



Challenges for Sustainability: Perspectives and Experiences from e- Social Science

Rob Procter

Research Director

National Centre for e-Social Science

rob.procter@ncess.ac.uk



National Centre for eSocial Science

e-IRG Workshop April 2007



The University of Manchester



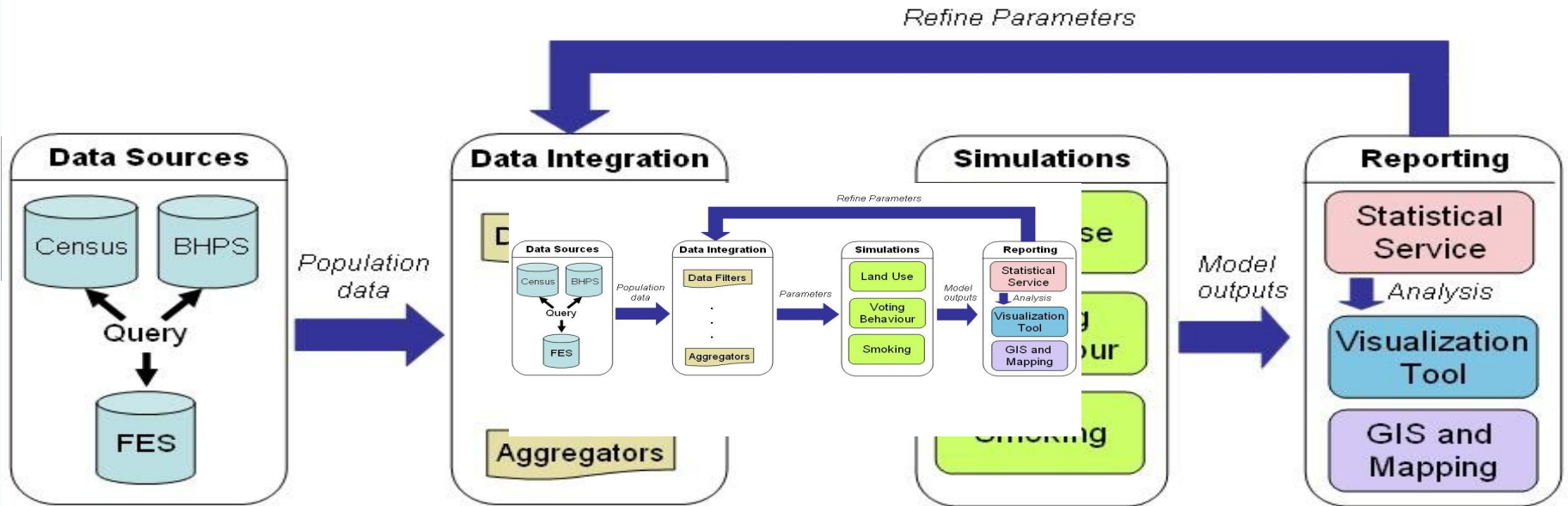
NCeSS Research Agenda

- Applications of e-Social Science:
 - Harnessing e-Infrastructure to tackle substantive problems in quantitative and qualitative research
 - Promoting innovation in research methods to tackle new and more complex social science problems

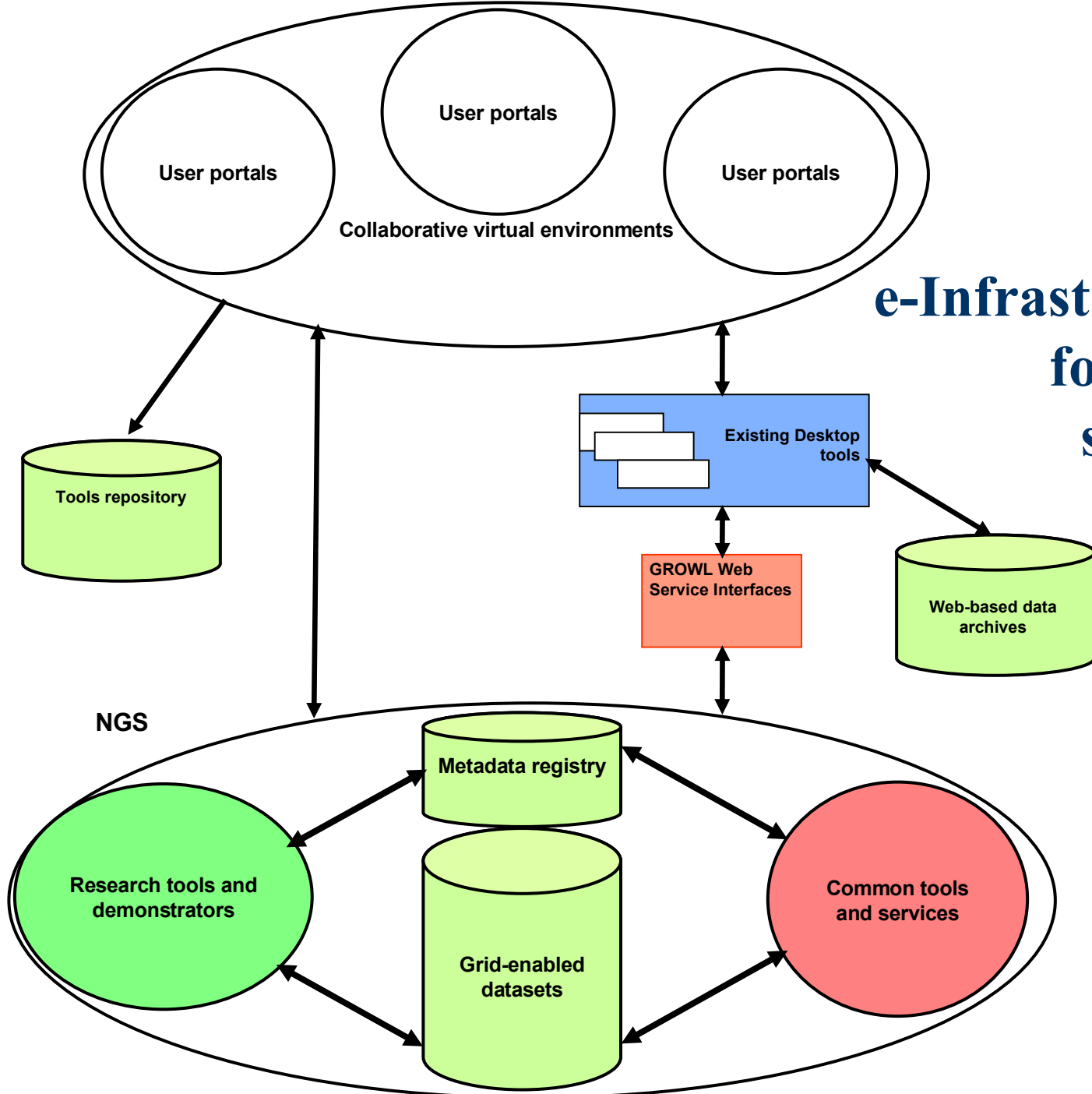
Supporting the Research Lifecycle

Review literature
and generate
hypothesis

Find datasets related to
proposed area of work



conclusions and discuss
with collaborators



e-Infrastructure for social sciences

NCeSS Research Agenda

- Socio-technical factors in design, uptake and use of e-Infrastructure:
 - How do we make it usable?
 - How do we engage with users?
 - How do we transfer successfully to new communities?
 - What are the research drivers?
 - How do we make e-Infrastructure relevant to the 'ordinary' researcher as well as those pursuing 'grand challenges'?
 - What are the barriers to growth and how are they to be overcome?
 - What distinguishes early and late adopters?
 - What are the impacts and broader implications?
 - What impact will e-Infrastructure have on scientific practices?
 - What are appropriate models for sustainability?
 - Is open source a credible supply mechanism?

Social Shaping Research

- Oxford e-Social Science Node: Ethical, Legal and Institutional Dynamics of Grid-Enabled e-Sciences (OeSS):
 - Aim is to understand the legal, ethical and social issues - such as confidentiality, privacy, data protection, intellectual property rights, accountability, trust and risk - raised by design and use of e-Infrastructures for scientific research
- Entangled Data: Knowledge and Community Making in e-(Social) Science:
 - This project investigated how and why groups of scientists do or do not collaborate using shared digital data sources.
- Accelerating Transition to Virtual Research Organization in Social Science (AVROSS):
 - The major goal is to understand reasons behind the low level of adoption of e-Infrastructures in the social sciences. Its main aim is to support optimisation e-Infrastructure developments and to provide guidance on how e-Infrastructures may be better deployed and exploited, by the social sciences and humanities research communities
 - international.fhso.ch/avross/avross

Social Shaping Research

- Adoption of e-Research Technologies:
 - The aim of this project is to study factors that may inhibit the wider diffusion and adoption of e-Research technologies and tools, and devise strategies for tackling them
 - http://www.nesc.ac.uk/esi/themes/theme_03/index.htm
- Barriers to the take-up of e-Infrastructure Services
 - The aim is to address barriers to the wider adoption of e infrastructure. It brings the expertise of a multi-disciplinary partnership that comprises leading members of the UK e-Science programme together with a wide variety of research communities in order to significantly improve our understanding of the barriers, and to devise and implement coordinated strategies to overcome them
- e-Infrastructure Use Cases and Service Usage
 - The aim is to articulate through the publication of use cases and the contribution of domain and Service Usage Models (SUMs), how the research community across different disciplines are actually or planning to engage with e-infrastructure. A deeper analysis of the use of existing e-infrastructure provision, both national and local, will also inform service development models which have hitherto been driven by the requirements of early adopters

e-Infrastructure: A Definition

- e-Infrastructure seeks to enable a decentralised research environment that:
 - Facilitates distributed collaboration
 - Provides incentives for participation at all levels
 - Encourages advancement of cross-boundary and interdisciplinary scholarship
- e-Infrastructure is a set of technical components, organisational practices and social norms that collectively provide for the smooth operation of collaborative, distributed scientific work.

Lessons from History

NSF Workshop on "History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructure" (2006):

- Initiation phase: technology and service innovations appear.
- Growth phase: technology transfer occurs across domains, often accompanied by technical diversity and competing systems. 'Reverse salients' are resolved. But early choices may constrain options available, leading to 'path dependence' - lock in effects, where inferior technologies become dominant - irreversibility, inefficiency. Innovations often fail when they attempt to transition rapidly from a small close-knit community of early adopters to a larger, more diverse community of novices.

Lessons from History

- Consolidation phase: mass adoption and convergence of standards. Network effects, where value increases exponentially with adoption may be a factor. Consolidation is complete when the service becomes a commodity resource.
- Splintering phase: since 1975, the model of monopoly utilities has been increasingly displaced by a deregulated, market-oriented approach.

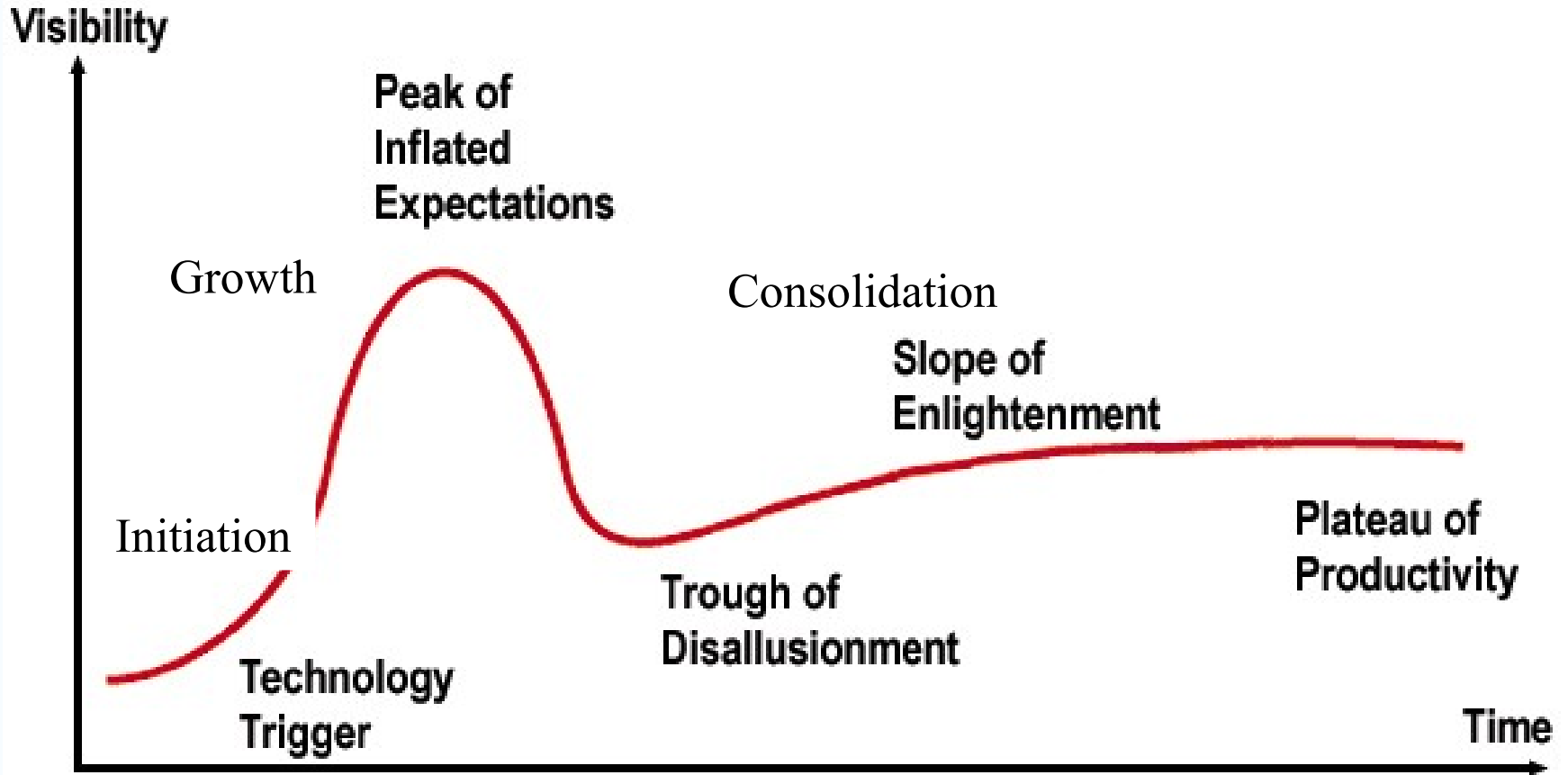
Lessons from History

- Successful infrastructures are a combination of 'top down' and 'bottom up' processes. They cannot be planned in any complete sense.
- They succeed because a stable socio-technical constituency - an ensemble of technical components (hardware, software, etc.) and stakeholders (people, interest groups, visions, values, etc.) - emerges.
- Socio-technical constituencies stabilise when stakeholders are able to strike a balance between their interests and those of the wider community.
- Each cycle of innovation is disruptive, there are winners and losers and socio-technical constituencies unravel.

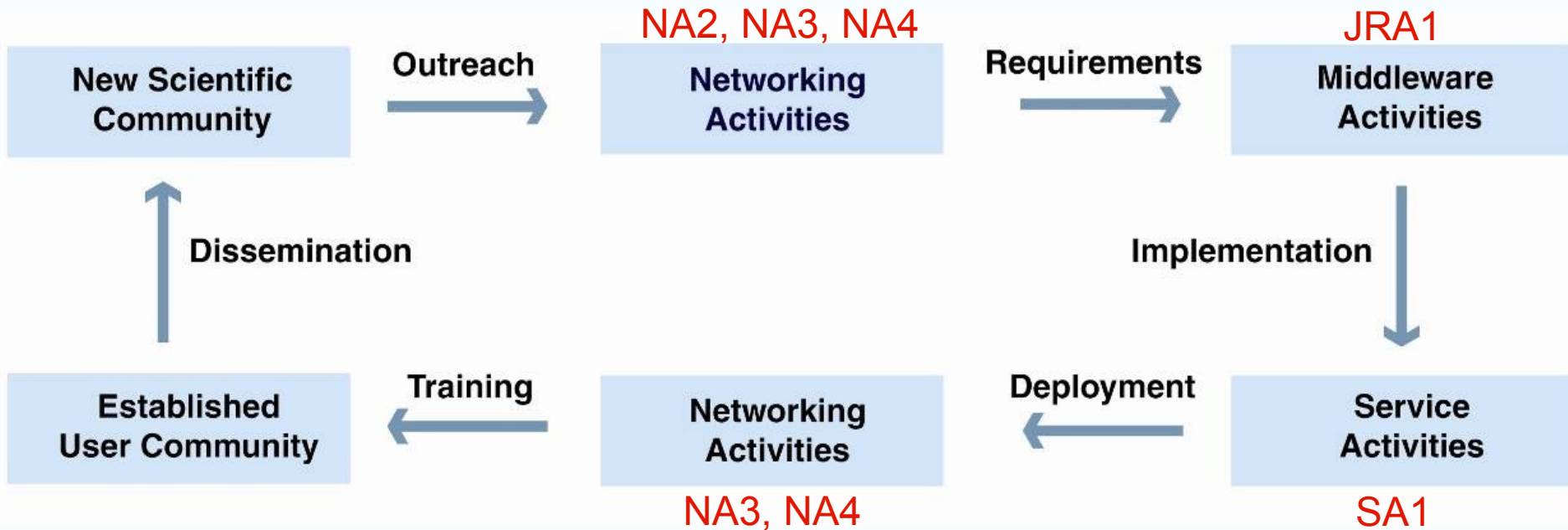
Lessons from History

- How does the current state of e-Infrastructure development map onto this historical model?
 - Competing standards - Grid, Web Services
 - Where are the reverse salients and path dependencies?
 - Is the Globus Toolkit a replay of OSI? Will more pragmatic solutions win?
- Are there aspects of e-Infrastructure which are not captured in the model?

The Gartner Hype Cycle



The EGEE Virtuous Cycle



Building effective user communities

A Possible Sustainability Model

- If we build it, early adopters will come and even help us build it.
- If we demonstrate its value, reduce the risks, lower the learning curve, etc., late adopters will also come.
- If everyone comes, then demand will ensure sustainability.
- But sustainability may be a risk for late adopters, so how do we square the circle?

Sustainability and the Social Sciences

- The social sciences has a long established practice in sharing data which has led to national repositories where researchers are required to deposit their data. These repositories are sustained by a mix of research council and other central funding.
- Sustainability of repositories has been a relatively straightforward issue. Stakeholders understand their role and what is expected of them; stakeholders are either resource providers or resource users.

Sustainability and the Social Sciences

- The sharing and re-use of resources which lies at the heart of the e-Research vision doesn't come for free.
- Proliferation in numbers and kinds of resources raises questions about how they can be sustained.
- For example, funding bodies are concerned that escalating commitments to sustain resources will consume an ever increasing proportion of their budgets.
- There are 'straws in the wind' suggesting that current funding arrangements will unravel.

Sustainability: A Problem of Misalignment?

- Existing mechanisms for curating resources so that they are and - remain - fit to share are incompatible with the new resource landscape (proliferating, heterogeneous) because they don't scale:
 - Greater community engagement is a potential solution but there are misalignments around research cultures, such as incentives and reward structures.
- Existing funding models are incompatible with the new resource landscape (proliferating, distributed, federated) because they are based upon a (few) providers-(many) users model:
 - We do not have a funding model - or the accounting and governance mechanisms to go with it - compatible with the VO concept. VOs do not align with the institutional and national entities which define the boundaries of current funding models.

Sustainability and Community Engagement

- It is difficult to imagine how new resources can be made sustained without engaging community effort on a large scale. What evidence have we that this might be achievable?
- Bloggers, wiki contributor and social bookmarkers in their millions are creating, editing and annotating content. New tools are changing who can collaborate and how.
- One factor is that Web 2.0 offers lightweight technical solutions which reduce effort and time to deployment, and so (to some extent) lessen risk.
- Can this model transfer into research with its different community demographics and cultures?



Welcome to myExperiment

Find a Taverna workflow, service or author:

..or browse: [authors](#), [tags](#), [services](#), [groups](#), [organisations](#), [projects](#), [research topics](#), [examples](#), [tutorials](#), [papers](#)

myExperiment is a collaborative platform for life scientists to manage, share and find [Taverna](#) workflows. You can find simple workflow [examples](#) showing Taverna functionality, or go deeper into [tutorials](#) showing best practice usage of most common [services](#) or [research topics](#). Other workflow [authors](#) have published their workflows and data, shared within their [groups](#) and [organisations](#). You can browse the aggregated work in funded or virtual [projects](#), such as [eScience UK](#), [myGrid](#) or [BioMoby](#).

If you [log in](#) or [register](#), you will get your own workspace for you and your affiliations. In your workspace, you can keep track of all your workflows and what data they have produced. You can arrange your workflows by your research topics and tasks, and browse related work by other researchers. You can share your workflows with your colleagues, and collaborate within your organisations and research topics.

At choice, you can [publish](#) selected workflows and/or data, making your work available for other myExperiment users. If you have published [papers](#) about your results, you can link up these with the related workflows, providing a resource page for your published work. At myExperiment.org, you can also create a public page to present your research interests and highlight workflows that represent your work.

Browse:

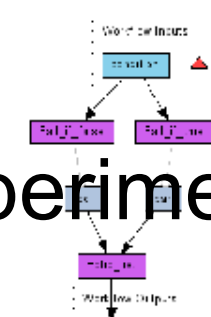
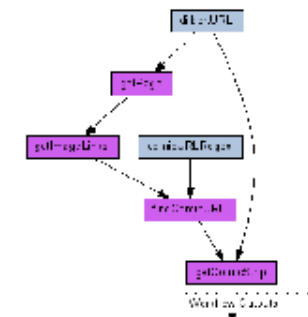
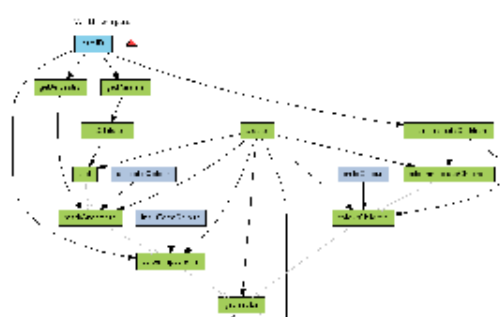
[Micro-array analysis](#)
[Sequence analysis](#)
[Sequence alignment](#)
[Protein sequencing](#)
...12 other [research topics](#)

[BioMart](#)
[BioMoby](#)
[SeqHound](#)
[SoapLab](#)
... 9 other [service providers](#)

[Alternate processor](#)
[Iteration strategy](#)
[Split and merge list](#)
[Nested workflow](#)
... 14 other [examples](#)

Selected workflows

These workflows have been highlighted by the myExperiment editors. You can suggest a workflow for highlighting by



myExperiment.org

Some Recommendations

1. Disseminate successful examples of e-Infrastructure-enabled research:
 - Fund more research into impact measurement.
2. Encourage community engagement:
 - Provide more funds for education and training.
 - Learn and borrow from domains such as Web 2.0.
3. More research on sustainability:
 - Work with stakeholders to define new funding models and the mechanisms to implement them.