The International DI Context: Digital changes everything …[1]

January 2014

[1] Prepared for DI Summit 2014 by the Project consulting team.

Contents

The International Environment – an Overview

There are large numbers of international reports on the impact of the data deluge and new technologies on the research environment from governments, research and scientific organizations and funding agencies. There are an equal number of different approaches to dealing with the emerging digital challenges. It would be a mistake to identify any one country as having "got it right" in terms of their approach to digital infrastructure; furthermore, any such action is context specific. However, most developed countries are further advanced than Canada in policy articulation, coordination, and customized delivery models. Essentially our competitor countries have been actively at work for well over a decade in addressing the issues of digital infrastructure.

Three things characterize the international actions that are reviewed below:

1. National DI initiatives have been framed by government policies – e.g. there is a significant "top down" approach that in turn engenders a bottom up response and engagement.
2. All recognize the need for multiple stakeholders to be engaged and the concomitant need for co-ordination.
3. From an early focus on physical infrastructure, the international DI discourse has increasingly used the lens of data as a national asset.

There are some respected structures, practices and approaches that could inform Canadian thinking as we go forward. A sample of these is highlighted below and described in more detail in the sections below. Of particular note are approaches that:

- Articulate a *vision, policy framework* and *dynamic strategy* for cyber-infrastructure, e-science, digital scholarship and research data infrastructure (e.g. EU Riding the wave; EU IRG Roadmap; UK Jisc; EU GRDI project, NSF Advanced Computing Infrastructure);

- Situate funding programs within *well-articulated policy frameworks* and guide by *multi-stakeholder coordination* (e.g. Australia);
- Promote *innovation* and multi-disciplinary approaches to e-infrastructure development (e.g. UK Jisc; UK e-science; US NSF);
- Designate a coordinating function for certain activities *among* delivery agencies (e.g. RCUK and UK EPSRC, NITRD in the US);
- Provide an overarching coordination mechanism *separate from* delivery functions (e.g. Australia, EU, PCAST in the US);
- *Incent institutions* to adopt improved practices in dealing with *research data* (e.g. Jisc and the UK Research Councils);
- Provide *new tools* and other forms of support for *tool development* for institutions and researchers (e.g. Australia ANDS; UK DCC);
- Provide *customized support* for a designated research communities (e.g. those just emerging into the computational intensive research space) and for exploration of new computational paradigms and technologies (e.g. Australia ANDS; UK Jisc; UK DCC);
- Focus on *capacity development* and training (e.g. UK DCC, NSF, NIH).

International Perspectives

Internationally, there has been an explosion of attention to the implications of digital technologies on the research enterprise and the concomitant evolution of policies, practices and programs of support for research.

> "Rapid and pervasive technological change has created new ways of acquiring, storing, manipulating and transmitting vast data volumes, as well as stimulating new habits of communication and collaboration amongst scientists. These changes challenge many existing norms of scientific behaviour."
>
> *Science as an Open Enterprise*
> *2012 Royal Society London*

The universal challenge is to realize the full value of data for research (as a producer and consumer of data), the economy and society more broadly.

> "To the extent feasible … digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze…"
>
> *Memorandum from the US Office of Science and Technology Policy*
> *February 2013*

Data have themselves become a critical national infrastructure; similarly the underpinning infrastructures of high speed networks, collaboration tools, computational systems, middleware, software, standards for interoperability, and skilled personnel who enable use of these systems are an increasingly critical national digital infrastructure.

> "eResearch infrastructure increases both the efficiency and effectiveness of research and opens up new and innovative research possibilities, across all disciplines including the ability to disseminate that knowledge. This reflects the fact that eResearch infrastructure is an integral part of modern research."
>
> "Planning for future eResearch infrastructure builds on the near ubiquitous deployment of high bandwidth research networks; on investments in, and excellent uptake of, high performance computing; and growing understanding and uptake of data management practices. The impact of collaboration tools and resources is being reinforced while the use of authorisation services is well established and growing."
>
> *2011 Strategic Roadmap for Australian Research Infrastructure*

Overarching Frameworks

While each nation frames its DI and e-research policies in its own context, the 2007 "Platforms for Collaboration" investment framework published in Australia provides a valuable example of how the underlying principles can be articulated in a simple and clear manner that in turn structures coordination and investment actions.

Developments in e-Research and cyber-infrastructure, both technological and social, are rapid and likely to continue for the foreseeable future. Hence, we can expect an increasing pace of change and an ongoing flow of new opportunities to enhance the quantity, quality and productivity of research efforts; noting always that some research is otherwise impossible and that the improvement in infrastructure and the ability to ask more demanding questions go hand in hand.

Combined with this pace of change, the independent goals and decision making processes of a multitude of governments and institutions will ensure that any future Australian e-Research infrastructure will be a combination of many activities with a need for significant co-ordination.

Therefore: A foundation principle is to enhance, strengthen and build on co-operative arrangements so that an increasingly coherent level of support can be provided to researchers, and their collaborations and communities.

The strengthening of e-Research activity and the pursuit of collaborative research within existing ICT intensive disciplines and more broadly within other disciplines hinges on the ease with which e-Research activities can be carried out.

Therefore: A foundation principle is to reduce barriers to adoption of e-Research by systematically supporting new e-Researchers as well as expert e-Researchers; by sourcing and supporting suitable tools and services.

e-Research can most usefully be contemplated as a permanent change in the way researchers work; so that the capabilities needed to carry out e-Research must be robust and enduring, reliable, always on, and commoditised when practical.

Therefore: A foundation principle is to provide robust and enduring services delivered by providers where the delivery of the service is the mission.

The funding available from the National Collaborative Research Infrastructure Strategy (NCRIS) for

Platforms for Collaboration (PfC) is significantly less than would be required to support all the Australian research which could benefit from e-Research services, so a means of defining priority is required.

Therefore: A foundation principle is to prefer e-Research infrastructure services which are of value across multiple research communities; and are of value to those communities

Platforms for Collaboration – Final Investment Plan 2007

Research Data Management

Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance.

Riding the Wave: How Europe can gain from the rising tide of scientific data
http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

From an early focus on physical infrastructure, the international discourse has increasingly turned its attention to focus on good stewardship and efficient utilization of data as a national asset – an asset for researchers, health, public policy, innovation, and society more generally. Both effective management of data through its life cycle and open access have become focal points of reports and policy statements. Examples follow.

In the US, the November 2013 report from the Council on Library and Information Resources entitled "Research Data Management - Principles, Practices, and Prospects" (http://www.clir.org/pubs/reports/pub160) notes that data management is one of the most important new strategic issues facing research universities, yet that very few universities have adopted policies for effective management of the data assets their researchers produce through public funding.

The European Global Research Data Initiative issued a report in 2012 on "Technological & Organisational Aspects of a Global Research Data Infrastructure: A view from the experts" that underscores the diversity and complexity of data managmnet http://www.grdi2020.eu/Default.aspx. A series of background expert papers analysed the state-of-the-art, the vision for the future, the challenges faced, and ways to tackle the challenges involved in different aspects of the data intensive environment: i) data analysis; ii) data discovery; iii) data interoperability; iv) data linking; v) data policy; vi) data preservation; vii) data provenance and trust; viii) data quality and curation; ix) data security; x) data storage; xi) data use – virtual research environments; xii) education and training in data handling and analysis at the interface between e-infrastructure and researchers and xiii) funding, sustainability, and governance.

RCUK in the United Kingdom has a released a short, but solid set of common principles on data http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx that deal, at a high level, with access, data stewardship, data author's rights and recognition.

The discussion on research data management is well advanced in Canada, albeit policy and practice has yet to catch up. Data Summit 2011 released a report with a comprehensive vision, principles and set of action items that are now being further advanced by Research Data Canada and associated organizations (see separate documents).

The February 2013 memorandum from the US Office of Science and Technology Policy stressed the importance of public *access* to the results of research – including research data:

> "The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.
>
> Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.
>
> Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather data underpins the forecasting industry, and making genome sequences publicly available has spawned many biotechnology innovations. In addition, wider availability of peer-reviewed publications and scientific data in digital formats will create innovative economic markets for services related to curation, preservation, analysis, and visualization. Policies that mobilize these publications and data for re-use through preservation and broader public access also maximize the impact and accountability of the Federal research investment. These policies will accelerate scientific breakthroughs and innovation, promote entrepreneurship, and enhance economic growth and job creation."

In Germany, the Alliance of Science Organizations[2] brings together the major funders, societies and research performing organizations dedicated to creating a sustainable, integrated, digital research environment that provides all German researchers with the broadest possible access to relevant published knowledge, research data and tools, no matter where or when they need it. Their coordination of activities and resources deals with: i) national licensing; ii) open access; iii) national hosting strategy; iv) research data; v) virtual research environments and vi) legal frameworks.

Global (Pan-National)

There have been a limited, but increasing, number of initiatives focusing on global approaches to e-infrastructure, in particular tackling the issue of research data which would appear to be the most intractable and difficult of infrastructural issues, especially in terms of integration, harmonization, and inter-operability. Prominent among these initiatives are:

- The **G8+O5 Global Research Infrastructure Sub-Group** on Data that issued a Draft Report, 28 October 2011 in which there is articulation of a future scenario for global data infrastructure and an identification of the challenges in achieving this vision in a short public document.

[2] Alexander von Humboldt Foundation, German Academic Exchange Service, Deutsche Forschungsgemeinschaft, Fraunhofer Society, Helmholtz Association of German Research Centres, German Rector's Conference, Leibnitz Association, German National Academy of Sciences Leopoldina, Max Planck Society, German Council of Science and Humanities.

- The **Global Research Data Infrastructures (GRDI)** project on *The Big Data Challenges* funded by the EU FP7. Its detailed 2012 report is designed to establish a framework for guiding the development of *ecosystems* of global research data infrastructures. Their concept of a digital science ecosystem is an interrelated set of data collections, services, tools, computations, technologies and communities of research. Two main challenges are addressed:
    - To effectively and efficiently support data-intensive science
    - To effectively and efficiently support multidisciplinary/interdisciplinary science

  The report outlines a strategic vision and addresses the technological (data), system, infrastructural, application, organizational and policy challenges that need to be addressed to achieve the vision. A series of ten recommendations are advanced (http://www.grdi2020.eu/Repository/FileScaricati/e2b03611-e58f-4242-946a-5b21f17d2947.pdf)
- The **FP7 iCORDI** program whose prime objective was to establish a coordination platform between Europe and the USA to discuss and improve the interoperability of today's and tomorrow's scientific data infrastructures of both continents and to extend this to the global levels.
- The **Research Data Alliance (RDA)** that aims to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be achieved through the development and adoption of infrastructure, policy, practice, standards, and other deliverables. Its first plenary meeting was held in March 2013. The initial founding bodies are i) Australian Commonwealth Government through the Australian National Data Service supported by the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative, ii) The European Commission through the iCordi project funded under the 7th Framework Program and iii) The United States of America through the RDA/US activity funded by the National Science Foundation. While Canada was invited to be a founding country, it declined (deemed premature at the time).

Other global initiatives promote the development of e-research capacity, e.g.:

- **Digging into Data** - Sponsored by ten international funding agencies (including Canada) aims to address how "big data" changes the research landscape for the humanities and social sciences. It supports a wide variety of projects that explore how computationally intensive research methods can be used to ask new questions about and gain new insights into our world.

European and National Actions

Europe

Europe has championed a strategic focus on cyber-infrastructure in its STI policies, with a particular focus on pan-European collaboration and integration. The EU has done much of the "heavy lifting" on the DI agenda, with varying levels of action from member and participating countries (note – Norway, Switzerland and Israel participate in a number of programs, although not formal EU members).

In a 2009 communication to the European Parliament, the EU highlighted the strategic role of "ICT infrastructures as a crucial asset underpinning European research and innovation policies" and called for a coordinated effort to foster world-class ICT infrastructures. Five domains of e-infrastructure were stressed:

- The pan- European high speed network
- National grid initiatives that provide a ‘one-stop-shop’ for a number of common grid-based services for their research communities
- Scientific data e-infrastructures – i.e. an ecosystem of European digital repositories, combining and adding value to national and discipline based repositories
- High performance computing
- Virtual Research Communities (VRCs) − structured international user communities

Since then there has been a large amount of action within the EU, including the publication of a “raft” of documents with policy and practice recommendations relating to e-infrastructure, including:

- 2010 – Report to the EU “Riding the wave: How Europe can gain from the rising tide of scientific data”
- 2011 – GEANT – “Knowledge without Borders”
- 2011 – Report on the Integration of Data and Publications (under 7th Framework Program)
- 2012 - EGI strategic plan developed through a large consortium, highlighting the need for a number of actions that support the “European Research Area” (EGI.eu was created in 2010 to coordinate and maintain a sustainable pan-European infrastructure to support European research communities and their international collaborators).
- 2012 - A Study for the EU on Authentication and Authorisation Platforms For Scientific Resources in Europe
- 2012 – An e-IRG[3] Study on Cloud Computing for research and science
- 2012 – Sienna – Roadmap on Distributed Computing Infrastructure for e-Science and Beyond in Europe

In addition, in 2012, the e-IRG issued a roadmap http://www.e-irg.eu/publications/roadmap.html that outlines Europe’s need for a single “e-Infrastructure Commons” for knowledge, innovation and science.

An example of multi-country collaborative action that has been triggered within the EU is the Knowledge Exchange (KE) partnership among Denmark, Germany, the Netherlands and the United Kingdom.[4] This partnership offers broad outlines for a possible action program for the four countries in a collaborative data infrastructure designed to enable researchers and other stakeholders from education, society and business to use, re-use and exploit research data to the maximum benefit of science and society. There is an explicit expectation of the involvement of all stakeholders from the scientific community in an action plan that targets four “drivers”:

- *incentives* and
- *training* in relation to researchers in their role as data producers and users of information infrastructures
- *infrastructure* and

[3] IRG = the e-Infrastructure Reflection Group. The e-IRG mission is to pave the way towards a general-purpose European e-Infrastructure.

[4] See http://www.knowledge-exchange.info/surfboard

- *funding* of the infrastructure in relation to further developments in *data logistics*.

The Horizon 2020 EU framework programs include an explicit priority for investment in research infrastructures, including e-infrastructures in the context of the larger objective of "developing the European research infrastructures for 2020 and beyond, fostering their innovation potential and human capital and reinforcing European research infrastructure policy".

- *Developing the European research infrastructures for 2020 and beyond* - The objective is to ensure the implementation and operation of the ESFRI and other world class research infrastructures, including the development of regional partner facilities; integration of and access to national research infrastructures; and the development, deployment and operation of e-infrastructures."
- *Fostering the innovation potential of research infrastructures and their human capital* - The goal is to encourage research infrastructures to act as early adopters of technology, to promote R&D partnerships with industry, to facilitate industrial use of research infrastructures and to stimulate the creation of innovation clusters. This activity will also support training and/or exchanges of staff managing and operating research infrastructures.
- *Reinforcing European research infrastructure policy and international cooperation* - The aim will be to support partnerships between relevant policymakers and funding bodies, mapping and monitoring tools for decision-making and also international cooperation activities.

The European research infrastructures policy has been driven within the European Research Area (ERA) partnership that was reinforced by the 17 July 2012 adoption by the Commission of the Communication on 'A Reinforced European Research Area partnership for Growth and Jobs' that called for structural changes across Europe in a partnership between Member States, Stakeholder Organizations and the Commission for a timely delivery of concrete measures to increase the level of excellence of Europe's public research system. The 2013 ERA Report on Progress calls for the following action in regard to "digital":

> "All Member States should ensure that conditions are in place to support: seamless online access to digital research services for collaboration, computing and accessing scientific information; the federation of electronic identities for researchers, which facilitates researchers' cross-border access to services and resources; and harmonised access and usage policies for e-infrastructures and digital research services in order to enable collaborations by multinational research consortia with both public and private partners."[5]

The pan-European approach has been reinforced by the work of the European Research Council (http://erc.europa.eu/

[5] Report from the Commission to the Council and the European Parliament. European Research Area Progress Report 2013. http://ec.europa.eu/research/era/pdf/era_progress_report2013/era_progress_report2013.pdf

The UK

E-infrastructure related activities in the UK are distributed among a number of organizations, including Jisc, the Research Councils, RCUK, and a number of centres, most notably the Digital Curation Centre (DCC).

Jisc[6] is a registered charity that works on behalf of UK higher education. It champions and plays a critical role in maintaining and enhancing the environment for digital scholarship/e-science in the UK, as well as in the integration of digital technologies in all aspects of higher education and research. Particularly valuable have been the incentives that Jisc has put in place for institutions to become actively engaged in the management of research data through targeted funds and in capacity development more generally. It also manages the high speed academic network *Janet* through a subsidiary.

The Engineering and Physical Sciences Research Council (EPSRC) currently manages the HECToR (High End Computing Terascale Resource) and holds a coordinating role for the UK Research Councils in HPC activities. Regional HPC centres and university clusters are connected by various grids providing a multi-tier capability.

The Research Councils UK (RCUK) provides an overarching role in catalyzing and supporting e-science. The recommendations of its 2009 review of the RCUK e-science initiative are reproduced in Appendix 1 and are indicative of the extent to which there is an overarching vision of how the overall ecosystem for e-science should develop. In addition, the Research Councils have developed more comprehensive policies in regard to data management and sharing than most nations, although this is a rapidly evolving field.

The U.K. Data Curation Centre (DCC) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community. With a stronger R&D focus than the internal operations of Jisc, it complements the Jisc activities through developing tools, guidelines and policies that facilitate institutional and researcher curation initiatives. The text box provides an overview of its genesis and activities.

Digital Curation Centre (DCC) http://www.dcc.ac.uk/

Purpose - a national centre for solving challenges in digital curation that cannot be tackled by any single institution or discipline. Digital curation is maintaining and adding value to a trusted body of digital research data for current and future use; it encompasses the active management of data throughout the research lifecycle.

Start-up - Launched in March 2004, following a successful response to JISC request for proposals.

[6] Historically, JISC stood for Joint Information Systems Committee but over the last decade it has evolved and as a company is now known as Jisc.

DCC Management structure - a consortium comprising the University of Edinburgh and its partners at UKOLN[7] (University of Bath) and HATII (University of Glasgow

Phase 1 – March 2004-February 2007. Projects and provision of generic services, some development activity and research. Target groups – i) those engaging in digital preservation and curation activities within UK Higher and Further Education (e.g. data specialists, records managers, librarians, archivists, researchers as data creators, and policy-makers and ii) the public and commercial sectors, international 'sister organisations' and standards working groups
Phase 2 – March 2007-February 2010. Building on phase 1, activity shifted towards increased and direct involvement with the active research community, as exemplified by the creation of an e-Science Liaison function and the conduct of immersive discipline case studies.
Phase 3 - March 2010 - February 2013. A shift away from the development of curation tools and a renewed focus on building capacity, capability and skills for data curation across the UK's higher education research community. This includes a series of 60 day engagements with institutions that include helping to make the case for good research data management practice, through the diagnosis of current practice and identification of forward priorities, to the use of DCC tools and training for building lasting data management strategies and infrastructure.

Resources and services - Briefing Papers, How-to Guides, Developing RDM Services, Curation Lifecycle Model, Curation Reference Manual, Policy and legal, Data Management Plans, Tools, Case studies, Repository audit and assessment, Standards, Publications and presentations, Roles, Curation journals, Informatics research.

On the other hand, there have been some significant failures in the ongoing support of DI in the UK, *inter alia*, what has been deemed as premature "unfunding" of the Arts and Humanities Data Service in 2008 as described in a report by Susan Hockey and Seamus Ross (http://www.methodsnetwork.ac.uk/publications/final-report.html) :

> "…demonstrated not only that ICT methods and tools are central to humanities scholarship, but also that there was 'a very long way to go before ICT in humanities and arts research finds its rightful and needed place'. The investment in ICT in the arts and humanities needs to be much greater and it needs to reflect better the particularities and needs of individual communities. Researchers who do not have access to the most current technological methods and tools will not be able to keep pace with the trends in scholarship. There is a real need for support and infrastructure for distributed research."

There is room for learning from both the successes and failures of other countries.

[7] Formerly known as The United Kingdom Office for Library and Information Networking, UKOLN is a centre of expertise in digital infrastructure, information policy and data management

The United States

The US commitment to cyberinfrastructure is of long standing. The 1991 High-Performance Act provided for a "coordinated Federal program to ensure continued United States leadership in high-performance computing." The Next Generation Internet Research Act of 1998, and the America COMPETES Act of 2007 extended the scope and federal commitment to strategic investment. Over time, the agenda has extended to embrace the impacts of the "data deluge" in all fields of research and the need for capacity building.

A seminal report was issued in 2003 by a "Blue Ribbon Committee" appointed by NSF to chart the way forward. That report continues to be a point of reference for the myriad of reports that have emerged since that time. Among the key observations were:

> "The emerging vision is to use cyberinfrastructure to build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments and that operate at unprecedented levels of computational, storage, and data transfer capacity. "
>
> "…This vision of science and engineering research involves significant educational dimensions. The research community needs more broadly trained personnel with blended expertise in disciplinary science or engineering, mathematical and computational modeling, numerical methods, visualization, and the sociotechnical understanding about working in new grid or collaboratory organizations."
>
> "…We propose a large, long-term, and concerted new effort, not just a linear extension of the current investment level and resources."
>
> "We estimate that the new funding will be distributed into four coordinated areas:
> - fundamental research to create advanced cyberinfrastructure ($60M);
> - research on the application of cyberinfrastructure to specific fields of science and engineering research ($100M);
> - acquisition and development of production quality software for cyberinfrastructure and supported applications ($200M);
> - provisioning and operations (including computational centers, data repositories, digital libraries, networking, and application support) ($660M).
>
> These are recurring annual figures."

As might be expected, given the complexity of the US S&T funding and policy worlds, many agencies are active in funding or exploring policy issues in this area, among them the National Science Foundation (NSF), the National Institutes of Health (NIH), the National Endowment for the Humanities (NEH), the Department of Energy (DoE), the National Oceanographic and Atmospheric Agency (NOAA), the Department of Defence (DoD), the President's Advisory Council on Science and Technology (PCAST), the National Academies of Science (NAS), the American Council of Learned Societies (ACLS) and higher education institutions themselves (e.g. through their involvement in Internet2 and in hosting some of the larger cyberinfrastructure resources). Fourteen federal agencies are grouped together under the umbrella of the Networking and Information Technology Research and Development (NITRD) Program that also provides a high level coordinating mechanism, reporting through the Office of Science and Technology Policy (OSTP) and the National Science and Technology Council (NSTC) to the President.

The following text box captures some of the key investments in the technological cyberinfrastructure in the US:

Key infrastructure components - USA

DoE

- The Energy Sciences Network, or *ESnet*, is a high-speed network serving thousands of Department of Energy researchers and collaborators worldwide. Managed and operated by the ESnet staff at Lawrence Berkeley National Laboratory.. ESnet derives its effectiveness from the extensive cooperation it enjoys with its user community
- Oak Ridge National Laboratory Leadership Computing Facility (***OLCF***) - Home to Jaguar, a Cray XK6 capable of 3.3 thousand trillion calculations a second—or 3.3 petaflops—the OLCF combines world-class staff with cutting-edge facilities and support systems. . With 18,688 16-core AMD Opteron processors, 960 NVIDIA accelerators, 600 terabytes of memory, a 10-petabyte file system, and input/output bandwidth of 244 gigabytes per second, Jaguar is the United States' fastest and most powerful supercomputer dedicated to open scientific research.
- Argonne Leadership Computing Facility (***ACLF***) - The ALCF provides the computational science community with a world-class computing capability dedicated to breakthrough science and engineering. The research being conducted at the ALCF spans a diverse range of scientific areas - from studying exploding stars to designing more efficient jet engines to exploring the molecular basis of Parkinson’s disease. The ALCF teams provide expertise and assistance to support user's projects to achieve top performance of applications and to maximize benefits from the use of ALCF resources.
- National Energy Research Scientific Computing (***NERSC***) Center - As a national resource to enable scientific advances to support the missions of the Department of Energy's Office of Science, the National Energy Research Scientific Computing Center (NERSC), operated by the Lawrence Berkeley National Laboratory, annually serves approximately 3,000 scientists throughout the United States.

NSF

- ***Blue Waters*** - operated by the University of Illinois. The Blue Waters supercomputer provides sustained performance of 1 petaflop on a range of real-world science and engineering applications. It is one of the most powerful supercomputers in the world. US researchers use the power of Blue Waters to tackle a wide range of challenging problems, from predicting the behavior of complex biological systems to simulating the evolution of the cosmos.
- The ***eXtreme Digital (XD***) program consisting of:
  - The Extreme Science and Engineering Discovery Environment (XSEDE) - responsible for integration of XD tier shared resources and services. XSEDE supports 16 supercomputers and high-end visualization and data analysis resources across the country. Its digital services, meanwhile, provide users with seamless integration to NSF's high-performance computing and data resources. XSEDE's integrated, comprehensive suite of advanced digital services federates with other high-end facilities and with campus-based resources, serving as the foundation for a national cyberinfrastructure ecosystem. Common authentication and trust mechanisms, global namespace and filesystems, remote job submission and

monitoring, and file transfer services are examples of XSEDE's advanced digital services. XSEDE's standards-based architecture allows open development for future digital services and enhancements. XSEDE also provides the expertise to ensure that researchers can make the most of the supercomputers and tools.
  - Technical Insertion Service - Evaluates and makes recommendations on insertion of software and other technologies into the XD environment

  - Technical Audit Service - Provides metrics on XD systems and operates XDMoD a publically available and easily usable tool for extracting data and monitoring XD systems .

Internet2

- A not-for-profit organization governed by an executive Board of Trustees representing a diverse membership of 220 U.S. universities, 60 corporations, 70 government agencies, 38 regional and state education networks, more than 65 national research and education networking partners representing over 100 countries
- The upgraded network uses 100 Gigabit Ethernet technology and will have 8.8 Terabits of capacity. The Internet2 Network is the *first national network* (in the US) to deploy 100 GigE waves on its entire footprint.
- The Internet2 NOC is supported by the Indiana University Global Network Operations Center (GlobalNOC), which provides world-class production and research and development support services to Internet2 Network participants

Recent reviews (2010 and 2013) of the network and information technology (NIT) R&D program (NITRD) by the President's Council of Advisors on Science and Technology (PCAST) have observed that the program has been effective in enabling a vibrant NIT industry, and has fostered seminal discoveries and advanced new technologies that are helping to meet many societal challenges. Those reports note, however, the difficulty of obtaining robust data on investments and the challenges of coordination among the various agencies. The 2013 PCAST report on NITRD also notes both the importance of high level strategy development and the increasing need for such action:

> "Determining a national strategy for R&D investment requires not only the dedication of people at the highest levels of government, but also sustained discipline-specific advice from leading experts in academia and the private sector. The 2010 PCAST NITRD report recommended the establishment of a broad high-level standing committee to provide this guidance. Since 2010, the need has only grown."

Glimpses of some of the issues and approaches being taken by the key funding agencies are captured in the following text boxes.

The National Science Foundation (NSF) 2013 - Advanced Computing Infrastructure

Vision - NSF will be a leader in creating and deploying a comprehensive portfolio of advanced computing infrastructure, programs, and other resources to facilitate cutting-edge foundational research in computational and data-enabled science and engineering (CDS&E) and their application to all disciplines. NSF will also build on its leadership role to promote human capital development and education in CDS&E to benefit all fields of science and engineering.

Strategic Directions

1. Foundational research to fully exploit parallelism and concurrency through innovations in computational models and languages, mathematics and statistics, algorithms, compilers, operating and run-time systems, middleware, software tools, application frameworks, virtual machines, and advanced hardware.
2. Applications research and development in use of high- end computing resources in partnerships with scientific domains, including new computational, mathematical and statistical modeling, simulation, visualization and analytic tools, aggressive domain-centric applications development, and deployment of scalable data management systems.
3. Building, testing, and deploying both sustainable and innovative resources into a collaborative ecosystem that encompasses integration/coordination with campus and regional systems, networks, cloud services, and/or data centers in partnerships with scientific domains.
4. Development of comprehensive education and workforce programs, from deep expertise in computational, mathematical and statistical simulation, modeling, and CDS&E to developing a technical workforce and enabling career paths in science, academia, government, and industry.
5. Development and evaluation of transformational and grand challenge community programs that support contemporary complex problem solving by engaging a comprehensive and integrated approach to science, utilizing high-end computing, data, networking, facilities, software, and multidisciplinary expertise across communities, other government agencies, and international partnerships.

NSF – A domain specific initiative for data management and access - DataONE - Data Observation Network for Earth http://www.dataone.org/about (NSF Grant #ACI-0830944)

DataONE is designed to be the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data. Its goal is to ensure the preservation, access, use and reuse of multi-scale, multi-discipline, and multi-national science data via three primary cyberinfrastucture elements and a broad education and outreach program.

Cyberinfrastructure elements:

- Coordinating nodes that provide network-wide services to enhance interoperability of the Member Nodes and support indexing and replication services
- Member nodes comprise a distributed network of data centers, science networks or organizations. These organizations can expose their data within the DataONE network through the implementation of the DataONE Member Node service interface. In addition to scientific data, Member Nodes can provide computing resources, or services such as data replication, to the DataONE community.
- Investigator Toolkits enable access to customized tools that are familiar to scientists and that can support them in all aspects of the data life cycle.

Community Engagement and Outreach

- Working groups focus on identifying, describing, and implementing the DataONE cyber-

infrastructure, governance, and sustainability models.
- User groups provide the opportunity for funders, users, developers, educators or any other stakeholder to gather and discuss the advancement of DataONE products and services.
- Education that spans formal graduate-level training in research and cyber-infrastructure development, to developing informal inquiry-based education modules that allow students of all ages to ask their own specific questions.

National Institutes of Health (NIH)

Recommendations of the Data and Informatics Working Group Report to The Advisory Committee to the Director 2012

The DIWG recommends that the NIH should invest in technology and tools needed to enable researchers to easily find, access, analyze, and curate research data. NIH funding for methods and equipment to adequately represent, store, analyze, and disseminate data throughout their useful lifespan should be coupled to NIH funding toward generating those original data. The NIH should also increase the capacity of the workforce (both for experts and non-experts in the quantitative disciplines), and employ strategic planning to leverage IT advances for the entire NIH community. The NIH should continue to develop a collaborative network of centers to implement this expanded vision of sharing data and developing and disseminating methods and tools. These centers will provide a means to make these resources available to the biomedical research community and to the general public, and will provide training on and support of the tools and their proper use.

The NIH Response 2013

To maximize utilization of the vast amounts of biomedical data and information that are being amassed, NIH has started to develop a series of activities grouped under its Big Data to Knowledge (BD2K) initiative that would create:

- improved data and software sharing policies, catalogs of research data, and data/metadata standards development to facilitate broader use of biomedical big data
- analysis methods and software development and dissemination
- enhanced training for biomedical big data
- proposed new centers of excellence

It has also indicated that it will launch the NIH InfrastructurePlus adaptive environment to advance high-performance computing, agile hosting and storage approaches, and modernization of the network, among other approaches.

For nearly all of the new data and informatics activities, workshops will be held in FY 2013 to refine the agency's plans and determine the data and informatics needs of the scientific community, with implementation scheduled to begin in FY 2014. As a first step toward developing a set of recommendations, a BD2K working group has issued a request for information (RFI) on the short- and long-term training needs of individuals who work with biomedical data. The group is also seeking examples of programs or strategies to cross-train scientists at all career levels as well as comments on evaluating workforce skills and knowledge and developing a diverse research workforce.

The US has invested heavily in computational facilities and high speed networks – both regional and national. Much of the public discussion has focused on the benefits of such investments for industrial competitiveness, the requirements of big science for extreme computing and more recently data intensive life sciences. The US has had a long tradition of strong institutionally based repositories for longitudinal social sciences data, often supported by provide foundations. More recently, the 2006 report of the American Council of Learned Societies *Our Cultural Commonwealth* and the NEH-led international initiative *Digging into Data* are important additions to our appreciation of the evolving needs of the digital environment.

Our Cultural Commonwealth http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf 2006

Vision – the transformative potential of an improved cyberinfrastructure with respect to the preservation and availability of our cultural heritage.

Five goals for an effective cyberinfrastructure – it should:
1. be accessible as a public good;
2. be sustainable;
3. provide interoperability;
4. facilitate collaboration;
5. support experimentation.

Recommendations
1. Invest in cyberinfrastructure for the humanities and social sciences, as a matter of strategic priority.
2. Develop public and institutional policies that foster openness and access.
3. Promote cooperation between the public and private sectors.
4. Cultivate leadership in support of cyberinfrastructure from within the humanities and social sciences.
5. Encourage digital scholarship.
6. Establish national centers to support scholarship that contributes to and exploits cyberinfrastructure.
7. Develop and maintain open standards and robust tools.
8. Create extensive and reusable digital collections.

Australia

The Australian e-infrastructure ecosystem has many of the essential components, the result of a concerted national initiative that had its genesis in extensive national discussions on research in the early to mid-2000s and a commonwealth government commitment to world class cyber-infrastructure to underpin its research and innovation agenda. Triggered by a five-year innovation plan "Backing Australia's Ability (BAA)" in 2001, there was serious background research and national discussion on what was required to create world class research infrastructures, including a particular focus on digitization and e-infrastructures that are designated as two critical enabling capabilities for Australia. The various initiatives that were launched – and that have had a profound impact on the overall research environment – were framed by sequential policy documents emanating from government following transparent national consultative processes.

There has developed a sophisticated (and somewhat complex) management, coordination and oversight system that includes the designation of certain universities to manage a national centre or program. Observers of the system have considerable respect and recognition for the frontier initiative of the Australian National Data Service that has provided leadership and capability. There are, however, some concerns with i) the problems inherent in asking individual institutions to take the lead role in managing major national programs and ii) the decoupling of the network, high performance computational and storage facilities and the national data service.

Australian e-Infrastructure Initiatives		
Initiative	***Date launched and funding***	***Description***
The National Collaborative Research Investment Strategy (NCRIS)	2007	Publication of *Platforms for Collaboration (PfC) – Final Investment Plan* that laid out principles, policies, and priorities for strengthening e-research capacity and activity in Australia. This triggered many of the e-initiatives in following years.
The National Computational Infrastructure **(NCI)**	2007	This high end computing facility is an initiative of the Australian Government, hosted by The Australian National University. NCI's mission is to foster ambitious and aspirational research objectives, and to enable their realisation through world-class high-end computing services. NCI's advanced computing infrastructure, comprising a petascale HPC system, a large-scale compute cloud (primarily for data-intensive services), and multi-petabyte high-performance storage, is funded through programs of the Department of Industry, Innovation, Science, Research and Tertiary Education, while its operations are sustained through the substantial co-investment by a number of partner organisations including ANU, CSIRO, the Australian Bureau of Meteorology, Geoscience Australia, a number of Australia's research-intensive universities, and the Australian Research Council. The work of NCI is structured into two main activities: Facilities and Services, and Planning, Access, Policy and Outreach.
The Pawsey Project (Supercomputing facilities - petascale)	May 2009 $80 million	As part of the "Super Science Initiative" in May 2009 the Commonwealth Government allocated $80 million to Western Australia's supercomputing hub iVEC to establish a petascale supercomputing facility. The aim is to enhance Australia's position in the international supercomputing field and to boost its bid for the Square Kilometre Array Project. The three components of the Pawsey project are a $5 million Linux cluster at Murdoch University (called 'Epic'), a $4 million hybrid-GPU cluster at the University of Western Australia (called 'Fornax') and a $33 million Cray supercomputer for Perth's Pawsey Centre for general research and astronomy

		data processing. The Cray Cascade system will initially be capable of 0.3 petaflops in 2013, expanding to "more than 1.2 petaflops" when a second phase of the machine is rolled out in 2014.
Australia’s Academic and Research Network **(AARNet)**	1989	AARNet was established in 1989 by a group of Australian universities and research institutions, with the aim of providing high-speed internet connectivity to serve the academic and research community. Today, AARNet serves 38 Australian universities and the CSIRO, as well as a number of health and other research institutions, cultural organisations, schools, vocational training colleges and specialist content providers.
The Australian Research and Education Network **(AREN)**	2002 $88 million	AREN is the high bandwidth communications network managed by AARNet and including access to a high bandwidth trans-Pacific backbone for research and education. The Commonwealth government has committed up to $88 million since 2002 to contribute to the AREN's development including acquiring access to a high bandwidth trans-Pacific backbone for research and education
The National Research Network **(NRN)**	2010 $37 million	NRN is a project extending and upgrading AREN through an investment of $37 million of Commonwealth government funds through the Education Investment Fund (EIF) to be partnered with other participants in in eight component projects. In August 2010 the government announced the appointment of the University of South Australia as Lead Agent for the Project.
The Research Data Storage Infrastructure **(RDSI)** Project	2011 $50 million	RDSI is hosted by the University of Queensland; $50M funded from the Education Investment Fund. Four primary nodes have been announced by early 2013 – one to be established by Intersect in Sydney, a Brisbane node at the Queensland Cyber infrastructure Foundation, the ANU will establish a node in Canberra and eResearch SA will set up a node in Adelaide. An additional node will be established by the University of Tasmania. These are the first of eight to ten nodes intended to underpin the national storage network, which by 2014 will offer Australian researchers access to around 100 petabytes of data collections. Nodes in the RDSI will be connected by AARNet but using a separate wavelength to other data traffic on the network to avoid the transmission of large data collections “flooding” the AARNet backbone.
The Australian National Data Service **(ANDS)**	Launched in 2008 and extended in 2009 ~30 million	ANDS was created by Australian government to o make Australian research data collections more valuable by managing, connecting, enabling discovery and supporting the reuse of this data o enable richer research, more accountable research; more efficient use of research data; and improved

		provision of data to support policy development Monash University leads the ANDS partnership with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).
The Australian Research Data Commons **(ARDC)**	2009 +	is being built by ANDS to make better use of Australia's research data outputs through: ○ a set of data collections that are shareable ○ descriptions of those collections ○ an infrastructure that enables populating and exploiting the commons ○ connections between the data, researchers, research, instruments and institutions.
Research Data Australia **(RDA)**		RDA is ANDS' flagship service and provides a comprehensive window into the Australian Research Data Commons. It is an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines.
The National eResearch Tools And Resources **(NeCTAR)** Project	2011 $47 million (commonwealth) $54 million (partners co-investment)	NeCTAR enhances frequently used research tools, develops exemplar digitally enabled laboratories and establishes virtual server and cloud infrastructure able to support the next generation of research 'apps'. In 2011 $47 million was provided to support the University of Melbourne led NeCTAR Project. In addition, the Australian Research sector has committed $54 million as co-investment to the NeCTAR project, resulting in NeCTAR injecting $101 million to Australia's research infrastructure
Oversight bodies : • Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education (**DIISRTE)** • The Australian eResearch Infrastructure Council **(AeRIC)** • The Australian Research Committee **(ARCom)**	Structural developments 2011-2012	Oversight is provided by the core government department and two advisory bodies: • Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education (**DIISRTE)** ○ The Australian eResearch Infrastructure Council **(AeRIC)** provides an overseeing guide to the e-infrastructure programs. In a 2012 Forum, AeRIC discussed the growth in eResearch infrastructure and services, what is available now and what is expected to come online in to 2013, lessons learned during implementation and how these experiences are informing future planning. ○ The Australian Research Committee **(ARCom)** was created in 2011-12 to provide integrated and strategic advice on investment across the science, research and innovation system, including in the areas of human capital, infrastructure and collaborative activities. It provides advice to the government via the Minister for DIISRTE