Bioinformatic and Biological Resources Consultation

February 2006

A consultation to develop a better understanding of the resource requirements of the bioscience community, in particular in the areas of bioinformatics and biological resources.

Background

BBSRC's 2005-08 Delivery Plan stated that BBSRC will invest £16M (pre full economic costing) in new tools and resources to equip researchers for 21st Century bioscience. It explains that the funds will be used to **develop the tools, technologies and resources needed to deliver BBSRC's vision of predictive and quantitative bioscience** including the provision of the e-tools, large data sets and informatics on which bioscience will increasingly depend. The Delivery Plan also recognised the need to provide proper support for bioinformatic and biological resources such as databases, genetic resources and culture collections which require long term maintenance and curation.

In June 2005, the Tools and Resources Strategy Panel agreed that the funding for tools and resources should be invested through three routes:

- A Tools and Resources Development Fund
- A Technology Development Research Initiative
- A fund to provide a more sustainable future for bioinformatic and biological resources (such as those outlined above)

The three routes were approved by Strategy Board in September 2005 and the Tools and Resources Development Fund and the Technology Development Research Initiative were launched in October 2005.

The Strategy Panel held an initial discussion of the funding stream for bioinformatic and biological resources at its meeting in December 2005. Members recognized that many such resources exist and that **funding is limited** (anticipate approx. £6-7M for 2006-2007). The Panel agreed that support should be directed towards a few **highly strategic investments** and decided that a consultation should be undertaken in order to develop a better understanding of the needs of the BBSRC community and how any support for resources would be best deployed. Finally, the Panel noted that the data sharing policy, currently under development, may need to be recognized in the scope of this funding mechanism.

Consultation

This section sets out the key areas in which the Strategy Panel would like to receive input. A series of comments and questions are included in order to prompt discussion. They are not exclusive and comment on other issues considered relevant is welcome.

1. **What bioinformatic and biological resources do you think should be supported?**

   Points to consider here include:

   a. It is proposed that funding:

      - should focus on resources critical to UK bioscience research and without which the community could not achieve its full potential.
      - should focus on resources which are / would be well used by a broad UK community.
      - should underpin the delivery of BBSRC's strategic goals. For example, does systems biology present any new resource needs? If so, what are they?

   b. What resources do you anticipate will be required for future bioscience developments?

   c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

   d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

   e. Are there any resources that are currently missing?

   f. Are there any resources that it would be inappropriate for BBSRC to support, for example for ethical or societal reasons?

   g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources which you consider could be relevant to this funding mechanism?

2. **How do you think should the funding be used ?**

   Points to consider here include:

   a. Should emphasis be placed on supporting resources previously established by BBSRC funding?

   b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)?

   c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)?

   d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges?

e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence?

f. Should funding be focused on publicly accessible resources only?

3. **What are the key components of '*more sustainable*' support?**

   Points to consider here include:

   a. Duration of funding;

   b. Location and management;

   c. Monitoring and assessment; and

   d. When would you anticipate that funding is no longer required (e.g. lack of use, operating on cost recovery basis; resource no longer of high strategic relevance to current UK bioscience research)?

4. **Other issues**

   Looking to the future and beyond the remit of this activity, are there any other resources (out with bioinformatic and biological resources) that could be important for future UK bioscience research?

Submitted comments

John Innes Centre

This submission has received wide consultation through the organisation covering 'wet' science, bioinformatics, computing and biological resource managers.

Bioinformatic resources requiring support:

We believe that databases and systems with wide application to the UK research community (e.g. ATIDB, SCODB, ENSEMBL, REACTOME) must be supported, where support also includes hand curation and ontologies. Alongside this, the delivery of large databases to the desktop should be considered. Mechanisms to achieve this could be through provision of either high bandwidth to sites with databases, such as EMBL, or seamless delivery to local servers to include indexing. Embedded, localised, support staff to facilitate the above and to drive the exploitation of these resources through direct interaction with bench science is also essential. Funding committees must be sympathetic to inclusion of these staff resources into grant proposals where they are justified. Funding of high performance central computing power (like HPCx) should also be considered.

Training in bioinformatics should be given appropriate priority and an assessment made as to whether courses currently available in bioinformatics are fit for purpose? Is there sufficient (any) bioinformatics content in undergraduate biology courses? Are MSc courses teaching appropriate material, such as including languages most relevant to the subject.

Open source tools, such as EMBOSS, should be supported and their use and utility extended, for example, inclusion of greater functionalities including repositories for data such as high throughput sequencing technologies (e.g. sRNAs). Open-source tools should take priority for funding over closed-source software. A central repository of open-source tools could be considered

The issue of data standards is as important for this consultation as it is for the recent Data Sharing consultation exercise. Projects for defining data standards should be considered and supported.

Biological resources and integration mechanisms requiring support

Biological resources are of high strategic relevance to fulfilling the mission of JIC. These biological resources contribute significantly to the science at JIC and the wider UK community. JIC commits significant CSG funds to staffing and maintaining these resources some of which have high strategic UK significance such as the BBSRC small Grain Cereal collections which are public open collections and serve a wide range of external user groups. The funding for the maintenance and development of these resources at an Institutional level has been a difficult issue for many years. The development of a future strategy for these collections requires extensive integration with wider resources such as DNA marker, maps, other omic platforms and bioinformatics as well as more phenotyping of the material for target traits. The rapid development of comparative biology is not perhaps best served by the existing network and organisation of current resources which have generally evolved in response to former requirements. Novel ways of bringing resources together either physically or virtually into more appropriate alignment to underpin the current and future needs of UK Biosciences could significantly improve strategic relevance of these resources. This view is supported by the inclusion of resources into a number of proposals entered into the recent BBSRC Crop Science Initiative. JIC welcomes the initiative by the Tools and Resources Strategy Panel to address the

sustainability of these resources to fulfilling BBSRC's mission statement and in ensuring they remain fit for purpose and deliver to the UK Bioscience community and beyond. The JIC is particularly pleased at the recognition being given to a data sharing policy as the integration of biological and bioinformatic resources should be priority area for support.

Plant genetic resources required by the UK bioscience community

Key resources

- Comprehensive reference collections of UK crop germplasm for all crop groups, cereals, legumes brassica etc.
- Diverse germplasm of crops and wild relatives from primary and secondary centres of diversity and all levels of the genepool.
- Typed mutant reference collections for crop and model species
- Mutant and TILLING populations for reverse and forward genetics
- Precise genetic stocks including aneuploids, alien introductions, intervarietal substitutions, recombinant inbred and doubled haploid mapping populations with relevant maps and markers

Microbial resources required by the UK bioscience community

JIC has submitted comments to the Microbial Sciences consultation already. For the sake of inclusivity we also state here that little thought appears to have been given to post-genomic resources (e.g. mutant collections) and their distribution. Culture collections should be Government-funded for long-term stability. Maybe more efficiently organised as a central resource into which individual research laboratories feed their collections and remain as sources of expert advice. Charges imposed (with differential between academia and industry).

Genomics resources required by the UK bioscience community

JIC currently holds a substantial collection of BAC libraries of crops and other organisms. These are of considerable international importance. We recover the direct cost of distributing materials but sustain the indirect costs. Support that ensured the full recovery of costs of maintaining and distributing resources such as this would be beneficial to the whole science community and could provide even greater benefit if underpinning funds enabled cost-effective expansion of the resources to cover new collections on behalf of this community. Cost-effective means of maintaining these resources in a secure and stable manner require development and investment to secure their long-term future.

Plant resources currently missing

The *Medicago* research community has highlighted the urgent need for co-ordination of effort and resources in the development of mutant collections and diverse lines to compliment existing resources and to act as the long term repository for mutants arising from the large scale US and European reverse genetics platforms. Such a resource stock centre for *Medicago* would be most effective if coordinated with similar developments in other UK crop legumes such as pea and clover.

Resources relevant for this funding mechanism

Bioinformatics

JIC maintains various databases, such as ATIDB and SCODB, as well as generic databases and analysis systems for semantic data integration, such as GERMINATE and COMPARAGRID; these are suitable for funding through this mechanism.

Genomics

JIC maintains genomic resources encompassing fourteen libraries that include Arabidopsis, Brassica, wheat and Brachypodium. These are distributed by the GeTCID facility.

Plants

JIC maintains and is developing a number of key germplasm resources that underpin a range of science at JIC and the UK biosciences community that are directly relevant for consideration under this funding stream. These include the **BBSRC** small grain cereal collections, wheat precise genetic stocks and an internationally important collection of Pisum. The latter includes the international reference collection for classical mutants and key public mapping populations. Additional support for these collections is required in the areas of phenotyping for traits of interest, completing DNA banks and associated seed stocks and further development of databases to ensure that the large scale data sets being generated on the collections are accessible.

Further resources developed and publicly available from the JIC or under development include the following;

Lotus japonicus - JIC has developed and hosts a public platform technology for TILLING in Lotus and a mutant collection jointly constructed with the Sainsbury Centre and **BBSRC** support. The database for this collection is currently maintained by the LMU in Munich in a web accessible format (http://www.zopra.de/).

Pisum sativum - Mutagenesis programmes currently running in association with Defra Pulse Crop Improvement Network and EU Grain Legume Integrated Project. These resources are being developed and maintained within the Genetic Resources Unit at the JIC.

Brachypodium sylviticum - Being developed as a model cereal to bridge the gap between rice and wheat. **BBSRC** has already contributed to a publicly available BAC library. Other resources being developed are an ecotype collection and mapping populations.

Brassica species- A TILLING population is being developed in the diploid *Brassica rapa* species, which will be available to the community through collaborations or for a fixed fee. Seeds from 3,000 M1 families are expected to be generated in the first round. In addition we expect to develop a Fast Neutron collection in *B. rapa* over the coming years comprising seeds from >25,000 M1 families.

How the funding should be used

The bullet points outlined in 1.a of the consultation document should be used as the key criteria and decisions should be based on merit, not history. This will includes resources previously supported by BBSRC funding (2 a.) as well as identified resources that are currently missing or under development where the UK has the necessary expertise and need to be involved in the development of new resources such as reference mutation collection for the model legume *Medicago truncatula*. Existing resources also require further development (2 b.). Evolution and

improvement of current resources should be preferred over the invention of completely new approaches unless the latter is necessary for serving a particular need. The rapid developments in high throughput and omic technologies are resulting in larger and more varied data sets being generated. It is vital that these data sets remain linked to the resources and are made availability to the research and other user communities in such as way as to enable cross species and genera comparative analyses to be made. The curation and analysis of these resources highlight the important interface between biological and bioinformatic resources and data sharing which should be a high priority for this new funding mechanism. For example, existing resources should be brought together in new ways, such as by allowing data integration by the installation of effective APIs to federate such resources.

There are potentially real gains to be made by looking at the synergies of co-ordination of efforts between resources where there is a high degree of complementarity such as the legume resource. This could be developed on a decentralised model whereby the necessary funds were provided to enable common database systems and web services to be constructed, deployed and maintained.

The case for funding static collections should be based, as for the others, on high strategic relevance, and they should continue to receive a low level of support if they have proven uses. If additional funding facilitated further development and added functionality of such resources to improve their relevance then this should be considered. It is hard to envisage static collections in the current climate with the exception of certain reference collections.

Public accessibility should be a prerequisite for eligibility for funding of resources. Opening up access and raising awareness should be more the onus of the host institution and a low priority for this funding mechanism.

Key components of sustainability

Trained management, continuity and experience are key components of sustainability but germplasm resources are by their very nature a long term commitment and the duration of funding is THE key component. Funding at local level can lead to competition with other demands that, over the long term fail to be sufficient to maintain the level of development of resources that is required. Central funding for such resources may help alleviate this problem if it is maintained at a sufficient level. This would naturally involve input into the strategy and priorities from a wider representation of users from the bioscience community possibly in the form of steering groups to ensure the resources are managed in a dynamic manner and remain fit for purpose.

A rationale, strategy and road maps are important, and clear objectives are helpful in assessing performance and monitoring strategic relevance. Sustainability should be achieved by a rolling programme of assessment of supported resource of this kind.

While monitoring and assessment are important, this should not be just about numbers in and number out. Quality assessments of utilisation such as citations, publications and involvement in collaborative projects are additional and often better criteria for assessing utilisation. While a focus on support and delivery into UK biosciences is important, it is important to extend the criteria to include involvement with European and international networks of such resources where co-funding may be possible.

Location can be an important factor in ensuring close integration and awareness of key user groups as well as synergy of effort in terms of staffing and physical resources, storage, glasshouses and field costs. Critical mass (e.g. Wellcome Genome Campus) can provide the synergy required to maintain databases (such as EMBL) and centrally provided tools. However many resources can only be maintained efficiently with local biological knowledge and are best distributed to where the 'wet' science is done.

Where appropriate, good virtual networks should be considered, in order to bring complementary resources into closer alignment.

In general, funding would no longer be required in the event of a resource falling out of use, or duplication. Cost recovery would militate against freedom of availability. Consideration must be given to the long term implications of withdrawing support if this would mean that a potentially useful resource would be impossible or difficult to reproduce if required in the future.

Dr Steve Rawsthorne, John Innes Centre
Posted 3 March

Institute for Animal Health

A distinction needs to be made between research centres, such as the IAH, whose role it is to carry out research and produce data, and data centres, such as the EBI or Sanger Institute, whose role is to curate repositories of that research data. There are clear examples where an institution can be both a research centre and a data centre, and the Sanger Institute is a good example.

1. What bioinformatic and biological resources do you think should be supported?

Genomic Sequence resources:

- A 90% sequenced genome is not a finished genome, and nor is a 60Mb genome that exists as 20,355 contigs. Although a genome sequencing project may be costed and timed at *x* years, **it should be recognised that further sequencing and funding may be required to finish the genome properly**.
- A genome sequence is a living thing, it can be updated and improved beyond the end of the project. **Resources and skills for further and better assemblies are essential.**
- Automatic annotation of a genome is the beginningof a successful annotation project, not the end. Long term, functional annotation of any genome involves several years of collaborative work between research centres, which carry out functional studies, and data centres, which curate the data. **Systems for closer integration between research and data centres are required.**
- Eukaryotic organisms with small amounts of evidence available for genome annotation may need further investment to gather evidence. For example, the chicken has roughly a tenth the number of ESTs available for human and mouse, Eimeria a factor of ten lower still. Effective genome annotation with small amounts of evidence is very difficult. **Resource to gather more evidence for poorly annotated species may be required.**

Microarray resources:

- Not everyone uses affymetrix. The ability to integrate microarray data, and link it to resources such as genome annotation, GO, protein interactions networks etc is very important. **Systems to allow the mapping of microarray probes to functional annotation databases are required.**
- Linking of microarray data to pathway data is also important. At my last count, KEGG contained 159 human pathways, 155 mouse pathways, 31 chicken pathways and 12 cow pathways. **Translation of functional information from well annotated species to less well annotated species is required.**
- Ideally with microarray data, you should be able to upload your data (probe information (ie sequence) as well as numerical data) in one place and instantly have access to pathways, interactions networks, ontologies, functional categories etc

Proteomics resources:

- The proteome is not just a translated version of the transcriptome. Proteins form complexes, are modified post translationally, are regulated post transcriptionally, and are activated by mechanisms such as phosphorylation. An increase in RNA abundance does not mean an increase in protein abundance or activity. **These models need to be incorporated into tools and databases.**
- The biological community require standardised tools, databases and skills for handling mass spec data

General resources:

- Proliferation of databases is confusing – applications for new databases should be encouraged to extend existing ones where appropriate. **Standardised database management technologies would be perfect.**

Systems Biology resources:

- Access to integrated database systems
- Access to mathematical modelling expertise. The centres for systems biology **must** seek to collaborate with research centres
- The capability to perform systems biology on **non-model organisms**

Software

- **Commercial software for bioinformatics remains inhibitively expensive. The availability of free software suites such as EMBOSS and bioconductor are god sends, without which much of the work at the IAH would not be possible. It is essential that cutting edge, open source, freely available bioinformatics software is funded and maintained.**

2. How do you think should the funding be used?

Out of the 304 bacterial genome sequences in the public databases, **only 48 have a version number greater than 1.** This does not reflect the amount of knowledge accrued about these genomes, it reflects the failure of that knowledge being curated and deposited in the public repositories.

Research centres do not have the time and resource to ensure that every piece of knowledge (eg about a gene function) is propagated through into the data repositories. Similarly, data centres do not have the resource to read every paper and update the databases themselves. The system of research centres publishing data as PDFs which are then text mined to provide information is inefficient.

Better systems for the transfer of knowledge from research centres into data repositories are required. This should not entail simply throwing more money at the data centres, as the research centres are producing the knowledge on a daily basis and need extra resource to be able to deposit this knowledge into public repositories.

Resource should be centred on existing resources and extending these existing resources, so that we have a few, large, integrated resources that have a common framework, rather than lots of small resources that may use different technologies.

"Should funding be focused on publicly accessible resources only?"

In answer to this specific question, absolutely, 100% yes.

3. What are the key components of *'more sustainable'* support?

Bioinformatics and biological resources should be centred on a few, large centres rather than lots of small ones, such that a common framework for the resources is used and the cost of administration spread out across those resources.

Initial funding should be available for the set up of the resource, then a "background" funding scheme put in place to allow the resource centre to maintain and develop that resource, but not at the same rate as during initial set up.

Resources that are no longer being developed, that are just static, should require a minimal amount of funding that could be incorporated into the FEC calculations of the resource centres.

Dr Geoff Oldham, Institute for Animal Health
Posted 8 March

Institute of Food Research

1. Bioinformatic and biological resources that require long-term support:

- **Databases of biological information** of community-wide interest, which have been generated by publicly funded research, or are otherwise in the public domain. At present, such databases are generally either:

- Major generic resources, sited and maintained at large, permanent organizations (e.g. EBI), **that receive long-term support from the host institution**, and wide usage and recognition in the bioscience community.

- Minor and/or more specialist resources, locally held by individual researchers, research groups or small consortia. Typically these are created in the first instance by short term grants, remain the responsibility of the research teams that generated them, **and have no formal commitment to long-term support.**

Of these, it is the smaller and/or specialist resources that are most likely to suffer from funding uncertainty and discontinuity. There is currently no system for accessing funds to support long-term database infrastructure, and the development of a mechanism for doing so is timely (assuming that this is what is meant by BBSRC's statement *"a fund to provide a more sustainable future for bioinformatic and biological resources"*). This will provide a much needed route for obtaining support for specialist databases which may otherwise be lost to the bioscience community entirely.

In practise, this means that the funds would of necessity be used to support:

- **Physical infrastructure** (including hardware enabling access to large databases from the desktop; high bandwidth connection to database sites; seamless delivery to local servers; high volume storage and backup).

- **Support staff** whose primary role is maintenance and curation of the databases (or other collections of biological resources, software tools, etc).

Research in the biological sciences increasingly depends on access to large, well organized databases and collections of software tools. In addition to the established mainstream resources, we see the importance and usage of specialist databases increasing in the future as user communities agree upon standards for data and metadata exchange, for example in the post-genomic technologies. A priority area for the future must be biodiversity informatics e.g. integrating post-genomic datasets with culture collection databases, thus ensuring the research community has instant access to full comparative data, as well as to properly authenticated cultures.

2. How the funding should be used:

Access to this funding should be based primarily upon merit and subject to competitive peer review, and should be granted only for the maintenance of publicly accessible resources. However this funding route needs to be separate from the usual research grant funding process – it is inappropriate for a proposal for maintenance and curation work to compete directly with a hypothesis driven research project.

The funds should not be restricted to resources established by previous BBSRC funding, but rather each case considered on its own merits. Since the resources concerned are intended to be accessible and of utility across the whole of the bioscience community, the funds should perhaps

arise from a collaboration between a range of national funding bodies (UK Research Councils, DEFRA, DTI) rather than the BBSRC alone. Long-term maintenance of nationally important resources could be regarded as a shared responsibility.

3. Key components of 'more sustainable' support

The duration of resource maintenance grants should not be limited to the "standard" three years; 5 – 10yrs would be more suitable. Within this framework, a programme of regular assessment could be used to monitor criteria such as frequency of usage, relevance to current bioscience research, etc., with the timetabled opportunity to withdraw funding if the resource falls into disuse, or is duplicated/superseded elsewhere.

Operating on a full cost recovery basis would inevitably lead to restricted availability and uptake, and arguably runs counter to the commitment to data sharing and freedom of availability of the results of publicly funded research.

With regard to location and management, it is clear that long-term, institutional commitment means that such sites are well placed to host and maintain large strategically important databases. However many other more specialised resources are best sited alongside local biological knowledge, and these should not be overlooked – it would be costly, difficult or even impossible to reproduce many of these resources if they are lost through lack of support.

Kate Kemsley, Head of Bioinformatics and Statistics,
Institute of Food Research, Norwich
Posted 1 March

University of Cambridge, Professor Sir Tom Blundell

The greatest need is for bright young research fellows to be based in the leading Departments supported by the BBSRC to carry out programmes in biocomputing and bioinformatics. This should not ignore the historical need for sequence and structural comparative analyses but it will need to focus on integrating these more traditional informatics tools with those for the 'omics - ray technologies, proteomics and metabolomics. The individuals should carry out with their own research programmes but they should be collaborative with others in the department. The historical trend has always been to have good computer and informatics scientists in biological departments and this need continues.

There is a continuing need also to encourage collaboration with engineers, physicists and computer scientists but this has its own momentum as they are now keen to engage with biological sciences in order to have research programmes which will keep their own graduates for Ph.D.'s. I feel the need to have people in house has not received sufficient emphasis.

Professor Sir Tom Blundell, University of Cambridge
Posted 1 March

University of Cambridge, Gos Micklem

Having read the currently posted comments I feel that I have nothing further to add. Michael Ashburner is a close colleague and I very much agree with his views and in fact with the other views posted too.

Gos Micklem, University of Cambridge
Posted 1 March

University College London, Christine Orengo

Dr Andrew Martin has already compiled responses on behalf of the Department of Biochemistry and Molecular Biology, UCL, below are some additional comments regarding support for databases.

The United Kingdom has made major impacts in the field of bioinformatics through the establishment of databases that are internationally renowned.
In fact, some of the most widely used databases are based in the UK and have few counterparts in the States or the remainder of Europe. For these resources to remain at the cutting edge and maintain their dominance, they clearly need support. However, it is also vital to nurture smaller less established ventures, which may grow equally in status, in time, and help to maintain the UKs leading role.

Small databases, set up by individual researchers to facilitate a particular type of research, should be funded from specific contributions allocated to the project. These databases may not be regularly updated once the project is finished and the data may only be of interest to a limited section of the biological community. These could perhaps be described as highly specialised or project driven resources.

Where the data is easily accessible to and valuable for a much larger section of the biological community, even where only one group is responsible for maintaining the resource, I would describe these as community based. Deciding at what point a project driven database becomes a community database, deserving more regular and reliable funding, is therefore a key question.

In judging this, the difficulty in deriving the data and the proportion of the community benefiting from the data should both be considered. Some small, project driven databases should be ensured continued support if they are the sole, or only one of a few resources, providing data that is key to understanding an important biological system. However, there should be some transparent mechanism for demonstrating the continuing need of the community for the database.

Whatever the type of resource there should be some obligation on the resource providers to compile meaningful and comprehensive statistics on who is benefiting. Web statistics are an obvious way to do this, but whilst the number of independent sites accessing the data is often presented, this is rarely broken down into useful descriptions of these sites or evidence of independent publications citing the data.

Consideration of other related resources providing similar data should also be compiled and presented so that any funding panel can more easily gage the gain or loss to the community of supporting or refusing the grant.

In addition to seeking endorsements through public consultations, that may largely trigger responses from those with vested interests, the research councils should establish more transparent and standard mechanisms for evaluating resources. For example all database providers could perhaps be obliged to provide the following statistics and information in their grant applications:

- Number of independent sites accessing the data; number of web page accesses
- Number of publications in peer reviewed journals citing the resource (broken down into publications associated with the resource provider and independent publications).
- Number of independent groups supporting the resource (indicated by letters of support or statistics compiled through registration of users).

- Number of other public resources providing links to the resource. Number of collaborations between the resource provider and independent groups, exploiting the resource.
- Information on other resources providing similar data.
- How suitable is the group to provide the data (e.g. possesses the most powerful algorithms for deriving this type of data, has the most comprehensive curation and validation protocols, integrates more types of data and provides more links to other resources than competitors).
- Would it be possible for other groups to easily replicate this data if the resource was no longer funded etc.

Finally, whilst it is important that resources be as automated as possible, it is clear that most algorithms used in deriving data are far from being 100% accurate and therefore it is not surprising that many of the most highly valued and widely used databases employ some degree of manual curation. The amount of manual intervention involved in compiling these resources should also be taken into account, then, as this clearly increases the value of the data and means that it would be harder to recreate the resource if there was a problem with continued funding.

However, despite the value of manual curation, it is extremely difficult, in fact almost impossible, to obtain funding for this type of post.
Funding can often only be obtained for developing novel algorithms and PDRAs employed to do this research are understandably unwilling and in fact should not be expected to be responsible for the regular updating and manual curation of a resource.

In summary, there is clearly a need for sustained long term funding for maintenance of some of the larger community based resources through joint research council funding. Most community databases will require longer term funding (e.g. 5 years at a time) and this should not necessarily be to develop new features but to keep pace with the increasing volumes of the data arising from genomics initiatives. More specialised or project driven databases are perhaps more likely to require shorter term funding and may want to focus more on developing novel features, likely to widen the scope of the database and broaden community interest in the resource.

Christine Orengo, University College London
Posted 28 February

Scottish Association for Marine Science

Bioinformatic and biological resources are amongst the most important underpinning components of the UK's scientific infrastructure. The provision of both bioinformatics data and live starting cultures have become synonymous with the activities of microbial culture collections or, as they are now frequently termed, BRC's (Biological Resource Centres). Microbial strains in collections are *de facto* biological standards. Without biological standards comparative studies are impossible. In biotechnology strains may vary considerably, if these are not preserved and documented then exploitation is at best problematic, or even impossible. We are aware of a company obtaining a "hit" in their screening program, but because the strain was not preserved/ documented it proved impossible to exploit this potential product.

The below submission to the ongoing BBSRC consultation on bioinformatic and biological resources is being submitted on behalf of SAMS (The Scottish Association for Marine Science). This organisation has the advantage of being a user of biological resources and a provider as it hosts the Culture Collection of Algae and Protozoa (CCAP) a constituent collection of the UKNCC. In addition, SAMS is the location of the European Centre for Marine Biotechnology (ECMB), an incubator for biotechnology interlinked with SAMS, which houses 'incubator' biotechnology companies exploiting marine microbial resources.

For simplicity the input is headed as suggested above in the Consultation section.

Should any clarification be required please feel free to contact Dr John Day jgd@sams.ac.uk

1. What bioinformatic and biological resources do you think should be supported?
1. b. What resources do you anticipate will be required for future bioscience developments?

- Well characterised high quality live material.
- Expertise in conservation/preservation, maintenance and culture optimisation
- Bioinformatic data: traditional taxonomic information; molecular data; other bioinformatic data.
- High quality bioinformatic data should be directly linked to stable conserved live material.

1. c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

- High quality bioinformatic data directly linked to stable conserved live material.

1. d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

- It is vital that material is held in designated collections in such a way as to ensure both phenotypic and genotypic stability.

1. e. Are there any resources that are currently missing?

- Large sections of the known biodiversity are currently not available from any BRC. This is because of a variety of factors including: they have not been isolated; they have as yet not been possible to culture or preserve.
- There have been no serious attempts in the UK to conserve *ex situ* threatened algal biodiversity. However, this approach has been successfully employed in Japan to conserve Charophytes for future reintroduction to restored habitats.

1. g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources which you consider could be relevant to this funding mechanism?

- The CCAP and other UKNCC collections are an obvious source of biological resources and ideal platforms for the development of a sustainable biotechnological sector.

2. How do you think should the funding be used ?
2. a. Should emphasis be placed on supporting resources previously established by BBSRC funding?

- Continuous support is absolutely vital to all BRCs.
- The current BRCs are the ideal platform on which one could build new bioinformatic, taxonomic, conservation or biotechnologically orientated projects.

2. b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)?

- Most collections now have some molecular expertise. However, the on-going developments in bioinformatics and genomics mean that they need investment to maximise their value to the biotechnological community.

2. c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)?

- There is a minor role for BBSRC to fund the maintenance of archive collections e.g. of taxa that have been cited in the scientific literature. However, these strains must still be available to the user community.
- If the specimens are not accessible and therefore not usable then BBSRC funding should be withdrawn and the resources relocate to a service collection.

2. d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges?

- Yes, there are many areas where consortia holding different biological resources could work together more effectively than single collections including: taxonomic, bioinformatic and preservation orientated projects.

2. e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence?

- This is vital nationally and internationally.
- One option would be to make submission of strains isolated and/or developed under BBSRC funding to a service collection compulsory. Funding could be top-sliced from any grant and funds allocated to the collection with the deposited material. This would need costing, but for a fee of ~£600 the material could be accessed to a collection and thus be available to the scientific community.

2. f. Should funding be focused on publicly accessible resources only?

- Yes, if the material is to be housed in a public collection it must be available to the user community. The obvious exception is the confidential deposition of material for safe keeping. This is usually held at a commercial rate by BRCs.

3. What are the key components of ' **more sustainable** ' support?
3.a Duration of funding;

- Long-term funding is vital for the conservation of biological resources and the curation and development of both bioinformatic and biological resources.

3. b. Location and management;

- Ideally, BRCs should be located where there are synergistic opportunities with users of biological resources or a centre of excellence. The UKNCC network has developed in this fashion with different cell-types held in the appropriate/relevant institutions around the UK.
- Critical mass is important. The investment by NERC in recombining the CCAP at SAMS is already starting to show added value and increased scientific productivity.

3.c. Monitoring and assessment;

- It is important that there are periodic assessments of any collections function and value to the scientific community. It is inevitable that different performance criteria are relevant to different collections. At CCAP the collection is reviewed as a component of SAMS at regular Science and Management Audits.

3. d. When would you anticipate that funding is no longer required (e.g. lack of use, operating on cost recovery basis; resource no longer of high strategic relevance to current UK bioscience research)?

- If quality of any collection falls below an acceptable level and the material is not being requested by the user community, then funding should be withheld until an action plan is initiated, or withdrawn if there are no plans for improvement.
- It is probable that some level of public funding will always be needed by BRCs. No BRC anywhere in the world is self-sustaining on culture provision charges and routine services.
- Improved in-house exploitation of the collections would generate additional funds that would partially cover costs.
- Another suggestion is that there could be a fee associated with culture deposition. This could help fund collections, but this would need to be covered from the original research grant funding the research. If implemented, this could become a significant income-stream if the deposition of cultures became an absolute prerequisite for acceptance for publication in any mainstream journal.

Dr John G Day, Scottish Association for Marine Science
Posted 28 February

Institute of Grassland and Environmental Research

Previous contributions to this discussion have addressed all the main issues we'd want to comment on. We particularly support the points made by:

- Andrew Martin (need for capability to download, not just search, data)
- Andy Law (extent of interfacing with other resources as an important criterion for support)
- Pete Downes (better chemical database resources)
- Colin Semple (requirement for some of the services formerly provided by HGMP-RC)
- and all the many, especially Keith Edwards and Michael Ashburner and Andy Law, who emphasised the crucial importance of a proper mechanism to supoort continued maintenance and development of databases for where there is proven usage and need

Helen Ougham,
Institute of Grassland and Environmental Research (IGER)
Posted 28 February

University of Aberdeen

What bioinformatic and biological resources do you think should be supported?

Support should include funding for microbial culture collections to enable them to continue their current essential role in support microbiology research programmes and to develop that role to achieve new requirements of BBSRC-based programmes.

Points to consider here include:

a. It is proposed that funding:

should focus on resources critical to UK bioscience research and without which the community could not achieve its full potential.

Microbial culture collections are essential for fundamental microbial research and for commercial applications of microorganisms. They provide long-term, quality-assured preservation of microbial resources and professional and reliable distribution of these resources. There are two major alternatives, which we believe are not acceptable. The first is reliance on foreign collections. However, BBSRC and the microbial research community cannot guarantee continuation or quality of this supply; new regulations (in particular material transfer agreements) significantly restrict supply; costs cannot be controlled (ATCC cultures are now several-fold more expensive than UK cultures); and movement across national boundaries raises problems. The second is a 'grey market' in cultures, with research groups exchanging cultures on an *ad hoc* basis. BBSRC again lacks control over sustainability of supply; quality assurance is non-existent, authenticity and culture identity are often questionable; there is no guarantee of equitable supply; and the resources required for preservation and supply are not cost-effective or fully understood. Examples can be provided of situations in which reliance on these forms of supply have led to embarrassing and serious wastage of research council resources.

Collections are essential for academic research groups and for industry. NCIMB, for example, preserves and provides cultures which are used widely in the pharmaceutical, food, bioremediation and environmental industries and provides quality-assured and secure preservation and storage of bacteria required for microbial research (e.g. type strains, novel isolates) and for commercial applications of bacteria (e.g. safe and patent deposits).

The current explosion of interest in microbial diversity, resulting from application of molecular techniques to characterise natural microbial communities, will significantly increase the number of novel isolates of potential scientific and commercial importance. Newly classified strains must be deposited in at least two accredited collections and demand for quality-assured preservation of industrial strains will increase.

The expertise within culture collections maintains traditional microbiological skills, which are becoming increasingly rare in academia, but also involves molecular techniques for microbial characterisation. This leads to additional services to the microbiological research community, notably identification services.

should focus on resources which are / would be well used by a broad UK community.

In the year ending July 2005, UK sales of bacterial cultures by NCIMB to academic and industrial users have were 251 and 157, respectively, an additional 661 sales to industry of QC strips and 32 patent and safe deposits. There were also 99 new accessions to the collection, demonstrating significant and continuing demand for these resources.

should underpin the delivery of BBSRC's strategic goals. For example, does systems biology present any new resource needs? If so, what are they?

Microbial culture collections underpin BBSRC strategic goals in microbial research and in biotechnology.

b. What resources do you anticipate will be required for future bioscience developments?

Collection services will expand through the increasing requirement for quality control and preservation of microbiological material used in and generated by genomic and post-genomic programmes, and the increasing realisation of the enormous number and diversity of microorganisms that have not yet been cultivated. Currently archiving and preservation of material arising from the considerable investment in genomic programmes (e.g. clone libraries, knock-out mutants) is not coordinated and generally left to networks of individual researchers on short-term contracts. Archiving of the 'microbiological' material arising from such programmes is arguably more important than archiving of sequence data.

c.What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

Functional and regularly updated web-based microbial databases which can be easily and remotely accessed and searched by the scientific community and be updated by the respective culture collections.

d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

A long term and reliable quality assured source of living microbial reference cultures provided by a co-ordinated national collection network. It is the paramount function of a national public collection of reference cultures to acquire, preserve, maintain and deliver quality products at affordable prices.

e. Are there any resources that are currently missing?

f. Are there any resources that it would be inappropriate for BBSRC to support, for example for ethical or societal reasons?

g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources which you consider could be relevant to this funding mechanism?

How do you think should the funding be used?

Points to consider here include:

a. Should emphasis be placed on supporting resources previously established by BBSRC funding?

Funding for collection activities, currently funded through BBSRC, should be continued. However, resources should be increased to address new requirements of BBSRC research programmes and developed in collaboration with activities and funding of other collections, to provide common strategies, policies and targets and to enable coordination of future requirements.

Funding would ideally be on a rolling contract, with regular reviews of progress. This would facilitate future planning, recruitment, training and retention of staff and development of new

services. Collection staff are highly specialised and require many years of training to develop the necessary expertise to maintain, preserve and distribute an increasingly diverse range of organisms. Such staff are frequently being lost due to poor remuneration, budgetary constraints due to diminishing core funding (in real terms) or uncertainty over the future of individual collections. Such losses can not be sustained if collection quality and integrity are to be maintained.

b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality?)

c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)?

As indicated above, there is a requirement for maintenance of microorganisms currently in collections. Demand for organisms within collections is high and justifies the considerable investment, including that from BBSRC research funding, in their original isolation and characterisation.

d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges?

See above regarding greater coordination of funding mechanisms for microbial collections, in addition to those funded by BBSRC.

e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence?

A vehicle for this already exists in the form of the UKNCC. This forum was set up in the 1990's to co-ordinate and centralise the activities and cultures available form specialist UK national collections, and to provide a better and more responsive user interface. However, its existence is under threat as under investment has stunted its potential. Further investment is required to revitalise this initiative.

f. Should funding be focused on publicly accessible resources only?

What are the key components of '*more sustainable*' support?

See above

Points to consider here include:

a. Duration of funding;
b. Location and management;
c. Monitoring and assessment; and
d. When would you anticipate that funding is no longer required (e.g. lack of use, operating on cost recovery basis; resource no longer of high strategic relevance to current UK bioscience research)?

Other issues

Looking to the future and beyond the remit of this activity, are there any other resources (out with bioinformatic and biological resources) that could be important for future UK bioscience research?

The University of Manchester, Norman Paton

Comments are made below in response to several of the questions listed, with comments restricted to bioinformatic rather than biological resources.

1. What bioinformatic and biological resources do you think should be supported.

a. Does systems biology present any new resource needs. Yes - most new experimental and computational methods will give rise to new resource needs, and systems biology, like functional genomics, for example, will give rise to new data resources and analysis tools, both of which may be candidates for long term support. The data resources may be data-type specific (e.g. for storing the results of novel experimental techniques, or for storing models) or community-specific (e.g. for bringing together diverse data sets on fungi).

b. What resources do you anticipate will be required? Informatics resources that will continue to develop in future will include: primary data repositories (for experimental results), secondary data repositories (for derived results - e.g. InterPro, Pfam, etc) and software tools. The latter may be specific to a single kind of data, or support integration of some form or other. There should be support both for research into the development of new databases, techniques and tools, and for maintaining successful examples longer term. These different categories of activity (research, development and maintenance) must be able to be assessed according to different criteria. Many of these will cut across the remits of individual research councils.

c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority? There are too many potentially relevant resources to make the provision of a list straightforward. Useful areas for investment, though, might include public data resources that would be useful destinations for data produced in UK labs, or that support communities with a strong UK presence.

e. Are there any resources that are currently missing? Sure. For example, there are not yet effective tools, databases or agreed standards for much of proteomics or metabolomics.

2. How do you think the funding should be used?

a. Should emphasis be placed on supporting resources previously established by BBSRC funding? There is something to be said for continuity and stability for certain forms of tools and resources. As such, one might anticipate a pattern in which funding bodies specialise on particular requirements and communities that are seen as representing good local practice. Any "emphasis" should, of course, be easily overruled within the peer review process by concerns about quality, value for money, etc.

b. What emphasis should be placed on further development of existing resources? This will often represent a sensible way to proceed. New functionalities will have to keep being added to existing resources to reflect changing requirements.

c. Should the BBSRC support resources that are being maintained but not developed. There should be no hard and fast rules on this sort of thing.

d. Is there a need for existing resources to be brought together in new ways? Sure. One can expect there to be benefit from loads of different forms of integration. For example, consistent description of samples and their processing in multiple kinds of omics experiment would provide support for integrative analyses that otherwise might be extremely fiddly. Furthermore, consistent programmatic access to data resources from web services would be helpful, as would consistent ways of advertising descriptions of what the services do.

f. Should funding be on publicly accessible resources only? As in (c) there should probably be no hard and fast rules. For example, a site should be able to make a case for support for resources that capture, manage and integrate data at that site. If they follow the emerging data sharing policy, then at least some of the data is likely to be made public in due course. However, good science increasingly requires good informatics, so sites should be able to make a case that they require custom software to support their science. Making an infrastructure widely useable by others (e.g. as open source software) will cost more than a purely local solution, due to the need for improved software engineering processes, documentation, multi-platform testing, etc. Sites should expect to be given credit for producing software and systems that others can use, and should expect reviewers to be positive about proposals that involve widely useful resources, but there should be no blanket rule, or tokenism will result (e.g. by sites releasing software that is impossible to use in practice).

3. What are the key components of more sustainable funding?

a. Duration. The length of individual grants probably is not the issue; the issue is the recognition that long term support of a valuable resource may be a suitable target for research council funds.

Norman Paton, University of Manchester
Posted 28 February

University of Cambridge, Dr Paul Schofield

The following response represents a personal view of the three authors who have been involved in bioinformatics and database development both inside and outside the UK for many years. We hope that this brief summary of our views is helpful to the BBSRC in formulating its policy on funding of the resources we need in the UK to remain competitive.

1. What bioinformatic and biological resources do you think should be supported?

We are primarily concerned with Resources important to functional genomics and the generation of mouse models of human disease. These fall into three broad categories:

- Biological resources; animals, DNA, cells etc
- Biological information, such as databases of gene expression, phenotype data, annotated genomic sequences
- Analytical and curatorial software

These types of resource are widely used by members of the biological sciences community as well as those engaged in functional genomics and include systems biologists, clinical researchers and biophysicists, and represent the infrastructure required for modern biology in the 21st century.

In addition to being resources which make available to the community data and reagents/animals, databases are tools for the dissemination of important data from their originating laboratories. This in itself provides greatly added value to the funding of experiments in these laboratories and for example in animal research reduces duplication, allows refinement of experimental hypotheses *in silico* before carrying them out, and facilitates the three Rs in laboratory animal research.

UK research currently relies heavily on biological resources funded from abroad. Some of these reside outside the UK completely such as the Mouse Genome Informatics (MGI) database (and are currently accessed for free), and some are funded within the UK but with foreign or international funds. It is notable that the major databases and informatics tools on which we rely as a community are predominantly funded by the European Commission or EMBL and by the US National Institutes of Health. Examples of this are the ENSEMBL database, the Gene Ontology Consortium ontology tools, and the Gene Expression Database GXD. This is striking because the expertise behind many of these globally important resources resides in the UK yet funding has not been available on the required scale or stability.

Sustainability of many UK-based databases is significantly compromised because the European Commission will not fund databases in isolation from current research projects as a matter of policy and consequently although there are a large number of important databases currently located in the UK, including

- Pathbase
- Eurogene
- Eurexpress
- European Bioinformatics Institute (EBI) associated databases, such as ENSEMBL, Array Express

which lack a stable funding platform.

What resources do you anticipate will be required for future developments?

The advent of systems biology and the study of regulatory networks, particularly regarding the understanding of disease processes, requires the public availability of detailed and reliable biological information in an integrated network, together with the tools to use that information for the generation of hypotheses and discovery. Particular needs may lie in regulatory interactions between genes and in mathematical or computational models underlying the understanding of systems.

In addition, there are a number of other resources which underpin database design and interoperability for which funding to provide resources is currently not available. An area of particular UK strength is the design of ontologies for the representation of different kinds of functional genomics data.

2. How do you think should the funding be used?

Should emphasis be placed on supporting resources previously established by BBSRC funding?

We do not believe such a bias to be appropriate. Rather, the most valuable resources should receive priority.

What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)?

Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges?

We believe two aspects to be particularly important in this respect:

1. Curation. The most successful and well-respected databases are often curated rather than relying on automatic annotation. However curation is frequently underfunded when setting up databases and funding for ongoing curation of existing databases is very difficult to obtain except in the few well-funded centres such as EBI. Funding support for curation is therefore needed, as is funding for curation tools to reduce costs associated with human curation where this is possible.

2. New functionality, better interoperability and broadening of data types represented in databases is highly desirable; for example the problem of distributing datasets from MRI and micro CT imaging data on model organisms. These issues are being actively addressed in the NCI Cancer bioinformatics grid (CaBIG) project which has recently committed $60M over three years into the integration of clinical and basic biological datasets concerned with cancer.

Should the BBSRC support static resources?

This is an interesting question and depends very much on the static resource and why it has become static. Databases may become static through lack of stable funding and remain important, they may become genuinely uninteresting. In the first category there are databases generated for particular projects both in the public domain and in industry where with the interest in the project has ceased or funding issues have forced closure. This kind of legacy database presents rather special problems. There would be an argument for a database 'deep freeze' especially if the database contained primary measurements from experiments which would be difficult or impossible to repeat. For example the large scale radiobiology experiments carried out over the 1960s and 70s which may be usefully reexamined in the light of future hypotheses and analytical tools. Pharma Industry databases may be a very valuable source of data which groups could use if there company has ceased to have a proprietary interest. Again by putting such data into the public domain the BBSRC would be generating great added value. There may be the interesting possibility of joint BBSRC and company funding to put such databases into the public

domain, which would be advantageous for all parties concerned. Recently the database and mutant mouse collection from Deltagen has been bought by the NIH and its database hosted by the Jackson Laboratory.

However if such 'saved 'databases are not guaranteed to be made available at least in the medium term then people will be reluctant to rely on them

Should BBSRC aim to open up access to resources and / or to raise awareness to their existence?

The answer must inevitably be yes, With respect to the comments about database integration above then a central portal at which all BBSRC funded databases would be accessed and where possible queried would both increase the use and usefulness of such resources.

From time to time, funding constraints force database resources to impose charges, often for a higher quality product while making a less comprehensive version available free to the academic community. Recent examples include Swissprot, the yeast database YPD and the Transfac database of transcription factor binding sites. Such trends place academic researchers at a significant disadvantage. Additionally, some companies have specialised in providing enhanced or novel information commercially - the Celera genome database is a good example of where some had access and some not. Provision of national funding for subscription to the most central foreign and private databases would make a major contribution to UK research competitiveness.

It is also likely that many large projects which have internal databases have not been able to make those databases available publicly because of the lack of funding instruments available within the UK to do so. This is a missed opportunity.

3. What are the key components of 'more sustainable' support?

At a time when systems biology and the detailed integration of data from a wide variety of sources, from protein structure to behavioural phenotyping, is at the cutting edge of the biological sciences we find ourselves unable to share data efficiently through databases and unable to guarantee the persistence of primary data in data bases and resources central to this new and multidisciplinary systems approach.

All databases by definition are international and, as stated above, the UK not only relies on databases in other countries but currently provides a huge input into the international effort. The UK is a leader in bioinformatics and punches greatly above its weight on the international stage. What is unfortunate is that although we are a leader in this field the funding we can obtain by and large comes from outside the UK. This is in marked contrast to the USA where for example the NCI and NIH will commit to rolling funding for the important primary databases and there is a degree of stability dependent on use and importance. The temptation seems to have been in the past that as databases are international then important databases are 'someone else's responsibility'. We could retain this attitude but gradually important databases and resources would become funded from other non-UK sources, scientists would be attracted abroad as has happened already, and UK scientists would lose the advantage of being able to disseminate their own data through databases tailored to that aim. The possibility of developing models for processes and systems would be made more difficult if the data on which those developments depend is in a database out of their control; for example not structured in the right way or unable to be queried in the right way.

The UK is missing long-term commitment to the provision of material resources with the closing of the HGMP, which gave UK science a competitive edge by providing a sophisticated bioinformatics infrastructure even to small laboratories. Some public databases have also been lost in recent years, such as Festing's inbred strains of mice and rats databases. The UK is

missing funding for the stability of important databases; database funding particularly needs to be made available to support the ongoing curation effort. This is a human activity which is vital to the accuracy and therefore the value of biological databases and requires curators. It is notable that the best and most used databases are those where human expert curation plays a major role.

The UK also lacks a central integration of the resources that exist. One thing which would be extremely cost effective would be the funding of a database integration group whose task would be to advise on best practice, and provide support for standardisation and co-ordination of database efforts. For databases to generate their best value interoperability and semantic consistency are vital. This has been well recognised by many biological sciences communities and the semantic web community as well. Such a centre could also provide access to expertise on programming tools, expert interfaces, interface design, tools for integration and semantic tools such as ontologies.

The UK is highly active in the area of ontologies and description frameworks, which are vital to bioinformatics. These require the bringing together of expert groups and need continuous curation. Funding for such expert curation is needed for the ongoing interdisciplinary meetings required to bring together the domain experts necessary to structure and develop such tools. This would be a rather efficient way of ensuring that the UK retains its preeminent position and is in a position to set standards for the rest of the world for the biological semantic web.

In summary we believe that this is a critical time for bioinformatics resources as we face a data deluge coming from the rapidly progressing large scale mutagenesis projects in the mouse. Without a firm bioinformatics infrastructure the UK will not be able to make the best use of data coming from these programmes, and reliance on funding and resources from outside the UK risks us losing our current strong position in functional genomics and bioinformatics in the medium to long term.

- Dr John Hancock, MRC Mammalian Genetics Unit, Harwell.
- Dr Paul Schofield, Dept of Physiology, Development and Neuroscience, University of Cambridge.
- Dr Duncan Davidson, MRC Human Genetics Unit, Edinburgh.

Dr Paul Schofield, University of Cambridge
Posted 28 February

Imperial College London

I respond on behalf of the following organisations at Imperial College: The Centre for Bioinformatics (of which I am the Director); the BBSRC/EPSRC Centre for Integrative Systems Biology at Imperial College (head Prof Douglas Young); and the Division of Molecular Biosciences (head Prof Paul Freemont). This response has been informed via my role as chair of the BBSRC Bioinformatics Coordinating Group,

Databases

We totally endorse the central role of bioinformatics databases to underpin many areas of biological research that has been highlighted by other submissions. There is a pressing need for support for the development of these resources across the entire range of developers, from the individual groups through to institutions. As we move into integrative systems biology, there will be an increased need for these resources to be maintained and robust links developed to related biological databases improved. More facilities will need to be provided for batch searching rather than the one-off enquires.

Other bioinformatics and computational resources

In addition to databases, the community also makes substantial use of programs, often available via web servers. These face the same problems of maintenance as databases. As we move into systems biology, in silico models will also become a resource that increasingly will be important both for other theoreticians and the wider community of experimental biologists.

Funding mechanism

Support for database and software resources has persistently proved difficult to obtain from research funding bodies. Applications that list the next sensible steps to develop the resource generally are rated as sound but do not compete with the more innovative research topics. Support for these resources must be handled by a different mechanism to standard research grants. Funding must be flexible, but the emphasis should be on a combination of maintenance coupled with sound next generation development. There should be separate funding streams for the innovative development of these resources on the lines of the current Tools and Resources funding initiative. The BBSRC should be able to support database and resources that are relevant to its remit irrespective of where previous funding came from: but it should have the opportunity to focus its resources on maintaining those resources it has previously funded. Funding should be for up to five years with flexibility as to the type of staff that can be employed. Projects that receive substantial funding require an external scientific review committee. Whilst many major bioinformatics databases meet the needs of many users, there remain some well funded databases that remain slow to respond to actual needs.

Professor Michael Sternberg, Imperial College London
Posted 28 February

Sainsbury Laboratory

David Studholme, February 2006.

As the Sainsbury Laboratory's Bioinformatics Manager, I present this response on bioinformatics issues on behalf of the Sainsbury Laboratory.

1. What bioinformatics and biological resources do you think should be supported.

There are three main types of bioinformatics resources to consider here:

- Public databases and webservers
- Bioinformatics software
- Bioinformatics personnel

Many of the larger established databases and webservers (e.g. those hosted at Wellcome Trust Sanger Institute and the European Bioinformatics Institute) have other sources of funding, and it would be inappropriate for BBSRC to fund these. However, many laboratory-based studies generate large datasets of more specialized interest (e.g. organism-specific ESTs, small RNAs, proteomics data). It often is not a trivial task to make these datasets available and to provide tools for the biologist to query the data; BBSRC should be prepared to support this component of any research project.

There are several core bioinformatics software tools that are widely used and re-used as components of analysis pipelines (e.g. the EMBOSS suite of sequence analysis programs). It is important that funding is provided for these resources, both for their intrinsic usefulness, and also to support a group of high;y qualified personnel. It is equally important that any software funded by BBSRC **should be released under an Open Source license**. Not only will this ensure that these publicly-funded resources are freely available to the whole community, but also it will enable and encourage other developers to contribute their time and expertise to augment the efforts of the core developers. **BBSRC should not support development of proprietary software**.

The third type of resource, is skilled personnel. In the Sainsbury Laboratory, we have found it extremely helpful to have dedicated bioinformatics support staff embedded in the laboratory. For optimal progress towards BBSRC's vision of predictive and quantitative bioscience, it is essential that biological sciences departments in all the universities and Institutes have access to effective bioinformatics support. There needs to be mechanisms for funding such staff. Therefore, **BBSRC should be sympathetic to requests for bioinformatics technical support**, for example as a component of a Research Grant application, that would enable departments to provide such support.

Aside from financial support, another potentially limiting factor is training. To realize BBSRC's vision, we will need large numbers of skilled bioinformaticians who can successfully bridge the gap between computing sciences and biology. There is a perception that a low proportion of the current crop of bioinformatics MSc-level graduates are ready to 'hit the ground running' in a biological research environment. As just one example, many UK Bioinformatics MSc courses do not cover the PERL (and BioPERL) scripting language, which is ubiquitously used by computational biologists. BBSRC should **consider developing or facilitating a "core curriculum"** of indispensable components of bioinformatics courses.

2. How should funding be used?

As a general principle, decisions should be based on merit of proposal rather than history.

In general, evolution and improvement of existing resources should be preferred over invention of completely new approaches. For optimal progress, there should be **an emphasis on integrating existing quality resources** rather than duplication and proliferation of disparate resources.

Software and databases should adhere to accepted standards wherever possible (e.g. standard file formats and ontologies, and compliance with http://www.w3.org/ for web applications).

Again, funding should be directed at Open and freely-available publicly-accessible resources wherever possible.

3. What are the key components of more sustainable support?

Although a critical mass of bioinformaticians (e.g. the Wellcome Trust Genome Campus) can provide an ideal home for wide-ranging databases and webservers, in some cases greater synergy can be attained by locating the informatics resource close to the relevant 'wet'science. For example, the ScoDB Streptomyces database is ideally located at Norwich, close to a large Streptomyces research community.

How has the Sainsbury Laboratory addressed its Bioinformatics requirements?

Like other biological research institutes, The Sainsbury Laboratory has found that its science has become increasingly reliant on computational technologies especially to support ambitious high-throughput post-genomic studies. To address this, since July 2004, the SL has employed an experienced bioinformatician, as well as a systems administrator, and is continuing to build up its computer hardware infrastructure.

As a result of this, the SL has access to the following expertises resources that many labs do not:

1. Expertise in automating repetitive and/or complex analyses.
2. Expertise in designing, building, and administering databases and tools for interacting with databases.
3. Expertise in web development and server-side tools for providing custom analysis tools.
4. High performance Linux workstations suitable for developing and using bioinformatics applications.
5. High performance clusters/servers for large 'number-crunching' tasks such as BLAST/FASTA searches or running simulations.
6. Expertise in systems administration to configure and maintain these workstations and servers.
7. Expertise in using existing bioinformatics tools.

A little more detail on each of these:

1. Expertise in automating repetitive and/or complex analyses.

With increasing emphasis on high-throughput technologies, biologists are often faced with repetitive tasks such as having to remove vector sequences, BLAST a large number of sequences against a custom sequence database, predict their secondary structures and then generate a summary table of the results. Expertise in a scripting language such as PERL or Python are required to set up analysis pipelines such as this, and would also allow setting-up of more complex pipelines. Also, expertise in a professional statistical analysis environment like R is extremely helpful.

2. Expertise in designing, building, and administrating databases and tools for interacting with databases.

The data from high throughput proteomics studies, large sequencing projects etc need to be organised and stored. Moreover, it is essential that the biologist can query and mine the data to address biological questions. Therefore biologists can benefit from technologies such as relational databases. The combination of a database engine (such as MySQL) with a scripting language (like PERL, or PHP) plays an integral role in such projects.

3. Expertise in web development and server-side tools for providing custom analysis tools.

In addition to implementing and developing algorithms and pipelines, a bioinformatician often needs to make these tools available to the bench scientist via a friendly user interface. One solution that we have found useful is to provide a web-based interface to our custom databases and analysis tools. This requires expertise in server-side scripting (e.g. PHP and PERL CGI scripts), webserver administration, and client-side technologies (such as DHTML and JavaScript).

4. High performance Linux workstations suitable for developing and using bioinformatics applications.

Our experience is that it is extremely useful for the bioinformatician to have a reasonably powerful workstation on which to develop and test scripts and applications before deploying them on servers. (We currently use dual core AMD64 processor workstations with fast SATA disks and 2Gb RAM. We use Debian GNU/Linux as our preferred operating system.) Almost all of the commonly useful bioinformatics-, development-, and statistics software that we use is freely available and open source.

5. High performance clusters/servers for large 'number-crunching' tasks such as BLAST/FASTA searches or running simulations.

Inevitable some computational tasks are too large to run on a single workstation. For example we sometimes need to run tens of thousands of FastA searches against complete genome sequences, which would take many days or weeks on a single processor. For these tasks it is essential to have access to a multi-CPU cluster. It is also essential to have a reasonably powerful web-server to host our custom web-based applications.

6. Expertise in systems administration to configure and maintain these workstations and servers.

Inevitably it requires manpower and expertise to keep the hardware and software and databases functioning and up-to-date. Whereas many labs may have limited IT support, we have found it extremely helpful to have in-house expertise in systems administration especially in networking and Unix/Linux systems administration working closely with our bioinformatics staff.

7. Expertise in using bioinformatics existing tools.

There are numerous useful (and some not so useful) bioinformatics available to the biologist; however, it is often not straightforward for the bench scientist to get the most out of these tools. The help of a bioinformatician available, who has experience of installing, configuring, and using a wide range of bioinformatics software, saves significant time and enhances the outcome.

WE recommend that when proposed work requires one or more of these capacities, that committees **look favorably on the computational support component of the proposal**.

David Studholme , Sainsbury Laboratory
Posted 28 February

Rothamsted Research

Preamble

This response represents collective inputs from three grouping of bioscientists in Rothamsted Research and should be considered alongside the associated responses that have been to the BBSRC consultations on Future Directions in MicrobialScience and Data Sharing.

The three groups of RRes scientists are:

Plant-pathogen Interactions Division - plant pathologists and microbiologists with interests in fungal, viral and nematode pathogens as well as soil microbiology. There are collections of taxonomic importance held in the Division as well as community-wide databases of some significance (DPV and PHI-base).

Plant and Invertebrate Ecology Division – invertebrate and plant ecologists who contribute to, manage and utilise two internationally valuable resources: the 40+ year Rothamsted Insect Survey and the National Willows Collection (aligned with biomass breeding)

Agriculture and Environment Division – soil and environmental scientists who contribute to, manage and utilise the 160 + year database and archives of samples associated with the Rothamsted "classical" experiments (cf eRA – electronic Rothamsted Archive).

Information about the Rothamsted Insect Survey, the Rothamsted "Classical" Experiments, eRA and the National Willows Collection can all be found at:

www.rothamsted.bbsrc.ac.uk/resources/LongTermExperiments.html

The above information is supplemented here with information about collections and data held in PPI Division

Nematology

The nematode slide collection represent a large collection of soil-dwelling free-living and phytopathogenic nematodes from around the world: many preserved specimens, "populations", and >20,000 slides. This is currently not curated and deteriorating. It needs digitising, cataloguing and linking to DNA sequence data which could be obtained from the preserved specimens and possibly from slides. This unique collection, one of only three worldwide, is in regular demand for consultation from UK, Europe and international scientists. Digitisation would greatly improve accessibility; DNA sequence information would extend relevance to a wider range of science. Professor Brian Kerry has produced a detailed case for support. The future of DNA-based identification systems is limited without taxonomic information and *vice-versa*. However, collaboration with ecologists and molecular phylogeneticists (e.g. Mark Blaxter, Richard Bardgett) would raise the profile of the collection and highlight its utility to the research community in the UK and internationally. There is also small working collection of phytopathogenic nematode lines that needs regular subculturing and quality control.

Plant and soil microbiology

Fungi

The PPI Division holds a large collection of plant pathogenic and saprophytic fungi, mostly isolated by Rothamsted scientists over many years. Isolates (> 600) exist freeze dried and on slopes and they are catalogued in a database. The recent OREGIN project (oilseed rape genetic

improvement – Defra funded) is generating a unique internationally-derived collection of pathogenic fungi from brassica species to provide plant breeder with access to characterised genetic diversity as a resource for disease resistance breeding. A database connects details of the isolate collection to a database holding brassica phenotype, genotype and genomic data. Funding for this project finishes in 12 months time and the full value of this resource will not be without ongoing support.

There is a unique collection of wheat-associated pathogenic and non-pathogenic fungi (including *Gaeumannomyces graminis*) which has been assembled by Dr Geoff Bateman and is in danger of being lost when he retires at the end of 2006. In addition, there is a large and unique collection of grain spoilage fungi collected by Dr John Lacey (long retired).

Zoosporic plant parasites

There is a unique collection of (ca. 100 isolates) obligate zoosporic plant parasites (*Polymyxa*, *Olpidium* and *Ligniera* spp.) and, in some cases, containing plant viruses that they transmit. This is held in dried plant roots containing resting spores. The collection was made by Dr Mike Adams who retires in 2006.

Bacteria

There is a large collection of rhizobia, many from the John Innes Collection, and single copies of lyophils representing part of the Rothamsted Collection of Rhizobia (most of the latter collection was sent to Aberystwyth or privatised by agreement with a commercial inoculant company in the 1980s). There is a smaller collection of other soil and rhizosphere bacteria made by Dr Penny Hirsch.

There is a unique collection of more than 50 populations of *Pasteuria penetrans* (obligate pathogen of nematodes) collected internationally by Dr Keith Davies.

Plant viruses and phytoplasmas

There is a reference collection of virus-infected plant material (45 accessions) and key phytoplasma isolates maintained in living plants (*Vinca*). These are of significance for diagnostic work conducted in support of developing country agriculture.

Antisera

There is an antiserum collection of >300 accessions including sera raised against viruses, bacteria and plant proteins. This is held at -20C and emanates from the work of Dr Phil Jones who retires in 2006.

Virus images

There is a massive collection of >50,000 virus images dating back to the birth of electron microscopy in the 1950s – currently Roy Woods (long retired) is going through the collection to secure the best and most representative images. This "classical" collection needs cataloguing.

Databases

The Descriptions of Plant Viruses (DPV) database (http://www.dpvweb.net/) is curated by Mike Adams (retiring in 2006). This is a much used repository for data on plant viruses and viroids as well as providing tools for manipulation and display of sequence data etc. The database is currently owned by the Association of Applied Biologists but how it will be maintained into the future is uncertain.

PHI-base is a new RRes-created resource that assembles data on genes involved in host-pathogen interactions (Winnenburg *et al*. Nucleic Acids Research **34**: D459-D464 2006). New informatic developments of this kind will need long-term curation.

BBSRC consultation questions:

1) What resources should be supported?

The highest priority is to provide continuity of support where there are well formulated plans or existing mechanisms in place to provide ready community access to high-demand, electronically documented and well curated biological resources closely associated with well-structured databases and informatic resources which are generating added value by being continuously added to by users of the resource. A good example of this situation is the Nottingham Arabidopsis Stock Centre – but other smaller, equally well integrated, professionally managed and valuable resource collections are worthy of support.

a. Does systems biology present new resource needs?

Yes, especially in relation to ecosystem functioning, plant health and investigation of disease complexes rather than simple interactions. Much more information is needed to identify environmental (soil) microorganisms, assign them to trophic levels and attribute functions to them.

c. Principal bioinformatics resources

EMBL, NCIMB databases

d. Principal biological resources

Sequence information linked to taxonomic identification.

It is essential that resources (ecosystems, culture collections, databases) are maintained and curated. BBSRC is one of the few sources of support for fundamental studies on natural ecosystems and agroecology. BBSRC support is therefore essential for the 'Classical' and other long-term experiments at Rothamsted, including the Insect Survey, the unique Sample Archive, and the Electronic Rothamsted Archive (e-RA). These resources are critical to UK bioscience in that they provide a unique and truly long-term view of agricultural systems. The resources are very well used by a wide community; many requests for samples and data are received and satisfied over a very wide scope. For example, the spectrum of recent research papers published range from the impact of nuclear weapons tests to studying the diets of pre-historic man. The resources certainly underpin BBSRC's strategic objectives in sustainable land use and platforms for systems biology at the land management level.

Detailed comments on the requirements for culture collections were given in the response to the consultation on Future Directions in Microbial Science – which it would be worthwhile consulting.

The intention to provide support fro bioinformatic and biological resources is extremely relevant to the internationally unique Rothamsted Insect Survey data and insect collections. Minimum support is forthcoming from the Lawes Trust to sustain data collection by additional support would enable better maintenance of databases, and improvement of functionality (eg graphics, mapping). Furthermore, additional support could be used to improve detection of insects using radar (ability to detect insects the size of aphids is highly desirable). Proper curation of data and samples is vital to achieving their full potential. These databases and sample collections are used extensively to underpin work on the population dynamics and genetics of insects which provide essential ecosystem services, are pests, are vectors of disease or are of conservation concern.

The work ranges from fundamental to applied with an interplay between biometrical and experimental investigations. The work has significance because insects are essential components of all ecosystems and without them, ecosystem services would collapse. It is vital to be able to predict the impacts of rapid environmental and socieconomic change on insect population dynamics and provision of ecosystem services in order to be able to prioritise preventative measures or mitigation. Rothamsted holds by far the most comprehensive data of its kind in the world and there is increasing interest in the non-agriculturally relevant components of the suction traps (eg potential vectors of human disease). Requests for collaborative access to data and samples are on the increase.

e. Resources currently missing

Soil and environmental microorganism databases

There is no facility to store frozen samples of soil and plant material, etc. Many modern techniques require such material, not air-dried or oven-dried samples. This is costly but essential for effective future applications. An adequately resourced strategy for the long-term storage of frozen samples should be developed.

g. Resources relevant to be funded

Support is urgently needed to maintain the Rothamsted nematode slide collection (see above), digitise it, create an electronic database, link it to sequence data and publicise it so that it can be fully used by the UK and international research community.

Currently BBSRC GSG funding supports part-time curation of the main PPI culture collections and could carry forward the OREGIN collection. To make full use of the other resources held by PPI further support is needed so that they are all held on one database and properly curated to allow provision of cultures on request, replacement of stocks and addition of new accessions.

2) How should funding be used?

a. emphasis on resources previously established by BBSRC funding?

Not necessarily – funding from Defra or EU is indirectly supported by BBSRC at their institutes and the resources gathered under projects funded by such sponsors remain valuable to the UK and international research community.

Responsive mode grants often support data management for the duration of the grant. Funding for longer-term data management is needed, and for the preparation of adequate metadata (description of the data and resources available) as short-term staff, working on grants, are unlikely to be around to explain the origins and content of the data.

b. What emphasis on further development of existing resources (new functionality)?

It is important for older established collections such as the Rothamsted nematode slide collection to be linked to DNA sequence data. There is also the issue of maintaining collections of key isolates, mutants and transformants for comparative and functional genomic studies. This requirement will increase as genome sequences of key target species become available (eg *Mycosphaerella, Fusarium, Leptosphaeria*).

c. Should BBSRC support static resources?

Yes, where appropriate

d. Should existing resources be brought together in new ways?

Yes, e.g. all PPI microbiology resources could be considered as one resource rather than several specialist collections.

BBSRC funds a considerable amount of long-term monitoring as part of NERC's Environmental Change Network. BBSRC and NERC should consider coordinating their support of bioinformatic and biological resources.

e. Should BBSRC open up access to resources?

Yes, publicity is lacking

f. Should funding be focused on publicly accessible resources only?

Yes

3) What are the key components of more sustainable support?

a. Duration of funding

This is certainly an important issue. The long-term experiments and Rothamsted Archive have minimal funding from the Lawes Agricultural Trust to sustain basic management of the experiments, sampling and data handling. Most recent research has been funded by short-term grants; for example, recent Defra-funded work on the sustainability of cereal yields ended after 3 years and before the work was concluded to our satisfaction.

b. Location and management

Sustainable support needs to be better co-ordinated. Thus, as stated above, co-ordination by BBSRC and NERC of their bioinformatics and biological resources would greatly improve sustainability and the usefulness of the resource. Resources are best kept within active research groups rather than being at a centalsent to some central location remote from key users.

Charges should not exceed actual costs unless real QA conducted – which is rarely feasible. Pathogenicity assays for example are not feasible in most cases and when archived samples are requested functionality may not be guaranteed.

c. Monitoring and assessment

Rothamsted maintains a record of the requests it receives for samples and data, as well as an approval system to ensure that samples and data are not wasted, and a database of publications arising from the use of Archive samples and the long-term experiments. Such monitoring is essential to ensure that resources are being utilised and funding not wasted.

d. When is funding no longer required?

The long-term experiments are regularly reviewed for their relevance and use. We have 'mothballed' several so that they require minimal if any maintenance but can be reactivated if they once again become relevant. Such reviews are necessary for any bioinformatics or biological resource. However, it is also important to note that the Sample Archive and long-term experiments become more valuable as they get older and the Archive and database grow. There will be a period in any new archive/database when it may appear to have outlived its usefulness. Maintaining it over this period may well prove to be immensely valuable but hard to justify, in that clear and factual expectations of future use may not exist.

When there is no longer interest in plant growth and soil health and when plant pathogens are no longer a problem in the UK – even when the UK imports all its food there will still be a need to protect the environment. Reference strains of key species will be required for post-genomic comparative and functional studies for the foreseeable future. In fact the need for representative collections may well increase as interest shifts from reference strains to embrace issues of biological and functional diversity.

4) Other issues - looking to the future and beyond

Resource management is critical to QC and QA. Any resource management policy must be coordinated with QC/QA policy.

BBSRC has just reviewed its data sharing policy. Clearly, bioinformatics and biological resources must be supported adequately and appropriately if data are to be effectively shared. The resource management strategy must be co-ordinated with the data sharing policy.

Compiled from contributions by: Penny Hirsch, John Lucas, Keith Goulding, Richard Harrington, Chris Rawlings, Brian Kerry

Professor Ian Crute, Rothamsted Research
Posted 28 February

The University of Manchester, Simon Hubberd

Tools and resources discussion – comments from Manchester-based Bioinformaticians

Manchester-based bioinformatics group leaders met on 22 nd February to discuss the Consultation document from BBSRC concerning the Tools and Resources programme, including representatives from the Faculties of Life Sciences, Schools of Computer Science and Informatics, the National Centre for Text Mining (NaCTeM) and the Biologists who actually use such tools and resources.

1. What bioinformatic and biological resources do you think should be supported?

We believe the principal bioinformatic and related resources or areas that need to be supported can be summarised generically as:

- Data capture tools/methods/resources
- Data storage
    - Relational or similar databases
    - Collections of flat-files in standard formats
    - XML repositories
- Data standards and related initiatives including ontology development
- Integration tools and middleware
- External analysis tools
    - Such as:
        - Microarray/proteomics/metabolomics/systems biology software
        - Structure analysis/prediction software
        - Toolboxes, such as those produced in MATLAB and R

More specifically, several key themes emerged which can be illustrated with examples.

1. Text mining Text-mining is a good example of an area which is growing rapidly, where access to data is currently limiting, as are associated tools and resources to get the most out of it all. Effort is required to increase efforts to integrate relevant ontologies with experimental data using text mining. Data derived from text mining (terms, interactions) will update and populate existing ontologies and will provide the "glue" between knowledge, text and experimental data. This theme of integration recurred throughout our discussions. One point raised was that funding could be invested in increasing access to large non-public domain data – which exists in the biological text-mining domain – for the benefit of the user community. This would allow text-mining groups to generate derived, secondary data, which could be made publicly available, unlike the private data.

2. A "resource compass" Another major theme was "Not knowing where to start"? This is a point that was well illustrated by biologists who feel overwhelmed by the shear volume of websites, tools and software. Where should they start? We believe a major effort should be put into developing a **registry** of available services which is curated by bioinformatic experts, in much the same way that sequence motif and profile databases or genome annotation projects are curated by experts. This would lead to a set of quality, documented, and ideally well supported services that could be integrated.

3. Programmatic interfaces to resources As well as curating the services, they need to made available programmatically. That is, not just as standalone tools that the user interacts with on the command line, or via a single web-form. Rather, we would hope that they would be wrapped as Web Services, logged in a central registry, so that more sophisticated (but easy to use) interfaces can be constructed to extract the data. A good example is the BioZon interface developed at

Cornell (http://www.biozon.org/). We considered a local example, which would be a systems biology database of available SBML models, which might be selected from for an appropriate task.

4. Portals The natural progression then is to support efforts to bring the integration about, potentially under the umbrella of technologies such as **portals**. This would allow users to bring together services available from all over the UK (or wider) informatics community, and they could choose the ones they want from the registry. It was suggested that a "friend of a friend" type scheme could exist, where recommendations could be made so that "we recommend service X if you want to compare A with B".

This model requires investment in multiple areas. It requires investment in a few key centres which can help develop integration middleware and support, as well as continuing to fund existing resources which are needed – both in terms of resources such as databanks and databases, and analysis tools to perform transformations, calculations and comparisons on the data.

2. How do you think the funding should be used?

We agreed the investment in some key centres makes sense, perhaps focussed around existing strengths in either thematic areas of biology (systems biology and functional genomics, sequence analysis and databases, text mining, structural bioinformatics, etc) and/or potentially around organisms in a similar way to the IGFs. Thus the development effort for a component of the toolset can be streamlined at one centre, where critical mass is probably already available. These centres would reach out to smaller, more focussed resources, or arguably much will already be available at that site. For example, we agreed it was important that tool and resource development remains very close to the research itself – i.e. groups developing microarray tools and databases should work closely with groups carrying out such studies, and ideally with groups working on the technologies themselves.

We discussed these ideas comparing two contrasting approaches: federation *vs.* warehousing. There were pros and cons on both sides, although the consensus was that federation might be the better long term strategy, even though the lead time to development was longer.

We agreed that point 2d) was important – that there was a great need for existing resources to be brought together in new ways. This again harks back to the key theme of data integration.

3. What are the key components of 'more sustainable' research?

We believe long term funding is required to achieve the long-term goals we have outlined. Some periodic pump-priming type funding (e.g. Tools and Res Development fund) is necessary and invaluable, but for the greater strategy either 5 year (or 3 + 2 rolling) funding is needed. This fits with the model of funding a few key centres, and getting the other satellite groups to work together with the large centres to make best practice available directly to the wider community, via standardised interfaces and registries of available tools and services. We discussed the HGMP as an example. This was a fairly poor interface to some out of date tools, but became very popular with biological community as it was a "one-stop shop" and was accessed in a consistent fashion. Importantly, however, it lacked the direct connection to research activity which we believe is essential to keep bioinformatics relevant and up-to-date.

This "hub and spokes" type model also fits in well with a funding strategy. A "sunset" clause can be built in, and would need the input of a BBSRC panel to judge the performance of the tools and services both at the level of "hubs" providing integration and interfaces (as well as local tools and services) and the satellite lab "spokes" which provide specialist tools and services. If a strong case can be made to keep the satellite tool going, via lobbying from the "hub", the user community, and support by an external review panel, then it can be keep going. Otherwise, if

necessary, it can be subsumed by the "hub" or else left to "die". Similarly, it may be necessary to know when commercial products become *de facto* standards and ought to be licensed into the services grid, rather than to continue funding tool development from the academic side.

A further gain from "hubs" will be to encourage biologists to get involved in using tools, and hopefully also to play a greater role in ontology development – where it is absolutely essential that they are involved. They will also serve as beacon web-sites/portals for international collaborations, advertising ongoing research activities in UK bioinformatics.

This has implications on the nature of funding. We discussed at length that the existing "postdoc" model does not fit well with this plan, and the need for Project officers, Experimental Officers, and Computer Officers on these types of grants is crucial. These are people with expert skills in the technological side, but their main role is arguably not going to be principally research. This needs to be taken in to account in any call. Similarly, to retain such individuals will be vital.

This document was circulated more widely, but contains direct contributions from the following: Dr. Simon Hubbard (FLS, Bioinformatician), Prof. Jun'ichi Tsujii (NaCTeM, Text-Mining centre, MIB), Dr. Sophia Ananiadou (NaCTeM, Text-Mining centre), Dr. Goran Nenadic (School of Informatics), Prof. Douglas Kell (School of Chemistry, MCISB), Prof. Terri Attwood (FLS, Bioinformatician), Dr. Simon Lovell (FLS, Bioinformatician), Prof. Andy Sharrocks (FLS, Biologist), Dr. Jane Mabey Gilsenan (FLS, Bioinformatician), Prof. Carole Goble (School of Computer Science), Prof Norman Paton (School of Computer Science), Dr Robert Stevens (School of Computer Science).

Simon Hubberd, University of Manchester
Posted 28 February

University of Nottingham

1. What bioinformatic and biological resources do you think should be supported?

Points to consider here include:

a. It is proposed that funding:

- should focus on resources critical to UK bioscience research and without which the community could not achieve its full potential.

  It is essential that UK research needs are met, but we should aspire to international standards. The latter could be effectively done through collaboration with the European Bioinformatics Institute (EBI).
- should focus on resources which are / would be well used by a broad UK community.

  This should generally be the case, but I think there will be occasions when we need development of new bioinformatics resources in parallel if not ahead of major research developments. This would be especially the case for bioinformatics infrastructure that would take a long time to put in place.
- should underpin the delivery of BBSRC's strategic goals. *Absolutely!* For example, does systems biology present any new resource needs? If so, what are they?

  Systems Biology does require some new resource needs. The area does have emerging information standards, but there is still a lot of work to be done to integrate post-genomic data with systems models. This also needs to be extended to proteomic and metabolomic data. Also, at present, systems biology databases are little more than repositories and far more sophisticated database schemas are required if we are to step up the pace of systems model development.

b. What resources do you anticipate will be required for future bioscience developments?

These are difficult to estimate, but should embrace cross-discipline initiatives.

c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

In terms of infrastructure, there will be increasing volumes of image data to manage in addition to the other datatypes currently stored (post-genomic etc). This will require large data storage devices, compute servers and fast datalinks. To be done responsiblyand efficiently, it would be better to place these facilities at sites that already have a track record in this area. Such facilities will require PDRA level informaticians to develop the resources and deliver a large training programme in their usage. To be fully effective, this resource development must include appropriate consultation with the user communities, and the biggest challenge will be persuading these communities to adopt data standards including the use of controlled vocabularies.

d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

Broader and deeper expertise in plant and large animal physiology. In addition

e. Are there any resources that are currently missing?

We are not aware of any satisfactory databases for biological images/videos or plant/animal behaviour at the individual or population level.

f. Are there any resources that it would be inappropriate for BBSRC to support, for example for ethical or societal reasons?

g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources which you consider could be relevant to this funding mechanism?

2. How do you think should the funding be used?

Points to consider here include:

a. Should emphasis be placed on supporting resources previously established by BBSRC funding?

Yes, where these are clearly addressing the research communities' needs.

b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)?

There is an ongoing need for more refined ontologies beyond the molecular scale to include multicellular organism biology and behaviour. The increasing volume of image data will require special data management.

c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)?

This has been a thorny issue for many years. There is an international need to support certain categories of data (e.g. biomolecular sequences) which goes beyond the means of the BBSRC and should be supported by some international body. The current challenge for people who have developed new (and valuable) databases is maintaining them beyond the lifetime of the original grant. It is very difficult to make a compelling case for database maintenance through responsive mode funding. Hence, a mechanism is required to ensure static resources can be supported, perhaps via RCUK and other funding agencies.

d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges?

There are benefits of scale in concentrating inter/national databases in a restricted number of physical locations.

e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence?

BBSRC-funded resources should be open-access and education/training should be an essential component of any new resource-building proposal.

f. Should funding be focused on publicly accessible resources only?

The case for this can be easily made.

3. What are the key components of ' more sustainable ' support?
Points to consider here include:

a. Duration of funding;

This is highly dependent on the type of resource being developed, so it is difficult to be prescriptive. For example t he main microbe collections for yeast are held by NCYC (National Collection of Yeast Cultures) based in Norwich and part of the IFR. There has been some discussion concerning their activities - and a requirement for improvement has been debated. The bacterial collections are more dispersed including NCIMB (National Collection of Industrial and Marine Bacteria), based in Scotland. A database of industrial yeast (brewing, baking, winde, biofuel) genome databases and resources is badly needed, and would be an extremely useful resource.

However long-term funding is essential for the maintenance of a range of databases, which will allow for future developments in data storage and access etc.

b. Location and management;

Restrict the number of locations to several national centres and the EBI, where adequate management can be maintained more cost effectively.

c. Monitoring and assessment; and

Perhaps assessment annually for the first 3 years and then every 3 years beyond that.

d. When would you anticipate that funding is no longer required (e.g. lack of use, operating on cost recovery basis; resource no longer of high strategic relevance to current UK bioscience research)?

Lack of use is OK if there is no genuine research need. It would be better to ensure that resources are developed such that the relevant community does wish to use it. A cost-recovery basis is satisfactory for static resources, but active database providers are usually looking to develop their services and further funding will be required for that. A focus on resources of only high strategic relevance should only be followed if the available funding is highly restricted.

4. Other Issues

Looking to the future and beyond the remit of this activity, are there any other resources (out with bioinformatic and biological resources) that could be important for future UK bioscience research?

Should bioscience databases link to other databases e.g. medical. If so a strategy would need to be developed on how would this be developed/managed etc.

Bob Webb, University of Nottingham
Posted 28 February

University College London, Dr Andrew C R Martin

This submission is made after consultation with other members of the department and is made on behalf of the department. Others in the department may also submit their own comments.

It has already been pointed out that databases can be divided into different categories: 'Institutionalized', 'Community' and 'Hobby'. The third term is somewhat pejorative as the definition includes some major resources such as CATH and SCOP which are neither Institutionalized not Community databases.
Another division is between those which can be largely automatically maintained and those which require huge amounts of manual maintenance.

Clearly it should be a requirement that when a new database is established, its maintenance should be as automatic as possible to maximize the likelihood that it will continue to be maintained. We have put a lot of effort into ensuring this is the case with the databases that we have developed. Nonetheless, it is often the case that the source data on which we rely may be a moving target and however automated the processing system may be, when source formats change, parsers have to change (for example the dbSNP XML format changed substantially last year). Thus even a completely automatically updated database may require some maintenance re-coding to cope with such changes and there is currently no way to access funds to do so. Clearly databases requiring large amounts of manual curation are extremely important, but I shall restrict the rest of my comments to databases that can largely be automatically maintained. A further division is between databases which set out to do some specific task (often adding knowledge to data) and those which are simply repositories of data. Some of these data may not have an obvious function at the present time, but who knows what may be useful in the future? Ideally all data collected by biologists should be archived, possibly nationally, but more likely at campus level. Data which experimentalists plan to keep may well be lost through lack of backups or the use of proprietary formats.

Previously discarded results from (for example) sequencing and microarray experiments should also all be archived using relevant data standards. Results which might not provide anything of relevance to the researcher who performed the experiments may well be useful to someone else. This would lead to a requirement for funding of central data handling, archiving and curation within universities. Deciding which resources should be supported is of course a major challenge. Just because a resource is a 'Hobby' database does not mean that it is not absolutely critical to the work of many hundreds of research groups around the world. Equally a database which is only accessed by a very small group of researchers may still be key to fundamental scientific advances. Selection of funded resources should clearly be driven by biological needs, but biologists may not be the best people to see the potential of a database. It is essential that there be a separate panel whose role is to assess requests for funding on benefit to both biologists and bioinformaticians. At a technical level, I have already suggested that wherever possible (and of course it is often not possible), databases should be maintained automatically. It should also be a requirement that databases should not just hide behind a web interface, They should make their data available, either for local download, or accessible via SOAP or REST-style interfaces. A database which hides solely behind a pretty HTML interface may be very useful to the Biologist interested in one-off searching, but its wider application is severely limited if the data are only accessible via screen scraping.

In summary:

1. There should be access to funding for developing both small and large resources without any requirement for hypothesis-driven research.

2. Once an activity has been funded there should be some form of fast-track funding scheme for maintaining that resource. This should include access to small pots of funding (perhaps 3-6 months) for occasional maintenance of automated resources.

3. There is a need to fund resources to archive all collected data at a campus level.

4. If it is possible to automate updating of a database that should be a requirement for funding. Some databases cannot be maintained automatically and manual updating of such resources must also be supported.

5. Access to data via download, SOAP or REST should be an absolute requirement. 6. There should be a separate panel that assesses value for both the biology and the bioinformatics communities.

Dr. Andrew C.R. Martin, Department of Biochemistry and Molecular Biology, University College London
Posted 27 February

Roslin Institute

Having read the comments on the web site, I heartily agree with most of them, particularly those of Mike Ashburner.

The need to constantly re-invent database and software resources in order to compete with hypothesis-driven response-mode grants has been a problem for many years - and one that BBSRC have seemed to wish would quietly vanish. It is deeply encouraging to see this review take place and I sincerely hope that something concrete and useful comes out of it.

I agree that the continued funding of these resources does need to be peer-reviewed. How that mechanism might operate I am not entirely sure.

I believe that the EPSRC run a scheme called 'platform awards' (?) which are designed to support strategically important resources - I could be wrong on that though.

One other criteria that has not been discussed but which is a useful indicator of resource utility is the degree to which a particular resources interfaces with other resources. This will become increasingly important in the next few years as we move more fully into an age of systems and comparative biology.

Dr Andy Law, Head of Bioinformatics,
Roslin Institute
Posted 27 February

CABI

1. What bioinformatic and biological resources do you think should be supported?

a. It is proposed that funding:

- ***should focus on resources critical to UK bioscience research and without which the community could not achieve its full potential.***
- ***should focus on resources which are / would be well used by a broad UK community.***
- ***should underpin the delivery of BBSRC's strategic goals. For example, does systems biology present any new resource needs? If so, what are they?***

b. What resources do you anticipate will be required for future bioscience developments?

b.1 A new taxonomy developed using molecular tools but based upon traditional concepts until a new understanding is fully demonstrated. Currently a polyphasic approach is necessary, as not all methodologies will provide a definitive identity of all organisms, this demands a co-ordinated and quality controlled approach to build the molecular databases.

b.2 Enhancement of information on biological resources that are held in the UK's microbial collections

b.3 Co-ordination of biological resource centres to reduce duplication of effort and guide them to provide the tools for biotechnology and an information resource that is able to support scientific development.

b.4 Strategic partnerships to bring critical mass, expertise and technologies together to make best use of them in a cost effective manner

b.5 Generation of more data, better access to it and improved data analysis

b. 6. Information resources including literature, full reports and papers, its access and dissemination

c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

Sequence data linked to functional genomics data and traceable to the source biological material

Validated and authentic information resources such as bibliographic databases teaching and training tools giving access to literature, abstract databases etc relevant to microbiology.

d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

d.1 *Ex situ* biological resource centres – public service culture collections (see below)

d.2 Molecular databases – need to be more comprehensive and accurate

d.3 Functional genomics information – to target properties for production/use or as targets for control – more data is required on microorganisms, created systematically and without duplication

e. Are there any resources that are currently missing?

A coordinated and comprehensive DNA bank for microorganisms (distributed)

Validated and authentic information resources such as bibliographic databases teaching and training tools giving access to literature, abstract databases etc relevant to microbiology.

f. Are there any resources that it would be inappropriate for BBSRC to support, for example for ethical or societal reasons?

None identified

g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources that you consider could be relevant to this funding mechanism?

UKNCC national collections and a network of underpinning resources linked to the OECD Global Biological Resource Network

The UK National Culture Collection (UKNCC) collections house microbial and cell line genetic resources. These are defined as genetic material of animal, plant or microbial origin, which has current or potential use. Genetic resources exist both *in situ* (living animals, plant and micro-organisms in their natural environment); and *ex situ* (outside their natural environment, e.g. in a gene bank, biological resource centre or culture collection). Genetic resources can be conserved in a variety of forms, including as whole plants, animals and microorganisms, seed, embryos, semen, and replicable parts such as genomes and DNA. The UK has several centres that maintain such resources but they are loosely linked, if at all, and many are somewhat adrift from user needs.

Genetic resources (GR) are an important resource for both the UK and other countries, and hold potential benefits for farmers, industry and the public at large. There is a need for the conservation and sustainable use of, and facilitated access to, animal, plant and microbial genetic resources, to support biotechnology, bioscience industry, education, sustainable agriculture and horticulture, and to support related environmental improvements, rural development, scientific research and conservation of heritage and biodiversity, now and in the future. A UK genetic resources policy is needed which requires:

- A plan to conserve and sustainably utilise the UK's genetic resources
- Joint programmes of *in situ* and *ex situ* conservation
- Framework for co-ordinated research, knowledge development and gap analysis
- A better understanding on the genetic resource needs of industry, education, research etc. an anticipated outcome of this consultation
- Priority list of actions to protect and utilise the UK's biological resources
- A molecular approach to its better understanding and use

Conservation of *ex situ* GR is currently the responsibility of research councils, collection holders and researchers. It is essential that such activities are co-ordinated and be complimentary to *in situ* conservation and a plan to achieve GR security through a gap analysis should be established.

Characterisation and evaluation of GR is a responsibility for levy bodies, research organisations, collection holders, and UK Government. It should have particular focus on GR that are of environmental, bioscience industry and educational importance. UK Government should seek to co-ordinate these efforts into a programmed approach. Within such an approach, the private

sector and levy bodies should focus on commercial benefits while UK Government should focus on public good benefits.

The UK Government invested in the establishment of the UKNCC in the mid 1990's and it has been struggling to maintain momentum since funding ceased in March 2000. The UKNCC is an affiliation of UK national public service collections with the prime goal to promote the member collection interests in order to aid their long-term sustainability. It supports member collections in the provision of high quality, efficient and effective services to the biotechnology and biological science communities. The UKNCC recognises its role in under pinning science research and development and its role in conservation and makes every effort to find support for these roles. The UKNCC provides a forum for the UK public service collections to develop common coherent strategic policies through dialogue with users and funding agencies but could do much more. It is failing because the UK culture collections are fighting for survival and the UKNCC has no resources to help them. Without investment the collections are unable to keep pace with developments and will be used less and less the downward spiral. The collections need to respond to the demands of the 21st century not that of the 20 th when the UKNCC was established.

The major strength and advantage of the UKNCC over other collections is the joint potential of their or their parental organisation's wealth of expertise. The UKNCC collection's focus on particular groups of organisms has allowed them to develop specific expertise holding a national resource in excess of 100 000 organisms or cell lines. The UKNCC Quality Management System, which includes three collections operating to ISO standards, enables the UKNCC to provide a high quality traceable service. The key to increased business will be the provision of unique value added products.

UKNCC Achievements

- The UKNCC MOU – an agreement to collaborate
- The UKNCC Quality Management System
- On-line catalogue databases
- Links on key collection sites e.g. World Data Centre for Microorganisms; National Focal Point for Access and benefit sharing
- Marketing materials and presence at major conferences
- UKNCC integrated catalogues in 3 volumes as hardcopy and CDROM
- The UKNCC Biological Resource book
- The UKNCC forum to discuss opportunities for joint ventures
- Taking a lead in the OECD Biological Resource Centre Initiative
- Participation in the process of implementation of the CBD
- Collaborative research projects; Light cycler; Cryopreservation
- Common approaches to national and international policy and regulation compliance
- International recognition as a successful way to produce effective collaboration of collections

The networking of collections in the UK has still further to go. The collating and updating of large integrated databases is old technology. New tools exist to link live databases across the Internet negating the central role of updating and maintenance of data. The UK collection databases need to be connected live over the Internet. Data held by the collections on properties, habitats, structure and taxonomic relationships could provide a useful tool for bioprospecting. The links between the UKNCC members are currently superficial although the medical collections are uniting under the Health Protection Agency, the remaining collections could similarly link. Poor linkage and the lack of shared resources and operations are not cost effective. A broader international link would be even more productive and enable a much greater utilization of the limited resources available to conserve and utilise the huge potential. The OECD BRC Initiative offers a way forward. The aim is to create a global network of collections operating to common high quality standards with common goals and policies. Each member collection will have to

reach the status of Biological Resource Centre independently assessed by and agreed mechanism and standards by a third party, a national certification or accreditation body. The costs to transform each candidate collection are not insignificant but it is possible that such costs could be offset by resulting new business. However, not many collections are business orientated making this unlikely for most collections unless there is further investment. The underlying ethos of the GBRCN is that the user benefits from the accreditation of culture collections through better access to authentic and reproducible materials in a transparent and traceable way. There must be benefits for the collection to provide incentive to change. There is an ever-increasing demand for authentic reference materials as more and more industries are adopting certification or accreditation as a means to demonstrate quality and competence. This may be the driving force for the business elements of a collection's strategy for long-term sustainability but it is also an increasing requirement to satisfy the funders of research who seek high quality science and solutions. The ability to demonstrate the competence to carry out and manage high quality research is being recognised by Research Councils and Government Departments. Third party evaluation through accreditation or certification may be the only way to demonstrate this.

It is imperative that organisms utilised in biotechnology are maintained in a way that will ensure that they retain their full capacity. Culture Collections must ensure a quality product providing standard reference material that will give reproducible results. To achieve this collections must apply quality control and assurance measures to maintain these standards, taking into account the needs of users and of the facilities and resources available. The need for common standards is evident, as the task of maintaining representative samples of microbial diversity cannot be achieved by one collection alone. Therefore, it is essential that a worldwide network of collections interact to provide the coverage required by the user. In order that a customer of such a network would get a consistent level of service and quality it is necessary to set standards for all collections to attain demonstrating a need to network.

The advantages of accredited culture collections forming a network can be split into two groups, those that give benefits to the users and those that benefit the collection itself although several could fall in both categories.

User benefits

- A one-stop shop where both biological materials and the information associated with them can be found
- Conformity of both quality and authenticity of biological materials but also of processes and procedures to access them
- Confidence that the materials are fit for purpose
- Assurance that national law, policies and procedures have been followed

Culture collection benefits

- Recognition that they operate to international scientifically based quality criteria
- An international mark of quality
- Raised profile
- Sharing of tasks
- Common policies and procedures
- Competitive edge
- Level playing field
- Common access to data enabling links to be made to other international initiatives without duplication of effort
- Common approach to data access, sharing and interoperability
- Improved data usage
- Collaborative research and development

Inevitably the introduction of the GBRCN will have further costs but more importantly introducing the requirements of the standard and accreditation procedures to the collections to achieve the status of accredited BRCs is not insignificant. It will be the responsibility of Nations to align their collections for better international collaboration. The BBSRC can assist the UK collections participate fully. Used correctly the improved quality management and better product can be used to attract investment in the development of culture collections and the outcome will be beneficial to all concerned. Amongst the potential new duties identified in the OECD discussions are clearinghouse roles and control points in biosecurity, biosafety and offering transparency in access to organisms being proactive in compliance with regulations.

There are a number of UK collections that have recognised the need to at least take a first step and they have acquired certification to ISO9000:2000 series. They have not utilised the OECD standards as a basis. They see the benefits as improved business, they are better able to win contracts and research projects, they have better argument to retain or secure core funding, a greater profile, they gain entrance to new markets, have greater credibility, and have better operations, they can demonstrate tracability, i.e. it is easier to trace errors and processes are easier to audit. They see the disadvantages, as additional staff needed, more recording of activities, time-consuming detail. CABI has taken a lead in this respect and has worked closely with the OECD BRC initiative to introduce the OECD system. CABI has part of its services accredited to ISO 17025.

Further information on the Organisation for Economic Co-operation and Development (OECD) can be found on the OECD web site under Biological Resource centres and in *Biological Resource Centres – Underpinning the Future of Life Sciences and Biotechnology* (http://oecdpublications.gfi-nb.com/cgi-bin/oecdbookshop.storefront). In addition to the drive to improve quality culture collections must now deal with the vast diversity of new genetic entities generated by life scientists as they seek to reveal the genomes of many organisms and to engineer new cells with novel properties. This increased demand is occurring whilst statistics of the World Data Centre for Microorganisms (WDCM) show a reduction in the number of registered collections (http://www.wfcc.info). There are fewer and fewer biosystematists at a time when we still have the vast majority of fungi to describe. Taxonomy is in decline in the UK and requires better co-ordination and infrastructure. For the survival of collections to meet the needs of the coming century partnerships, sharing tasks and responsibilities, and coordination of effort is paramount.

Genomic studies are generating extraordinary amounts of information and taxing the capabilities of informatics for analysing and using data. It is vital that data generated from authentic organisms a problem demonstrated by today's databases that contain erroneous data that can undermine research. A recent study (Bridge et al., 2004) revealed of 206 named sequences of the ribosomal RNA gene cluster in fungi and up to 20% of these was considered unreliable.

The BBSRC can assist the UK microbial collections to:

- Contribute to the co-ordination of efforts to conserve biodiversity and to provide access to natural and engineered biological resources.
- Assist in the development of a co-ordinated international system for decision making to guide appropriate acquisition, maintenance and distribution of biological resources so as to avoid unnecessary duplication of effort while preserving critical levels of biodiversity.
- Modernise to incorporate the latest developments in web-based electronic communication, bioinformational science and informatics technologies.
- Co-ordinate and unify catalogues and databases to meet the requirements of science in the developing post-genomics era.
- Develop new systems and technologies for the long-term maintenance and distribution of large numbers of diversebiological resources.
- Co-ordinate curation, as well as development and networking of informatics tools for data analysis, comparison and visualisation.

- Ensure that the scientific community has access to affordable products and services.
- Establish strategic partnerships to take advantage of the rapidly developing new technologies.

2. How do you think should the funding be used?

Points to consider here include:

a. Should emphasis be placed on supporting resources previously established by BBSRC funding? ***- Support has been very niche driven; it is essential that a holistic view be taken. There is need to review those existing resources and rationalise them to some extent.***

b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)? ***– This should be done but must only after considering strategic partnerships to access new technologies and establish the functionality with minimum investment***

c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)? ***– No but support the opportunity to transfer from stasis to active development***

d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges? ***Absolutely e.g. the UKNCC collections and other key UK resources at the same time creating a molecular resource – DNA and sequence bank that complements and extends others in the public domain***

e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence? ***A key function of the BBSRC is to encourage the use of existing resources***

f. Should funding be focused on publicly accessible resources only? ***Yes, but incorporation of resources generated by BBSRC funded programmes bringing them into the public domain***

Suggested funding to support the development of UK culture collection resources to better support the UK's business, research and fundamental life sciences

Upgrading UK public service collections to the level of BRC

A funded programme to enhance the quality and bridge the gaps of public domain molecular information databases

A focus on the further development of a new taxonomy using molecular tools but based upon traditional concepts until a new understanding is fully demonstrated Currently a polyphasic approach is necessary, as not all methodologies will provide a definitive identity of all organisms, this demands a co-ordinated and quality controlled approach to build the molecular databases.

Enhancing and sharing of the data on the UK's *ex situ* biological resources

Co-ordination of biological resource centres to reduce duplication of effort and guide them to provide the tools for biotechnology and an information resource that is able to support scientific development.

The establishment of strategic partnerships to bring critical mass, expertise and technologies together to make best use of them in a cost effective manner

Generation of more data, better access to it and improved data analysis

3. What are the key components of ' more sustainable' support?

Points to consider here include:

a. Duration of funding; ***Dependant on activity but wherever possible a movement toward self sustainability***

b. Location and management; ***As appropriate to the activity but there needs to be some rationalisation and consolidation of activities in the case of the UK microbial collections revisiting the Whittenbury report of 1994 (commissioned by the OST) would be useful.***

c. Monitoring and assessment; and ***Utilisation of independent third party assessment schemes including certification and accreditation***

d. When would you anticipate that funding is no longer required:
d1. lack of use
d2 operating on cost recovery basis
d3 resource no longer of high strategic relevance to current UK bioscience research

David Smith,
CABI
Posted 25 February

UKNCC (United Kingdom Culture Collection)

On behalf of:

CABI Bioscience UK Centre
Culture Collection of Algae and Protozoa
European Collection of Cell Cultures
National Collection of Industrial, Food and Marine Bacteria
National Collection of Pathogenic Fungi
National Collection of Plant Pathogenic Bacteria
National Collection of Pathogenic Viruses
National Collection of Type Cultures
National Collection of Yeast Cultures

1. What bioinformatic and biological resources do you think should be supported?

a. It is proposed that funding:

- ***should focus on resources critical to UK bioscience research and without which the community could not achieve its full potential.***
- ***should focus on resources which are / would be well used by a broad UK community.***
- ***should underpin the delivery of BBSRC's strategic goals. For example, does systems biology present any new resource needs? If so, what are they?***

b. What resources do you anticipate will be required for future bioscience developments?

b.1 A new microbial taxonomy developed using molecular tools but based upon traditional concepts until a new understanding is fully demonstrated. Currently a polyphasic approach is necessary, as not all methodologies will provide a definitive identity of all organisms; this demands a co-ordinated and quality controlled approach to build the necessary molecular databases.

b.2 Enhancing data on existing biological resources

b.3 Co-ordination of biological resource centres to reduce duplication of effort and guide them to provide the tools for biotechnology and an information resource that is able to support scientific development.

b.4 Strategic partnerships to bring critical mass, expertise and technologies together to make best use of them in a cost effective manner

b.5 Generation of more data on biological materials, better access to it and improved data analysis

c. What are the principal bioinformatics resources required by the UK bioscience community? Which ones should take priority?

Sequence data linked to functional genomics data and traceable to the source biological material

d. What are the principal biological resources required by the UK bioscience community? Which ones should take priority?

d.1 *Ex situ* biological resource centres – public service culture collections (see below)

d.2 Molecular databases – need to be more comprehensive and accurate

d.3 Functional genomics information – to target properties for production/use or as targets for control – more data is required on microorganisms, created systematically and without duplication

e. Are there any resources that are currently missing?

A coordinated and comprehensive DNA bank for microorganisms (distributed)

f. Are there any resources that it would be inappropriate for BBSRC to support, for example for ethical or societal reasons?

g. In light of the proposals above and your consideration of the questions posed, please can you identify examples of resources, which you consider could be relevant to this funding mechanism?

UKNCC national collections and a network of underpinning resources linked to the OECD Global Biological Resource Network

The UK National Culture Collection (UKNCC) collections house microbial and cell line genetic resources. These are defined as genetic material of animal, plant or microbial origin, which has current or potential use. Genetic resources exist both *in situ* (living animals, plant and microorganisms in their natural environment); and *ex situ* (outside their natural environment, e.g. in a gene bank, biological resource centre or culture collection). Genetic resources can be conserved in a variety of forms, including as whole plants, animals and microorganisms, seeds, embryos, semen, and replicable parts such as genomes and DNA. The UK has several centres that maintain such resources but they are loosely linked, if at all, and many are somewhat adrift from user needs.

Genetic resources (GR) are an important resource for both the UK and other countries, and hold potential benefits for farmers, industry and the public at large. There is a need for the conservation and sustainable use of, and facilitated access to, animal, plant and microbial genetic resources, to support biotechnology, bioscience industry, education, sustainable agriculture and horticulture, and to support related environmental improvements, rural development, scientific research and conservation of heritage and biodiversity, now and in the future. A UK genetic resources policy is needed which requires:

- A plan for conservation and sustainable utilisation of the UK's genetic resources
- Joint programmes of *in situ* and *ex situ* conservation
- Framework for co-ordinated research, knowledge development and gap analysis
- A better understanding of the genetic resource needs of industry, education, research etc. (an anticipated outcome of this consultation)
- Priority list of actions to protect and utilise the UK's biological resources
- A molecular approach to its better understanding and use

Conservation of *ex situ* GR is currently the responsibility of research councils, collection holders and researchers. It is essential that such activities are co-ordinated and be complimentary to *in situ* conservation and a plan to achieve GR security through an integrated *in situ* and *ex situ* programme.

Characterisation and evaluation of GR is a responsibility for levy bodies, research organisations, collection holders, and UK Government. It should have particular focus on GR that are of environmental, bioscience industry and educational importance. UK Government should seek to co-ordinate these efforts into a programmed approach. Within such an approach, the private

sector and levy bodies should focus on commercial benefits while UK Government should focus on public good benefits.

The UK Government invested in the establishment of the UKNCC in the mid 1990's and it has been struggling to maintain momentum since funding ceased in March 2000. The UKNCC is an affiliation of UK national public service collections with the prime goal to promote the member collection interests in order to aid their long-term sustainability. It aims to support member collections in the provision of high quality, efficient and effective services to the biotechnology and biological science communities. The UKNCC recognises its role in under pinning science research and development and its role in conservation and makes every effort to find support for these roles. The UKNCC provides a forum for the UK public service collections to develop common coherent strategic policies through dialogue with users and funding agencies but could do much more. It is failing because the UK culture collections are fighting for survival and the UKNCC has no resources to help them. Without investment the collections are unable keep pace with developments and as a result they will be used less and less and their downward spiral will continue. The collections need to respond to current demands.

The major strength and advantage of the UKNCC over other collections is the joint potential of their, or their parental organisation's wealth of expertise. The UKNCC collection's focus on particular groups of organisms has allowed them to develop specific expertise holding a national resource in excess of 100,000 organisms or cell lines. The UKNCC Quality Management System, which includes three collections operating to ISO standards, enables the UKNCC to provide a high quality traceable service. The key to increased business will be the provision of unique value-added products.

UKNCC Achievements

- The UKNCC MOU – an agreement to collaborate
- The UKNCC Quality Management System
- The UKNCC Website http://www.ukncc.co.uk
- On-line catalogue databases
- Links on key collection sites e.g. World Data Centre for Microorganisms; National Focal Point for Access and benefit sharing
- Marketing materials and presence at major conferences
- UKNCC integrated catalogues in 3 volumes as hardcopy and CDROM
- The UKNCC Biological Resource book
- The UKNCC forum to discuss opportunities for joint ventures
- Taking a lead in the OECD Biological Resource Centre Initiative
- Participation in the process of implementation of the CBD
- Collaborative research projects; Light cycler; Cryopreservation
- Common approaches to national and international policy and regulation compliance
- International recognition as a successful way to produce effective collaboration of collections

The UKNCC was created through monies allocated by the OST and administered through the BBSRC. A total of 1.54 million was invested which supported: the continued survival of National Collection of Marine Bacteria (NCIMB); the transfer of the National Collection of Food Bacteria to NCIMB; purchase of UKNCC server and database development; marketing and UKNCC web site design; establishing a Quality Management System, funding a light cycler project; operating costs for the UKNCC Secretariat; the establishment of the national collection of Pathogenic Viruses; BBSRC management costs; a market study; production and marketing of catalogues including CD version and posters; UKNCC Group Meetings costs; publication of the Uses and properties manual; and an MBA student project.

The initial investment in setting up the UKNCC was not targeted at strengthening the member collections or the infrastructure to network them. The networking of collections in the UK has still

further to go. Operating a single database is inefficient and inappropriate today and consequently the UK collection databases need to be connected live over the Internet. The UKNCC data and web site is now significantly out of date. Data held by the collections on properties, habitats, structure and taxonomic relationships could provide a useful tool for bioprospecting. The links between most of the UKNCC members are currently superficial although the medical collections are uniting under the Health Protection Agency, the remaining collections could be drawn closer together. Poor linkage and the lack of shared resources and operations are not cost effective.

A broader international link would be even more productive and enable a much greater utilization of the limited resources available to conserve and utilise the huge potential. The OECD BRC Initiative offers a way forward. The Organisation for Economic Co-operation and Development (OECD) has been discussing the establishment of a Global Biological Resource Centre Network (GBRCN) to facilitate access to high quality genetic resources. They reported in 2001 *Biological Resource Centres – Underpinning the Future of Life Sciences and Biotechnology* (http://oecdpublications.gfi-nb.com/cgi-bin/oecdbookshop.storefront). The UK needs to contribute in a practical way and the costs of adoption of the accreditation system demanded, upgrading of facilities, data sharing and operating costs of the global network will need to be met. The aim is to create a global network of collections operating to common high quality standards with common goals and policies. Each member collection must reach the status of Biological Resource Centre, assessed independently by agreed mechanisms and standards by a third party, a national certification or accreditation body. The costs of transforming each candidate collection are not insignificant but it is possible that running costs could be offset in part by resulting new business. However, not many collections are business orientated, making this unlikely for most collections without further investment. The underlying ethos of the GBRCN is that the user benefits from the accreditation of culture collections through better access to authentic and reproducible materials in a transparent and traceable way. There must be benefits for the collection to provide incentive to change. There is an ever-increasing demand for authentic reference materials as more and more industries are adopting certification or accreditation as a means to demonstrate quality and competence. This may be the driving force for the business elements of a collection's strategy for long-term sustainability but it is also an increasing requirement to satisfy the funders of research who seek high quality science and solutions. The ability to demonstrate the competence to carry out and manage high quality research is being recognised by Research Councils and Government Departments. Third party evaluation through accreditation or certification may be the only way to demonstrate this.

Inevitably the introduction of the GBRCN will have further costs but, more importantly, costs of introducing the requirements of the standard and accreditation procedures to the collections to achieve the status of accredited BRCs is not insignificant. However, used correctly the improved quality management and better product can be used to attract investment in the development of culture collections and the outcome will be beneficial to all concerned. Amongst the potential new duties identified in the OECD discussions are clearinghouse roles and control points in biosecurity and biosafety and offering transparency in access to organisms being proactive in compliance with regulations.

There are a number of collections that have recognised the need to at least take a first step and they have acquired certification to ISO9000:2000 series. They have not utilised the OECD standards as a basis. They see the benefits as improved business and are better able to win contracts and research projects. They have greater justification for retention or securing of core funding and a greater profile. They gain entrance to new markets, have greater credibility, have better operations and can demonstrate traceability, i.e. it is easier to trace errors and processes are easier to audit. They see the disadvantages, as additional staff needed, more recording of activities and time-consuming detail.

Genomic studies are generating extraordinary amounts of information and taxing the capabilities of informatics for analysing and using data. Biologists and biotechnologists will spend the next

few decades understanding and exploiting the information provided by these genome-sequencing efforts. These sequence data and their by-products – e.g. genome libraries – have to be preserved and made easily accessible. The quest to obtain information on each of the thousands of genes, gene products and other characteristics of each organism highlights the daunting task of storing, maintaining and disseminating this information faced by BRC data banks. It is vital that data are generated from authentic organisms, a problem demonstrated by today's databases that contain erroneous data that can undermine research. A recent study (Bridge et al., 2004) revealed that of 206 named sequences of the ribosomal RNA gene cluster in fungi, up to 20% were considered unreliable.

To cope with the massive expansion of biological resources, including living biological materials and data on genomics, the UK Culture Collections need to:

- Contribute to the co-ordination of efforts to conserve biodiversity and to provide access to natural and engineered biological resources.
- Assist in the development of a co-ordinated international system for decision making to guide appropriate acquisition, maintenance and distribution of biological resources to avoid unnecessary duplication of effort while preserving critical levels of biodiversity.
- Modernise to incorporate the latest developments in web-based electronic communication, bioinformational science and informatics technologies.
- Co-ordinate and unify catalogues and databases to meet the requirements of science in the developing post-genomics era.
- Develop new systems and technologies for the long-term maintenance and distribution of large numbers of diverse biological resources.
- Co-ordinate curation, as well as development and networking of informatics tools for data analysis, comparison and visualisation.
- Ensure that the scientific community has access to affordable products and services.

It is essential that such efforts are co-ordinated and a UK policy for the conservation, management and utilization of microbial diversity is put in place,

2. How do you think should the funding be used ?

The points to consider included:

a. Should emphasis be placed on supporting resources previously established by BBSRC funding? ***- BBSRC should work with others so that resources essential for UK research and development remain available and keep pace with current demands,***

b. What emphasis should be placed on the further development of existing resources (i.e. building in new functionality)? ***– The BBSRC should help develop existing resources and support the creation of strategic partnerships to create the functionality with reduced new investment***

c. Should BBSRC support static resources (i.e. those which are being maintained but not developed)? ***–Yes, if there is a need for them in their current form but particularly if they provide the pipeline for others to develop***

d. Is there a need for existing resources to be brought together in new ways in order to make them fit for purpose to meet current and future research challenges? ***Absolutely e.g. the UKNCC collections and other key UK resources at the same time creating a molecular resource – DNA and sequence bank that complements and extends others in the public domain***

e. Should BBSRC aim to open up access to resources and / or to raise awareness to their existence? ***A key function of the BBSRC is to encourage the use of existing resources***

f. Should funding be focused on publicly accessible resources only? ***Yes***

Suggested funding to support the development of UK culture collection resources to better support the UK's business, research and fundamental life sciences

Upgrading UK public service collections to the level of BRC

A funded programme to enhance the quality and bridge the gaps of public domain molecular information databases

A focus on the further development of a new taxonomy using molecular tools but based upon traditional concepts until a new understanding is fully demonstrated. Currently a polyphasic approach is necessary, as not all methodologies will provide a definitive identity of all organisms, this demands a co-ordinated and quality controlled approach to build the molecular databases.

Enhancing and sharing of the data on the UK's *ex situ* biological resources

Co-ordination of biological resource centres to reduce duplication of effort and guide them to provide the tools for biotechnology and an information resource that is able to support scientific development.

The establishment of strategic partnerships to bring together critical mass, expertise and technologies and to make best use of them in a cost effective manner.

Generation of more data, better access to it and improved data analysis

3. What are the key components of 'more sustainable' support?

Points to consider here include:

a. Duration of funding; ***Dependant on activity but wherever possible a movement toward self sustainability only achievable if the collection can generate a sound business that extends beyond culture sales and traditional services as these themselves are not profitable.***

b. Location and management; ***As appropriate to the activity***

c. Monitoring and assessment; and ***Utilisation of independent third party assessment schemes including certification and accreditation***

d. When would you anticipate that funding is no longer required: d1. lack of use
d2. operating on cost recovery basis
d3. resource no longer of high strategic relevance to current UK bioscience research

David Smith,
UKNCC (United Kingdom Culture Collection)Secretariat
Posted 25 February

University of Dundee

This consultation was highlighted to academic staff in the Faculty of Life Sciences, University of Dundee. The following points were raised:

Working at the interface between biology and chemistry, access to better chemical tools and databases that could form a resource for academics would be money well invested. Consideration should be given to an open source chemistry database that could be used by academics world wide. This could also house the raw data from low, medium or high throughput screens and biological data on receptor, target, cell and animal studies. The database should be suitable for substructure searches and other chemical queries. Editors of all chemistry journals should insist on final products of syntheses having a "chemical accession" number deposited with this database so that it becomes populated with novel chemotypes.

ACS with Chemical Abstracts and SciFinder Scholar has a monopoly on such information at present and have prohibitive costs often preventing/restricting academic access to this information. If one thinks back to the origins of genetic databases and structural databases, molecular biologists and structural biologists only started to deposit data en masse when this became a prerequisite for publication.

Pete Downes, University of Dundee
Posted 20 February

University of Cambridge, Michael Ashburner

For the consideration of your Consultation I attach a paper that is relevant.

I would, if you would allow, like to comment on two aspects of your brief, databases and physical resources.

1. Databases and informatics resources.

Modern biological research is totally dependent on the availability of well structured open source databases. There can be little question that this dependence will increase in the future.

These databases are, today, supported by three general mechanisms:

(a) Institutionalised databases, maintained by large 'permanent' Institutes such as the EBI in Europe, the NCBI in the US and the National Institute of Genetics in Japan. Although often dependent on outside funding, these databases have institutional support and commitment. They include the nucleotide sequence databases (aka Genbank), the protein sequence database (UniProt), the protein structure databases (PDB and MSD), the array databases (GEO, ArrayExpress) and so on. It is noteworthy that nearly all of these are international collaborations, both of institutions and in many cases of funders. Their important defining feature is that they have institutional support, they are not dependent on one or a few PIs. Should their present funding or leaders disappear it is their institutions that have - and accept - the responsibility for their future.

(b) Community databases. These are databases vital to a community and, typically, supported by short term grant funding and being the responsibility of a small group each of which is, again typically, an academic. Here I include all of the model organism databases (FlyBase, SGD, WormBase etc) with the single exception of the mouse database (MGD), which is institutionalised within the Jackson Laboratory. These databases typically have a single funder (most are funded by the NIH), and they are extremely vulnerable to the whims of funding and to the status of their PIs. They have no institutional support. For example if I were to give up my leadership role in FlyBase nobody in my Department or University would feel any responsibility for keeping FlyBase going. Yet these databases are _vital_ to the biomedical community (and this is _far_ wider than, for example, the Drosophila community - indeed we think more non- drosophilists use FlyBase than do drosophilists).

(c) Hobby databases. These are, typically, products of a single enthusiastic researcher and are, typically, designed around a particular research - rather than community - interest. They may get grant support for their first instantiation but probably not beyond that. I could name hundreds of these databases (see the Jan 2006 Nucleic Acids Research Database issue and the associated table from Galperin).

The UK and Europe have been appalling and very backward in giving support for databases in groups (a) and (b). By contrast, the NIH has a special funding mechanism, called P41, specifically designed for community resources. This mechanism has, for example, supported FlyBase continuously since 1992. P41 funding is peer reviewed and is competitive (highly), but is designed to be long term (funding is usually for periods of 5 years at a time). To my knowledge no UK or European agency has ever had the wisdom to follow NIH's example. It is P41 grants that support FlyBase, SGD, WormBase, dictBase, MGD, the Gene Ontology etc etc.

Database infrastructure funding must be competitive, it must go through peer review, but it cannot - and must not be made to - compete with research funding. The UK simply must design a way of supporting long term community infrastructure.

Yes, the UK's support should build on our strengths, but these community projects - however funded - are of international importance. It follows, in my view, that there should be international collaboration between funding agencies.

2. Stock Centres Much of what I have written also applies to stock centres Stock centres. This are of vital importance to the scientific community. The UK and Europe Drosophila community relys on the US funded Drosophila stock centre in Bloomington. Again, the US (through the NSF) has taken the support of stock centres seriously, in a way that no UK or European funding agency has done - see the NSF's Biological Collections Program.

In Europe we had a European Drosophila Stock Centre in Umea, Sweden, for many years. But the shortsightedness of the European funding agencies meant that this had to close and these stocks are now held in Japan. The new European Stock Centre, which holds unique and vital stocks, is in Hungary but this has almost _no_ funding, despite being an international resource. I am sure other communities could say similar things about the state of affairs for their organisms. Like community databases stock centres are not just for Christmas. Yes, their funding and performance must be peer reviewed and be competitive, but there must be an agency committment to the long term.

I could go on at great length about these subjects, and will if you invite me to do so. What I would like to see is, at least, a national cross-Council strategy with respect to these issues. Better, I would like to see a European strategy - but I remain a realist.

Michael Ashburner, University of Cambridge
Posted 17 February

University of Leeds

There is a real need for maintenance funding for strategic bioinformatic resources. My comments below refer to these resources only.

What resources should be supported?

The nature of the field and the way it has been funded over the years has produced a number of major resources at larger centres (e.g. at EBI, UCL etc.) and a plethora of more minor resources. With limited funding it is tempting to focus on major resources which are most cited and most WWW clicked. Nevertheless, these resources tend to have other sources of funding, and the loss en masse of the minor resources would result in a major loss to the bioscience community and a major waste of previous investment.

With the limited funds available I would recommend a committee be constituted to manage the UK BBSRC bioinformatics resource base. Such a committee could give minor grants for resource maintenance (applied for on the basis of usage and utility and reviewed by peers and the committee).

Such a committee might consider it appropriate to manage the resource base more actively. A great deal can be gained through integration. An example of this is the Interpro database - which is now a one stop shop for this type of analysis, and is certainly much greater than the sum of its parts. Funding for a resource could be conditional on appropriate efforts to integrate with other resources. Coordination with the current grid pilot projects would be essential in this aim, but funding should not be limited to the current scope of those projects.

How should funding be used?

See above. Funding should certainly be only for publicly accessible resources, but not limited to those with previous BBSRC funding. I think active management of the resource base and integration is perhaps the key.

The components of sustainable support.

1. Active management.
2. Relatively small scale funding when needed for maintenance and minor development (major developements should be funded by normal grants).
3. Integration.

David Westhead, University of Leeds
Posted 17 February

Medical Research Council Human Genetics

I think it would be very helpful to the bioinformatics community for the BBSRC to provide flexible resources of the kind laterly offered by the MRC HGMP-RC (now sadly gone). Academic users could register with the HGMP-RC free of charge and have access to multi-processor UNIX servers for computationally intensive analyses. Through the servers users had access to current versions of the sequence databases, large 'scratch' areas for dumping hefty intermediate result files, and a range of widely used algorithms (BLAST, HMMER, EMBOSS, et al). Users could then roll their own perl (etc) scripts to develop their own custom-made analysis pipelines. I know many colleagues and graduate students who found this invaluable and there is no equivalent at the moment that I'm aware of.
What would be most unhelpful would be yet another fancy-schmancy web interface that does not provide a flexible (i.e. scriptable) environment.

Colin Semple, Head of Bioinformatics, MRC Human Genetics
Posted 14 February

University of Bristol

The BBSRC is very good at generating resources but less good at either maintaining them or providing facilities to give open access. As a result many newly created resources are either thrown away when space becomes limiting or they never become publicly accessible. It's equivalent to continually building roads and not allowing anybody to drive on them! Its easy to see why this situation exists; as such maintenance work is not hypothesis driven science it tends not to get funded by the Committees, but there is an argument to be made that such resources are even more important, as it such resources that provide the glue that holds the science together. It is therefore important that BBSRC finds some way to support such facilities, resources and expertise. The problem that I see is that this situation has gone on for so long that any initiative designed to address the problem would be grossly over subscribed and would I think lead to disappointment. The alternative is to make any initiative community driven; they should after all serve the community (if they don't then they should not be supported).

Keith Edwards, University of Bristol
Posted 2 February
