on user feedback, a second version of the portal is due for release in late January 2010. Whilst future minor revisions will be on-going, we plan to release new versions of the portal on a two-yearly cycle in order to integrate new features and major revisions arising from user feedback. These revisions will involve a period of review within the user base, development work, and usability

testing prior to implementation. We will also continue new versioning of the internal file format (NDF) to ensure continued interoperability of raw file formats.

iv) **Domain-specific Technological Development:** Whilst many of the features of the portal are generic to neurophysiology we plan to undertake technically-oriented development work in order to configure the system to specifically support two application areas with high potential for data sharing and analysis using community-generated services. The first will be to support data files of recordings of EEG and event-related potentials. Whilst the current platform already has the capacity to support EEG, additional specific analysis tools and data capture / playback configurations need to be implemented to meet the requirements of this user group. This is the largest application area in human neurophysiology and has some specific requirements, such as power spectral analysis in selective frequency bands and visualisation based on the International 10-20 system. The second area will be time-series data that can be extracted from video images arising from optical recording (e.g. voltage- or calcium-sensitive dye imaging). Optic recording technology is increasingly being used in place of electrodes and in order to support this CARMEN will need to integrate suitable video display, methods for data extraction from areas of interest, and spatiotemporal analysis routines. We will also need to respond to changes in data volumes. Developments in recording methods means that the volume and complexity of data generated is rapidly increasing. For example, the consortium currently generates data from 64 electrode planar arrays and 100 electrode Utah arrays and through collaboration we have recently been uploading data from 512 electrode arrays (Litke, CERN/Santa Cruz) and 4,096 electrode on-chip arrays (Berdondini, Genoa). The dimensionality of future datasets is likely to increase.

v) **Interoperability and Integration with Other Resources:** Development work will be undertaken to enable interoperability and interfacing to other complementary neuroinformatics resources. This will involve links with resources that hold time-series data, as well as interfacing with structural and genetic repositories in order to work towards a comprehensive federated network that can support a broad range of neuroscience data. For example, a database of cell morphology and regional gene expression has been implemented through the US-based BIRN Network and it would be extremely valuable to link these structural data to recordings made from the areas or individual cells. This integration will be facilitated by involvement of the International Neuroinformatics Coordinating Facility (INCF) with which the project has close association (see letter of support from INCF Director Professor Sten Grillner). We will also develop a partnership with the JISC-funded NeuroHub project which exploring ways to facilitate data collection in neuroscience. A joint plan will be developed to ensure interoperability within the two projects (see letter of support from Professor Anne Trefethen).

3. Demand for the Resource and Benefit to Users

Neuroscience is a major branch of biological research and is undertaken in many hundreds of laboratories world-wide. Neurophysiological recordings (or other time-series data) constitute a major part of research within this field and underlie our understanding of the activity of the brain and nervous systems. Whilst this effort can continue as isolated labs, a detailed understanding of the way in which neural information is encoded in neuronal networks can only arise by synthesizing data across multiple groups. The CARMEN resource has the potential to be an important bioinformatics tool in this challenge. Importantly CARMEN is not simply a data store: it includes routines for the analysis of these datasets, and these have been contributed by many different computational (or computationally oriented) neuroscientists. As the facilities have been integrated on the CARMEN portal, there has been growing demand for its services. The portal was first released to the consortium and to selected collaborators in March 2009 and over the initial 10 months use has grown to 75 registered users and 5165 uploaded files. We expect to achieve significant further engagement over the remaining period of the project and have adopted a number of strategies to achieve this (see *Approach to Acquisitions* below).

The current and future demand for this resource will be generated by a number of different user groups with differing benefits:

i) Users who wish to work in an e-Science collaborative environment. The benefits to these users is the ability to accelerate the research process (including real-time working) by using existing code and data rather than re-creating existing methods and data, allowing more effective

comparative work through better data sharing, and enabling collaboration across geographically distributed groups;

ii) Experimentalists who wish to share their data. This group will benefit from gaining greater value from their research and the ability to share and aggregate data. This is likely to have the greatest benefit for data which are rare or of high inherent value because of the effort required in obtaining them. Currently within the consortium this includes *in vivo* recordings from task-trained non-human primates and *in vitro* recordings from human and non-human primate cortical tissue.

iii) Data analysts who wish access to data in order to conduct higher order meta-analyses, or to synthesise or test theories and models. The availability of a large resource in a single location will be of immense value for this research which is currently considerably hampered by the lack of suitable data. The rapidly growing field of neuroinformatics is likely to increase this demand.

iv) Service developers (computational neuroscientists) who require a platform on which their analysis tools can be deployed. At the moment this can be done through personal web sites or toolboxes (e.g. the FIND Matlab toolkit) but a major limitation is that implementation is limited by the format of services or data. CARMEN overcomes this limitation of interoperability by using an internal file format (NDF) that allows interchangeable data formats for implementation of services.

Our experience within the current CARMEN consortium has highlighted both the difficulties and benefits of developing and employing the resource. The consortium comprises 19 investigators (plus an increasing number of external collaborators) with complementary expertise in experimental neuroscience, e-Science, computational neuroscience and modelling. The consortium also has considerable expertise in delivering large-scale e-Science and neuroscience projects managed over distributed locations. This combination of expertise and motivations has been critically important for the success of the project. The aim of the original e-Science grant was to enable the sharing between groups of neurophysiology datasets (which are inherently difficult and expensive to generate), and the development of techniques which would allow these datasets to have a useful life beyond the end of the project which created them. This has been a challenging task but most of the major challenges have been overcome. The CARMEN resource is expected to result in significant high impact outputs. Even at this early stage we have demonstrated considerable value to a number of high quality projects with outputs including: **(i)** standards and schema that can be adopted by the wider bioinformatics community. For example the MINI metadata schema which was published in *Nature Precedings* (npre.2008.1720.1) and is currently being evaluated by INCF for adoption as the basis of an international standard for neurophysiology; **(ii)** publications reporting services and bioinformatics infrastructure provided by CARMEN, e.g. Fletcher *et al* (2009) Neural network based pattern matching and spike detection tools and services in the CARMEN neuroinformatics project. *Neural Networks* 21:1076-1084; Ince *et al* (2009) Python for information theoretic analysis of neural data. *Frontiers in Neuroinformatics* 3:article 4:1-15; and **(iii)** publications that have employed the infrastructure or analysis services developed in CARMEN in order to conduct high impact research projects, e.g. Hennig *et al* (2009) Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity. *J Neurosci* 29:1077-1086; Schultz *et al* (2009) Spatial pattern coding of sensory information by climbing-fiber evoked calcium signals in networks of neighboring cerebellar Purkinje cells. *J Neurosci* 29:8005-8015.

Significant demand for the CARMEN resource is expected to come from the BBSRC portfolio. The BBSRC OASIS database lists 37 currently active projects (i.e. with estimated end dates after 9/09) with a combined award value of £19.2M, that involve collection of time-series data that have the potential to be supported by the CARMEN infrastructure. These include single cell electrophysiology (extracellular and patch clamp), multielectrode array recording, evoked potentials and EEG/EMG. It is predicted that 40 further neuroscience projects generating time-series will be funded during the four years of requested support. Significant number of projects also funded by other government agencies (MRC, EPSRC and NIHR) and medical charities (notably Wellcome Trust). We will target support to these projects through discussions with each PI on the project.

4. Uniqueness and Strategic Relevance

(i) Uniqueness within the UK and Internationally.

CARMEN is unique in delivering a web-based resource for neurophysiology that provides extensive functionality and a repository function. There are other projects which hold neurophysiology data permitting others to use them (e.g. the Collaborative Research in Computational Neuroscience (CRCNS) database http://www.crcns.org), but CARMEN is considerably more sophisticated in providing: (i) services for processing these data; (ii) a security system which permits users to control who can and cannot share data, metadata, services, and workflows; and (iii) a computational capability allowing large datasets to be analysed without moving large volumes of data around the internet. Whilst CARMEN has been a UK-led project the provision of this resource has a strong international dimension and it has contributed to the global infrastructure required to underpin modern neuroscience research. A number of the neuroscience projects which have been supported involve collaboration with groups outside the UK.

Development of CARMEN resource has been conducted in close collaboration with the International Neuroinformatics Coordinating Facility (INCF) in Stockholm and its national nodes (notably those in the UK, Japan and Germany). It is timely that the project is being undertaken at a moment when this major international initiative is providing a framework to coordinate effort and support the necessary standards and protocols for interoperability. INCF has initiated a development programme on time-series data in which CARMEN is a lead project and it is also facilitating partnership arrangements with related work being undertaken within the Bernstein Network for Computational Neuroscience, Germany. INCF is also working with suppliers of acquisition software coordinating efforts to support open access to file formats, continuing the objectives of the Neuroshare initiative. There are no equivalent commercial resources, although a number of commercial organisations have expressed interest in using the infrastructure, particularly within the pharmaceutical sector.

(iii) Strategic Relevance to the BBSRC

Neuroscience is a core area of activity for both BBSRC and MRC and CARMEN will deliver a resource that will be widely applicable to many funded projects. It directly addresses the objective to gain a greater integrative understanding of biology through the use of tools and experimental data as set out in the *Bioscience for Society Ten-Year Vision*. Furthermore, by providing a platform for sharing and implementation of analysis tool for time-series data CARMEN fits with the *mathematical tools for biology* element of BBSRC's programme. CARMEN also addresses the priority area of Technology Development for Bioscience and the support for increased International Collaboration which are highlighted in the current priority document. Importantly CARMEN will help deliver the BBSRC data sharing policy in respect of projects employing time-series data.

5. User Access and Acquisitions

(i) User Access Arrangements

Access to the CARMEN resource is through a web portal and is, therefore, available to any registered user with an internet connection. We have implemented registration and log-on arrangements, and prospective users can have a level of security associated with them, allowing them to be members of particular groups. Those requesting read-only access can receive automatic registration, while users wishing to submit datasets or services will require to submit credentials to validate their research status. This is currently limited to them using an e-mail address from a recognised academic or research institution to which their registration confirmation can be sent. However, we will review this policy if evidence from periodic audits of new submissions suggest that data or services are arising from potentially unreliable sources.

(ii) Approach to Acquisitions

In developing the publicly accessible data repository it is important to note that experience in the field of neuroinformatics has repeatedly demonstrated that there is far less motivation to contribute data when the resource is only a data store of published data (except where this is enforced by the publishing journal, e.g. the fMRI Data Center and its link to the *Journal of Cognitive Neuroscience*). The effort required to upload and annotate already published work is a major barrier to participation when the benefits are unclear. By building the public repository on top of a collaborative research

environment we believe that CARMEN significantly reduces the barriers for engagement. Therefore, we propose to continue our current approaches to acquisitions which are:

a) *ad hoc* acquisitions that arise from engagement of users who use the resource in order to undertake analysis and collaboration. Data may be made public immediately or can be kept private until publication or completion of analysis. We will strongly encourage users to make data public at the earliest opportunity and propose to enforce a policy that data residing on the platform will become public after a fixed period (currently suggested to be two years). Data contributing to published work will be tagged with the publication details as part of the quality assurance procedures. When data are made public full metadata will also be exposed.
b) Targeted acquisitions of specific high value, high quality data. For example, within the consortium we are currently building a database of human and non-human primate cortical recordings and recordings from high density electrode arrays which will be of great value to network modellers. In order to overcome to effort barrier described above and encourage submission it may be necessary to offer assistance with data entry to researchers who originate these high value data. We will similarly target important analysis tools particularly those which are documented as being interesting/useful in the literature.
c) Support for development of specific interest groups who can share similar tools or combine datasets to perform large scale analysis. By encouraging collaboration using the CARMEN platform a database of a specific data type will develop, thereby attracting others in that field to participate. For example we have established a group interested in multielectrode array recordings from the retina: Sernagor (Newcastle); Hennig (Edinburgh); Eglen (Cambridge); Wong (St. Louis); Feller (Berkeley); Litke (CERN/Santa Cruz). Interest groups can also be formed around development of specific types of services (e.g. information theoretic analysis).

6. Publicity, Engagement, and Training Programme

(i) Raising Community Awareness

The CARMEN resource is already strongly connected to a proportion of the potential user community both in the UK and internationally and our user base beyond the development consortium has grown rapidly. It was officially launched at the national meeting of the British Neuroscience Association in April 2007 and was demonstrated at the BNA meeting in April 2009. It has been demonstrated at congresses of the Federation of European Neuroscience Societies (FENS), US Society for Neuroscience, Computational Neuroscience (CNS2009), and International Congress of Neuroinformatics. It has close links to the neuroinformatics community through INCF and its national nodes, and the UK node has supported workshop events around CARMEN activity. Our programme of engagement activities will continue this promotion and publicity in order to access a wider part of the community. Information about the project will be provided through the web site (www.carmen.org.uk) and those registering interest will be added to a mailbase to receive an electronic newsletter. We will also continue our strategies of direct approaches to individuals with high value datasets or services, and the creation of selective interest groups (see 5(ii) above).

(ii) Training Policy and Activities

Support manuals, step-by-step examples, on-screen help, and web-based FAQs are being developed as part of a highly intuitive portal. This should be sufficient for most read-only users and data contributors. Service submission may eventually be equally intuitive, although currently and for the foreseeable future we will provide direct 'help desk' guidance and periodic workshops to support deployment of web services. Training courses will also be targeted to groups likely to lead to significant engagement with service provision, e.g. the new intake of PhD students on the Neuroinformatics and Computational Neuroscience DTC in Edinburgh, or meetings of the UK Spike Train Analysis Network or UK Mathematical Neuroscience Network.

7. Project Management (see accompanying diagram)

(i) Management Structure

The current Executive Group which has successfully delivered the EPSRC CARMEN pilot project comprises a mix of neuroscientists and computer scientists. Its members are: **Prof Jim Austin** who is Lead of the Advanced Computer Architectures Group at York and brings expertise in grid systems for management and analysis of time-series datasets; **Prof Colin Ingram** who brings engagement to the experimental neuroscience community both as Director of the Institute of

Neuroscience at Newcastle and Honorary Secretary of the BNA; **Prof Leslie Smith** who is head of the Department of Computing Science and Mathematics at Stirling and has expertise in computational neuroscience; **Prof Paul Watson** who is Director of the Informatics Research Institute at Newcastle and UKRC Digital Economy Hub on Inclusion, and brings a track record in developing distributed data base systems; and **Dr Tom Jackson**, who has a background in neural networks and has project managed the CARMEN computer architecture. Given the smaller size of the continuity funding and the greater emphasis on user engagement it is proposed that the Executive Group managing the CARMEN resource be reduced to only Ingram and Austin with one other member of the user community (selected on the basis of the level of engagement at the time of award). The remaining members of the current Executive Group will be invited to join the Advisory Panel. Other individuals will be invited to join the Executive Group when specific initiatives are undertaken or additional funding awarded. The Executive Group will meet bimonthly and maintain regular contact through e-mail, and will be supported by the Project Secretary. Decision making will be informed by regular audit of use and submissions to the resource; by user feedback polled from the web site and from proactive surveys; and by guidance from the Advisory Panel. The Executive Group will have responsibility for: (i) delivering a programme of activities to increase user uptake of the resource; (ii) prioritising the technical developments of the portal; (iii) engaging with stakeholder groups, including neuroinformatics and e-Science policy and standards groups; and (iv) formulating and implementing a sustainability plan, including preparing further applications for funding.

(ii) Advisory Structure

The current CARMEN Advisory Panel comprises 12 members drawn from across fields of bioinformatics, computing science, commercial users, neurophysiology product companies, and academia. The current membership is: Phil Boden (Lectus Therapeutics); Peter Cowley (Quarndon Cognition); Goutham Edula (Astra Zeneca); Sten Grillner (Karolinska Institute); Chris Larminie (GlaxoSmithKline); Nicholas Le Novere (European Bioinformatics Institute); Savas Parastatidis (Microsoft); Helen Parkinson (European Bioinformatics Institute); Greg Smith (Cambridge Electronic Design); Casey Stengel (Neuralynx); Shiro Usui (RIKEN Brain Sciences Institute, Japan); and David Willshaw (Edinburgh University). We propose to continue with an Advisory Panel of similar size, although the membership will be modified to reflect the changing emphasis of the project, whilst maintaining a balance of national/international, academic/commercial, and neuroscientists/informaticians. We will also ensure engagement with national and international advisory groups, particularly with INCF as the main group developing standards and coordinating international efforts in neuroinformatics. In this respect Griller (INCF Director) and Willshaw (UK Node Coordinator) will provide direct links and Ingram is Chair of the INCF Oversight Committee on metadata and part of the INCF working group on time-series data. The Advisory Panel will receive biannual reports on activity and will meet annually to provide input on strategic direction.

8. Long Term 'Options Analysis' for Sustainability

Development of the CARMEN resource is very timely as the INCF was established at the same time as the project and is actively fostering the development of neuroinformatics resources, while advances in cloud computing are greatly reducing the cost of access to compute power and large scale data storage. Furthermore, there is growing desire amongst public research funders and journals for researchers to provide access to original data files. CARMEN can offer the infrastructure to support this curation for the specific area of time-series data. We hope to be able to sustain a service that is free at the point of use since this will foster engagement, and to achieve this one option to be explored is negotiating a small surcharge (proportional to data volumes) on all publicly funded projects that generate suitable data that may be deposited with the resource. Across a large number of projects this will generate a revenue sufficient to support on-going development and maintenance (including rental of compute resource). International adoption of the resource can provide access to other funding streams which may be administered through INCF. During months 35-40 of the project the Executive Group will undertake a review of usage, demand, funding opportunities, commercial partnerships, and research council priorities in order to formulate a Sustainability Plan which will be presented to the Advisory Panel and to the BBSRC in order to generate a recommendation for the future of the CARMEN portal.