Summary Report – Digital Infrastructure Summit 2014

Conclusions of Summit 2014, held in Ottawa on January 28-29, in Ottawa, Ontario

Prepared by the Leadership Council for Digital Infrastructure

in Association with

the Study Team of Janet Halliwell, J.E. Halliwell Associates Inc., Dennis Rank, Dennis Rank and Associates Inc., and Greg Jodouin, Pace Public Affairs and Community Engagement

March 4, 2014

The Leadership Council January 2014

Co-Chairs
- Steven Liss, Vice Principal Research, Queens University
- Jay Black, Chief Information Officer, Simon Fraser University

Honorary Co-Chairs
- David Barnard, President, University of Manitoba
- Chad Gaffield, President, Social Sciences and Humanities Research Council (SSHRC)

Participants
- Jim Ghadbane, President and CEO, CANARIE
- Kathryn Anthonisen, Vice President, CANARIE
- Jill Kowalchuk, Interim President, Compute Canada
- Pam Bjornson, Chair, Research Data Canada (RDC)
- Walter Stewart, Coordinator, Research Data Canada (RDC)
- Gerald Beasley, Chair Board of the Canadian Association of Research Libraries (CARL)
- David Baker, Executive Director, CASRAI
- Clare Appavoo, Executive Director, Canadian Research Knowledge Network (CRKN)
- Donna Bourne-Tyson , Board Member, CRKN
- Geoffrey Rockwell, Researcher, University of Alberta
- Ray Siemens, Director, Research Dissemination, Federation for the Humanities and Social Sciences (CFHSS)
- Robbin Tourangeau, Senior Director, Strategic Initiatives, Council of Ontario Universities (COU)
- Lori MacMullen, Executive Director, Canadian University Council of Chief Information Officers (CUCCIO)

Observers
- Robert Davidson, Vice President, Programs and Planning, Canada Foundation for Innovation (CFI)
- Jane Aubin, Chief Scientific Officer/Vice-President of Research, Canadian Institutes for Health Research (CIHR)
- Jonathan Faulkner, Manager, Strategic Investment Planning and Reporting, CIHR
- Ted Hewitt, Executive Vice-President, Social Sciences and Humanities Research Council (SSHRC)
- François Simard, Manager, Research Portfolio, SSHRC
- Kevin Fitzgibbons, Associate Vice President, Corporate Planning and Policy (Interim), Natural Sciences and Engineering Council (NSERC)
- Philippe-Olivier Giroux, Policy Analyst, Policy and International Relations NSERC:
- Nathaniel Valk, Policy Analyst, Policy and International Relations NSERC:
- James Fulcher, Director, Policy and Strategic Planning, SITT, Industry Canada
- Briand Garland, Policy Analyst, Policy and Strategic Planning, SITT, Industry Canada

Acknowledgements

Table of Contents

Executive Summary - DI Summit 2014

DI Summit 2014

On January 28-29, 2014 in Ottawa, the Leadership Council (LC) for Digital Infrastructure (DI) conducted the Digital Infrastructure Summit 2014. The roughly 140 participants included representatives from universities, researchers, service providers, industry, and funding bodies. The intent was to chart the DI path forward and address shared roles and responsibilities, and thus was a working meeting in which concrete action items were developed and discussed.

What is DI?

There has been a cultural shift in how digital "infrastructure" is conceived. Instead of being considered as only machines that compute the data and high speed networks that allow access to the data ("machines and pipes"), it is now seen as the entire ecosystem that surrounds our ability to capture, manage, preserve, and use data – the core element of digital infrastructure (DI). In other words, **data are infrastructure,** as are **the highly skilled personnel** who facilitate access to data, computational power and networks. The Canadian advanced DI ecosystem will be an integrated and interrelated set of facilities, services and capacities that provide Canada with the capability to conduct data-intensive, top-level research, and to intensify collaborations nationally and internationally. DI users will be drawn from all research communities, including academic, government and private sector.

With data and the infrastructure for data stewardship at the core of the DI ecosystem, these will be supported by the hardware, software and people that provide single-site, distributed and nationally centralized computational, networking and analytical tools.

The Underlying Problems

Three underlying problems framed the Summit discussion of approaches and solutions.

1. **Research data are a national asset, but not treated as such.** Research data were developed through significant investments by Canadian governments over many decades and in many fields. Modern research increasingly results in massive data sets (e.g., in particle physics or astronomy), but there are also innumerable smaller and highly diverse datasets that together represent equally large amounts of valuable data. These data were produced with the intent of being used to benefit Canadians, but are often not preserved or are inaccessible to potential users. For better stewardship of this national asset we need research data management (e.g., curation, preservation, access provisions); data storage infrastructure; the provision and training of skilled knowledge management personnel; and relevant software and middleware development.

2. **Inadequate governance/coordination.** The Canadian DI ecosystem is a patchwork quilt of many players, with insufficient coordination. The players include the funders of research (the TC3+), the DI service providers, and Canada's research enterprise, including universities, professional

associations, researchers, academic libraries, national standards organizations, and other user organizations. Developing clarity of roles and responsibilities and coordinating funding investments are two key actions to move us forward. Otherwise, we risk fragmented approaches, and sub-optimal socio-economic returns on our extensive research investments.

3. ***Lack of an overarching federal policy and planning framework.*** Rapid technological change has resulted in new ways of acquiring, storing, manipulating, and transmitting vast volumes of data, and radical improvements in our ability to extract insights from those data and innovate with them. To optimize Canadian benefits we need an overarching policy framework that articulates the central role of DI in the research landscape and the need for system-level adjustments to the new DI environment, including a structure approach to coordination and integration, appropriate planning and investment horizons, and refined support structures for all DI elements.

The Time for Action is Now

Canada needs a comprehensive, integrated, sustainable DI ***ecosystem*** – recognizing the interrelated, interdependent nature of DI components and supported by a national DI vision to foster excellence in research, education, innovation, and policy. This vision will be driven by the reconceptualization of ***data as infrastructure***, along with the supporting human and financial resources needed for their effective and efficient use across disciplines and sectors.

The need is seen to be pressing, to the point that immediate action – even if not perfect – is much preferred over more discussion and consultation. Canada's approach to managing data as a national asset should be:

- Collaborative, coordinated, clustered, and horizontal – a different way of doing things, not just a better way of doing things, and certainly not another silo;
- Experimental, recognizing that there will considerable refinements required as new transformative approaches are taken to research, innovation, learning, and collaborations;
- Open, transparent, engaged, and multidisciplinary/multi-sectoral;
- Responsive to the needs of individual use communities – no "cookie cutters";
- Dynamic and nimble; and
- Recognizing of the fact that collaboration is not cost-free: it requires substantial commitments of time, energy, and in-kind resources, even if direct financial costs are not huge.

Key Priorities Going Forward

Develop Government Policy. The international experience plainly shows that a comprehensive federal policy is critical for driving a DI agenda forward. This policy must address the entire DI ecosystem, wherever and however the services are delivered. Industry Canada, in consultation with the TC3+ and the LC, must articulate an overarching policy to frame the strategies and tactics of all players.

Align All Key Stakeholders.   While there is a strong history of cooperation among DI organizations, the existing alignment and coordination among key players, including CANARIE, Compute Canada, and Research Data Canada (RDC) are insufficient to the task. Improved alignment would include integrated and collaborative DI strategic planning, refined funding policies (especially for foundational "platform" and/or generic DI components), and alignment of project and program timelines. The DI should be inclusive of the needs of non-academic users of research data, especially within government and the private sector.

The Leadership Council Should Continue to Coordinate DI. The LC should continue in something close to its current form for at least the next 4-6 months (but understanding that the successor or "new" LC will likely exist for at least two more years) to ensure that the momentum from DI Summit 2014 is not lost. It should be nimble, agile, resourced at a reasonable level, and facilitative rather than operational or "managerial".

Some near-term actions were identified for the LC, including: refining and adding to its membership, developing a Charter, identifying a neutral Chair/Champion, developing communications and accountability mechanisms for reporting back to the community, and forming Working Groups for key DI components and DI pilot projects.

Develop Concrete Pilots and Demonstration Projects. The community expects concrete progress to be made without delay. The LC was tasked with coordinating pilot and demonstration projects, reviewing their progress on a regular basis, and disseminating information on their progress to the community. Pilots and demonstration projects will advance the DI agenda even while the larger policy framework is in development (and in fact will help inform that policy design).

Some key pilots agreed by the participants and already initiated are: developing a catalogue of existing services to identify gaps and avoid duplication, creating a pan-Canadian RDM network, piloting RDM protocols and technology requirements, developing RDM metrics, and investigating a "DI Trust".

Develop Mechanisms and Metrics that Will Foster Positive Change. The new DI will require a significant learning curve for researchers, digital librarians, and the providers of computing and network services. Human and financial resources must be devoted to developing the requisite expertise, transferring this knowledge to DI users, and guiding users through the expanded DI universe.

Metrics that measure Canada's success in achieving an integrated DI should also be developed such that they encourage positive behaviours – for example, providing and sharing significant datasets should be seen as significant contributions to the research landscape, encouraged through institutional policies and practices, and rewarded within academic tenure and promotion systems.

1. Introduction

On January 28-29, 2014 in Ottawa, the Leadership Council for Digital Infrastructure DI (LC-DI) conducted the Digital Infrastructure Summit 2014. This event brought together roughly 140 decision-makers, influencers, resource providers, researchers and other stakeholders with critical interests in Canada's DI. These individuals discussed how to ensure the availability of a comprehensive, integrated, and sustainable advanced DI ecosystem to support research, education, and innovation in Canada. Participants included representatives from universities, researchers, service providers, industry, and funding bodies. The intent was to chart the path forward and address shared roles and responsibilities, and thus was a working meeting in which concrete action items were developed and discussed.

Extensive consultation was done with key stakeholders prior to holding Summit 2014, including a crowdsourcing campaign run on the website of the LC-DI. In addition, participants had access to a great deal of background material prior to Summit, including:

- The consultation paper developed by the Granting Councils and CFI (TC3+);
- An environmental scan of the state of Canadian DI, identifying important current Canadian initiatives and key players;
- An analysis of how a number of international jurisdictions have addressed their own DI ecosystems;
- A draft Policy Framework to aid in conceptualizing DI priorities and in helping assess different options for DI; and
- A detailed Roadmap intended to be a "think piece" on the types of action items and timeframes required to design and implement a coordinated DI ecosystem.
- A reference list of a number of other resource materials

All of the above is available on the Digital Leadership Council website at http://digitalleadership.ca/resources-2/

During the morning of the first day of Summit 2014, short presentations were made by representatives of the LC-DI, the Natural Sciences and Engineering Research Council (NSERC), the Social Sciences and Humanities Research Council (SSHRC), the Canadian Foundation for Innovation (CFI), the Canadian Institutes for Health Research (CIHR), Genome Canada (GC), and the consulting team and facilitator engaged by the LC-DI to support the Summit. The remainder of Day 1 saw two rounds of breakout sessions in which 13 tables were charged with discussing two specific topics:

- Developing a robust Canadian research data management infrastructure; and
- Priorities for immediate actions related to DI policies, strategies, financing, and planning.

On Day 1, each table had assigned seating to ensure a healthy mix of participants with different perspectives, needs, and expectations of DI, and included both a Chair and a rapporteur. Day 1 finished with a plenary recap.

During the half day session on Day 2, the breakout sessions addressed seven key themes that arose on Day 1, with participants free to join whichever group most interested them:

1. New (or leveraged) DI services and tools; middleware development
2. Funding: possible recommendations around funding issues
3. Data management plans (implementation/facilitation issues)
4. Priority pilot projects: what, who, why
5. Metrics and evaluation methods
6. Inculcating a culture of data stewardship
7. Private sector, academy, and government: further the dialogue

More details on DI Summit 2014 and the full background papers may be found at:
http://digitalleadership.ca/2nd-annual-digital-infrastructure-summit/.

2. Characterizing the DI Ecosystem

Overview

The Canadian advanced digital infrastructure (DI) ecosystem is the collectivity of facilities, services and capacities that provide the research community (academic, government and private sector) with the capability to conduct data-intensive, top-level research in their respective fields, and to intensify collaborations nationally and internationally. With data and the infrastructure for data stewardship at the core of the DI ecosystem, it is supported by the hardware, software and people that provide single-site, distributed and nationally centralized computational, networking and analytical tools. This e-infrastructure is itself connected with comparable systems internationally. The term ecosystem is used to emphasize the fact that the component parts are of interrelated, interdependent, and require ongoing interactions for effective functioning of the system (or indeed, functioning at all). See Figure 1.
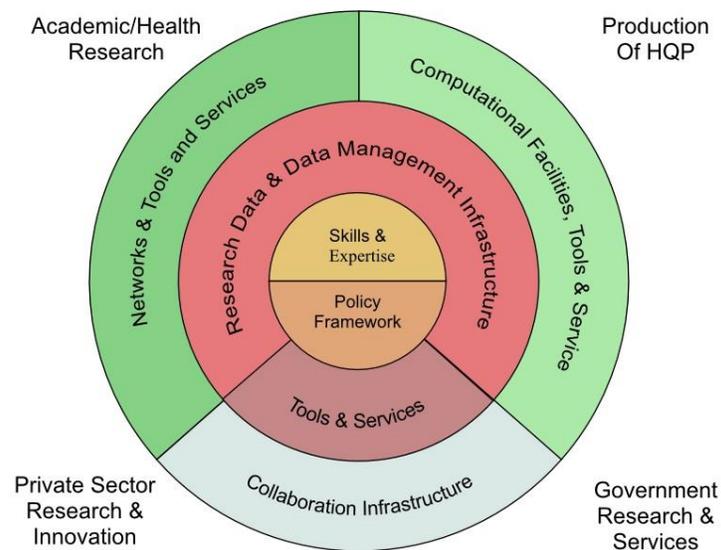

Figure 1. **The elements of the Canadian DI Ecosystem**

What is the problem?

While Figure 1 represents an integrated and connected DI ecosystem, the fact is that Canada has some distance to go to achieve that vision, despite some excellent foundational elements being in place. There are continuing problems as yet:

- Fragmented approaches in the absence of long-term vision
- Inconsistent funding and inadequate planning horizons
- Focus on equipment rather than people; little attention to data as "infrastructure"
- Policy gaps

In the words of one participant:

We have ten dimes but not a buck!

This presents a challenge in how we ensure that foundational DI infrastructure components deliver maximum benefit at any given time. Investments in the DI ecosystem have too frequently been piece-meal or project-based, creating disincentives to long-term system planning. Incentive structures have been competitive and insular and have therefore not fostered pan-Canadian collaborations.

Three overarching "problems" with the Canadian DI ecosystem were identified as the prime targets for discussion and action at DI Summit 2014.

Data management/stewardship

Research data management/stewardship is as yet the weakest link in the Canadian DI landscape, despite the massive increases in the amount of data being created daily through the research process. There is currently no agreed-upon strategy and/or the capacity to protect this valuable public asset, with little capacity to support access, use and reuse by a wide range of users. To date investments in digital infrastructure have been more focused on technology, without concomitant attention to data management and storage, curation and preservation, the provision of skilled knowledge management and data personnel, and relevant software development. The report from the Summit 2011 and the work of Research Data Canada (RDC) have laid a comprehensive foundation for this; the recent discussion paper from the TC3+ is a significant move in addressing some aspects of this problem[1].

Governance/coordination

There is a pressing need for greater collaboration between the funders of research (the TC3+), the providers of all aspects of the national digital infrastructure, and the critical partners in Canada's research enterprise, including universities, professional associations, researchers, academic libraries,

---

[1] TC3+ Consultation - Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf; Tirer profit des données massives: vers un nouveau cadre stratégique pour l'évolution de la recherche numérique au Canada http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_f.pdf

national standards organizations, and other research performing and user organizations. The collective actions of the TC3+ are an important step in this direction, but without increased collaboration and coordination, we risk fragmented approaches, and sub-optimal alignment of activities and investments. While structural change is not necessarily the answer, there have been extensive and urgent pleas for a more structured approach to dialogue, coordination and integration.

Policy and planning framework

Canada lacks a cohesive national strategy that provides an integrated planning and funding framework for all the elements of the DI ecosystem. As an example, two important components of Canada's digital infrastructure, Compute Canada and CANARIE, are funded under different structures and for differing time periods. There are asymmetries in mandate and performance expectations. There is also a lack of clarity as to how infrastructure for research data management should be effectively delivered and aligned. The lack of appropriate planning and investment horizons for all elements of the DI ecosystem engenders a "short-termism" of approach with investment cycles that are out of step with the pace of change in research and technology development, and with many DI elements developed through small-scale independent projects rather than at a national infrastructure level.

3. Key Messages from Summit 2014

3.1 Canada Needs a Comprehensive, Integrated, DI Ecosystem

There was near-unanimity that Canada needs a comprehensive, integrated, sustainable DI ***ecosystem*** – the term is important – supported by a compelling national vision for how DI can foster excellence in research, education, innovation, and policy development. This will be driven by the reconceptualization of ***data as infrastructure***, along with the supporting resources needed for their effective and efficient use across disciplines and sectors. Infrastructure can no longer be seen as simply "machines and pipes". However, the existing alignment and coordination among key players, including CANARIE, Compute Canada, and RDC are not seen to be sufficient to the task as presently structured and operated, notwithstanding the excellent history of cooperation amongst these and the many other Canadian players with stakes in DI.

As a result of these issues, the need to address DI design is seen to be urgent, to the point that immediate action – even if not perfect – is much preferred over an additional round of discussion and consultation. There was a feeling, moreover, that Canada can now take the opportunity to use this paradigm shift to fully leverage digital resources to foster research, innovation, knowledge, culture, education, and learning. DI is not a technology-only issue: it is more fully about concepts, cultures, people, and processes. Canada's approach to managing data as a national asset should be:

- Collaborative, coordinated, clustered, and horizontal – a different way of doing things, not just a better way of doing things, and certainly not another silo;
- Experimental, recognizing that there will considerable tweaking and refinements required along the way, especially as new transformative approaches are taken to research, innovation, learning, and collaborations;

- Open, transparent, engaged, and multidisciplinary/multisectoral;
- Responsive to the needs of individual use communities – no "cookie cutters";
- Dynamic and nimble; and
- Recognizing of the fact that cooperation and collaboration are not cost-free: they require substantial commitments of time, energy, and in-kind resources, even if direct financial costs are not huge.

3.2 Factors Needed to Facilitate DI Success

The international scan showed that:

- 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;
- All other jurisdictions recognize the need for *multiple stakeholders* to be engaged and the concomitant need for co-ordination; and
- From an early focus on physical infrastructure, the international DI discourse has increasingly used *the lens of data as a national asset*.

There was agreement with these three factors being key facilitating elements in advancing the nature, cohesiveness and effectiveness of Canadian DI. Many of the action items advanced below resonate with these insights.

3.3 Research Data Management (RDM) Must Be at the Core of DI

Of all components of the DI ecosystem, research data management (RDM) is at the core – without data, there is no need for either network access or computing power. If data are not preserved, shared, reused and optimized, research productivity is unnecessarily impaired, and valuable assets are underutilized or lost. Issues to consider in RDM are discussed below.

3.3.1 A core DM resource must be created

Research Data Canada (RDC) has demonstrated the ability to bring all of the data stewardship players around the table in a constructive, action-oriented context. It should be recognized as the lead on the data stewardship component of the DI ecosystem and be afforded a more robust, secure, predictable, and sustainable funding mechanism.

Discussion

Some form of National Data Service is required, as is found in many other countries (e.g., the Australian National Data Service).

In Canada, and unlike the provision of computational resources and networks, there are multiple policy and service delivery players for data stewardship; in this context management of the data dimensions of DI is hugely more complex. RDC is the most logical player to form the nucleus of a Canadian national data service to coordinate data repositories (whatever form they may take) and to deal with the multiple policy, educational and implementation issues.

Certainly other important organizations in the area of RDM such as CARL (which has expertise in data harvesting and archiving) and CRKN (which has expertise in data licensing and rights) should be active participants in designing the new initiative, as should university VPRs and Provosts. Compute Canada can play a role in storage (which method(s) are used).

The private sector can also play a role (e.g., for cloud storage).

Regarding funding, the current RDC structure, while robust in terms of goals and operations, suffers from lack of sufficient and predictable funding, which in turn limits its authority and effectiveness. There should be investment in RDC to a similar level as that for CANARIE and Compute Canada.

3.3.2 A high level strategy for RDM is required
Canada must develop a high level strategy for RDM implementation to achieve the integrated goals articulated in the policy framework. Such a plan inevitably entails shared responsibility of federal and provincial governments, institutions and other organizations, but needs to start at the federal level with articulation of a strategy for:
- Data input and access;
- The structure and distribution of responsibilities for delivering the services – computational, networks, data services and storage (include reference to "last mile" implementation of networks, and long term archival data storage);
- Responsibility for international agreements (e.g. OECD);
- Responsibility for international operational linkages on data (e.g. research Data Alliance);
- Responsibilities for developing and communicating standards;
- Distribution of responsibilities for funding of various components, with recognition of the diverse funding sources;
- Responsibilities for monitoring and reporting on the state of DM and achievement of policy objectives.

3.3.3 RDM policy leadership is critical
The TC3+ should develop and implement RDM policy, and tweak and monitor it as they go along.

Discussion
Consider a two-prong implementation strategy based on: (1) Principles/policy; and (2) Long-term funding policies to support RDM. Both would have a long-term view of RDM to generate Canadian socio-economic benefits.

3.3.4 RDM culture must be developed through appropriate policies, incentives, and metrics
The culture of effective data stewardship and data sharing needs to be brought to the fore. There are currently few effective incentives at the level of either the individual researcher or the institution to encourage effective RDM, and in fact a number of disincentives. Nor are the benefits well understood.

Good data stewardship is not just a researchers' responsibility, but is also needed at institutional, organizational, national and disciplinary levels.

Discussion
The TC3+ should provide for the overarching RDM policy direction. In addition, scholarly associations could play a significant role in changing the culture of data stewardship and concurrently what should be measured as important research outputs in tenure/promotion discussions.

Metrics are required to measure RDM progress against objectives during the design phase, and to encourage appropriate behaviour once implemented. Cross-disciplinary and cross-sectoral usage should be encouraged in these metrics, as should consideration of the design of data management plans at the institutional level.

This topic is important enough to warrant a pilot project; e.g., first developing a logic model for the benefits afforded by DI, and assessing how to encourage critical aspects of RDM (e.g., including database deposits and/or citations, or even publishing data management plans in a researcher's CV), so that the tenure and promotion cultures recognize good data stewardship, and the benefits of good stewardship (e.g., cost efficiencies, opportunity for validation of research results) for institutions are recognized.

3.3.5 DI rests on a new, expanded definition of digital "infrastructure"
DI Summit 2014 clearly demonstrated that expansion of the definition of "infrastructure" to include the non-physical aspects is far from completely accepted or understood within the community. Extensive effort is needed by all stakeholders to inculcate this understanding – especially that **the data themselves** are the most critical part of the DI infrastructure, as well as related aspects such as human and financial resources for RDM.

3.3.6 RDM training is an essential component
Effective RDM will require not only research into best practices and training researchers (including faculty, graduate students, and eventually undergraduates and Research Office staff) in the value and benefits of good data stewardship, but also training a new generation of research librarians and other individuals who can guide users in using the system both on the digital literacy side (e.g., storage and preservation, access, metadata, etc.), and on the analytic side (e.g., appropriate algorithms, interpretation).

Discussion
As RDM expands in scope and improves in effectiveness, the current shortage of training opportunities, resource materials, and financial and (especially) human resources will become more glaring.

"Training the trainers" is also required, and should be addressed by RDC, Compute Canada, and CARL, with support from TC3+, scholarly associations, librarians/libraries, CANARIE, and Compute Canada.

A networked model for training and expertise such as that used by the UK Digital Curation Centre would serve Canada well; regional expertise networks should be encouraged.

3.3.7 Life-cycle issues need to be addressed
The life-cycle sustainability of datasets will eventually become a pressing concern – what data should be included? What should be preserved? For how long?

Software and middleware platforms for lifecycle data management need to be developed.

Discussion
There are essentially two types of Big Data.
- Data that come from large collaborative enterprises such as TRIUMF, CLS, ONC, SNO, etc. Generally these organizations have data management expertise (although data Interoperability and accessibility to others certainly remains an issue).
- The aggregate of all the data from smaller research players which is equally valuable but ill-served at present in terms of its infrastructure, expertise resources and adoption of good practices. It is also far-more challenging in its diversity.

3.3.8 Data storage is a crucial component
DI will require significant additional data storage capacity.

Discussion
There is a need for a Canadian academic cloud storage facility for massive data storage, based on current initiatives and understandings of data volumes.

Compute Canada should provide at least some significant amounts of storage, but not data curation. It was noted that a number of institutions (e.g. OCUL) are investing in storage capacity – focused primarily on the research outputs of their own faculty. Issues of national access and interoperability remain.

3.3.9 Methods to foster compliance need to be developed
Some forms of project and/or institutional funding could be tied to compliance with RDM protocols and best practices.

Discussion
To foster compliance, institutions could require CFI funding applications to demonstrate non-overlap with existing DI infrastructure.

While in time it can be expected that the TC3+ will require compliance with their policies on research data management, draconian policies can become "paper tigers". It will be important to demonstrate that RDM policies directly benefit the user communities. Also, in the absence of DM infrastructure and professional DM expertise in place at the time the TC3+ policies are implemented, a policy will be honoured in the breech.

3.4 A Culture of Connection and Sharing Must be Developed

Discussions highlighted the importance of evolving the research culture with incentives for sharing data, physical resources and best practices. There was also recognition that this is not a narrow discipline or academic issue, but that forward strategies must incorporate the breadth of data-intensive research and research user communities.

3.4.1 A culture of sharing physical resources

Promote and incent the use of shared physical resources.

Discussion

Problem - Researchers have a culture that defaults to implementing new local resources rather than re-purposing existing shared commodity resources.

Approach – provide incentives for individual researchers to share physical resources; also there needs to be leadership at the institutional level.

3.4.2 Shared data are to be encouraged

The value of data is enhanced by its preservation and accessibility for reuse and validation of research findings.

Discussion

Problem - There are no incentives for researchers to preserve/sustain (or even share) the data that is produced from a research project.

Approach – Articulate data policies at the level of the TC3+ and institutions. Research offices and library practices will reinforce higher level policies. Also scholarly and scientific associations can play a significant role in educating their communities.

3.4.3 Shared best practices are to be developed

Share best practices. Some Canadian initiatives have developed "best practices" and resources. For example, some best practices may be available from organizations such as CBrain that have developed facilities that are "catholic" in terms of the nature of data input and analytics used.

3.4.4 Support for researchers will maximize effective and efficient usage

Provide support for researchers. This will assist in changing the culture. There is currently a paucity of tools and support expertise to assist researchers who are prepared to engage in data stewardship.

Discussion

Some of the systems that would be effective for supporting researchers are:

- Simple on-line tools for DM planning;

- Information on IP/use/licenses;
- Inventory of repositories (Databib and RDC);
- Inventory of educational opportunities (already prepared by RDC);
- A DCC-like centre of expertise and services;
- Policy notes on data citation for use by T&P committees;
- Wide dissemination of information on what exists and where to get it (see also recommendations for inventories and catalogues).

On access to cycles and networks, as well as analytical tools – there is a need for production ready tools (experimental, specialized); currently they are not available or far too expensive.

3.4.5 Design a concierge function for "one stop shopping"
Implement concierge-type functions. Many users need guidance in finding and using appropriate infrastructure and data management approaches. Implementation of a concierge function for DI would help guide users to the most appropriate datasets, HPC resources, analytic expertise, and network support. This function will increase in importance as more DI elements are developed. The Compute Canada model of providing local expertise has been effective.

3.4.6 Pilots will be a catalyst for culture change
Catalyze change through pilots. It was agreed that there is a need to develop pilot research projects to assess and advance best practices in DI design (some of these are described in more detail in Section 4.3) and improve culture and connection. These can be supported by TC3+.

3.4.7 Develop an inclusive agenda
Approach data stewardship, big data and DI as an inclusive agenda involving all disciplines and sectors.

Discussion
Government Big Data users should be immediately brought into the DI design phase. The involvement of the private sector, both as service providers, and ultimately as innovation developers (through use of the data), should also be addressed. For example, the financial and social networking communities are also Big Data users.

3.5 Address Financial Dimensions
Agencies and funding models must evolve to match the new reality of an interdependent research ecosystem. There is a need for a common approach to funding, to simplify and develop more straightforward approaches to funding projects and programs involving both operating and infrastructure support, and to align Compute Canada with CANARIE's funding models (i.e., strategic, longer-term funding), and improve the structure of the financing (e.g. what kind of co-funding is required).

Most participants agreed that the current funding models for some of the major DI elements create significant problems, and especially that of Compute Canada and that there are few vehicles for support

for the system-wide elements of RDM. Further, some small incremental funding was needed for certain aspects of the design and implementation of DI. Although the latter amounts are not likely to be significant, they are not currently "line items" in the budget of any of the key organizations. This gap needs to be addressed.

3.5.1 Revisit approaches to funding generic or foundational DI
The criteria for funding the foundational DI elements must be revised to better support and maintain platforms, other core components, functions, tools, and people.

Discussion
Project-based funding, using current TC3+ criteria, are often not appropriate for assessing the core DI elements, which: e.g., (1) are generally catholic with respect to both discipline and sector (and thus usually not within the purview of any individual granting council), (2) require longer-term, sustainable support for planning, operations, and maintenance (national planning needs incentives and follow-through; operations difficult to find through co-funding mechanisms), but flexible enough to respond to emerging needs; (3) require support for attraction and long-term retention of top HQP for essential support functions; (4) need "evergreening" of software and operational elements. (This is not to say that project-based funding has become unimportant, only that it does not serve DI foundational elements well.)

Even within discipline-specific tools (e.g., specialized software and middleware), the development of tools interoperable across the entire discipline is not well supported by current funding mechanisms.

The potential mis-match between research and infrastructure applications has long been a sore point, and a more comprehensive and integrated DI will make this even more pointed. A critical piece is to first define who is responsible for these funding mechanisms.

3.5.2 Revisit co-funding requirements
Revisit co-funding requirements for core DI platform elements.

Discussion
Unlike local and regional infrastructure, national DI platforms do not have the same ability to source matching and/or even co-funding. The current matching funding formulas (e.g., for CFI, GC) is problematic for capital support, and dysfunctional for operational support of national platforms.

3.5.3 Lengthen planning horizons
Longer-term, sustainable, predictable funding is especially needed for Compute Canada and RDC

Discussion
Address the need for an improved planning horizon for DI by extending the funding horizon to 6 years (perhaps a rolling 6 year commitment) with a mid-term review at 3 years. Avoid destabilization of core national DI through uncertainty of policy commitment.

Compute Canada additionally suffers from a co-funding requirement that is not in Canada's best interests for developing a DI that serves Canada's long-term needs; other options should be investigated.

3.5.4 Address DI gaps to ensure maximum utility
Address funding gaps for some important DI elements such as middleware, algorithms, metadata, and analytics.

Discussion
Although existing services fill some of the gaps at present, an expanded DI will require a more comprehensive set of tools. Improving these services will help assure acceptance by the end-user communities (initially mainly in academia, but government and private sector users will shortly follow suit).

3.5.5 *Align funding cycles*
Align and harmonize the funding cycles of CANARIE and Compute Canada.

Discussion
This would facilitate joint planning and delivery of middleware and software, new algorithms and analytics, and implementation initiatives. Ideally this would include harmonization of related federal/provincial funding agreements.

With RDC mandated and resourced as the convener of the RDM components of the DI ecosystem, its funding cycle should match as well. This will also enable CANARIE, Compute Canada, and RDC to achieve greater alignment in strategy, funding, and operations.

3.5.6 Address indirect costs on institutions
Institutions made it clear that achieving an integrated DI, including RDM requirements, will create additional indirect costs for them. A specific solution was not offered, but this issue must be addressed within DI funding mechanisms.

3.5.7 Industry Canada is to be the lead Actor
Industry Canada is the logical actor to take the lead in revising funding mechanisms for foundational and/or generic DI components. Revisions should be made on a principle- and policy-driven basis.

DI will be a significant driver not only for research, but also for innovation and education. These aspects suffer from "market failure" when attempting to support some DI elements solely within the TC3+ academically-driven framework.

4. Going Forward

4.1 The Leadership Council

With very few exceptions, there was agreement that some form of coordinating body is required, at least in the short-to-medium term, and that this body should be nimble, agile, resourced at a reasonable level, and facilitative rather than operational or "managerial". Terms such as "moral authority", "promote", "coordinate actions by others", "track progress", "advance the solution", "champion", "communication", and "cross-cutting" are indicative of the role that such a body would play – in fact very much as the Leadership Council role has been to date, and continuing to avoid creating another layer of bureaucracy. The term "Coalition" was sometimes employed to designate the somewhat expanded role and membership of the LC in going forward; while this term did not meet universal acceptance, the essential idea of some form of coalition did.

In terms of participation in the Leadership Council itself, there are views that it is "too large" on the one hand, versus on the other that it should be expanded to include other stakeholders (e.g. AUCC, industry and government research units). These conflicting opinions might be reconciled by the notion of a broader stakeholder group supported by an active and accountable executive committee that meets frequently and is itself supported by a light secretariat. It is clear, however, that the community wishes for continuation of the current structure until such time as a modified Leadership Council is put in place.

There is less consensus on the appropriate shape of a coordinating body in the longer term, likely reflecting the fact that much will depend on the articulation of an overarching DI policy from the federal government and the future status of Compute Canada, CANARIE and RDC. There is also some divergence in approach on the locus of leadership. In one case there was a strong statement that it not come from the technology layers; in other comments a stronger role for Compute Canada and CANARIE is seen.

4.1.1 Continue the LC in the immediate term
 The LC should continue in something close to its current form for at least the next 4-6 months (but understanding that the successor or "new" LC will likely exist for at least two more years) to ensure that the momentum from DI Summit 2014 is not lost.

Discussion
A critical early action is to establish the LC's role, legitimacy, and recognition from other key stakeholders. Given the often contradictory opinions of "who should lead", unanimity cannot be

expected. However, action must be taken regardless, as Industry Canada at the very least – and virtually the entire community – firmly expects forward momentum.

4.1.2 The LC's role will be to be a focus for coordination

The LC will act as an overall focus for the coordinating efforts of all relevant organizations, but with small, nimble, short-term working groups of variable membership created as need be. Notional working groups are discussed in the Roadmap. The LC should communicate outcomes of the working groups and pilots on a regular basis.

The draft Policy Framework and Roadmap should be used to help guide future LC/Coalition actions. These documents, and the two environmental scans, were well-received by Summit participants and should continue to be available for the community as resources.

4.1.3 Refine the LC membership

Refine the LC/Coalition membership.

Discussion

Immediate additions in membership should include:

- Strengthening of academic presence (perhaps a Provost or A-Provost), ideally with individuals at the Presidential level helping move DI forward (these individuals have not, to date, been active DI supporters);
- AUCC;
- The health and health research sector;
- Private sector representatives (perhaps one industry and one an association such as i-CANADA);
- The provinces and territories;
- Government departments using Big Data.

4.1.4 Create an Executive Committee

Create an Executive Committee that meets regularly (once a month). The LC should receive Minutes of those meetings and meet at least 3 times a year or at the call of the Chair. More frequent meetings are required over the next six months as early action items are implemented and findings/recommendations are reported back to the community.

4.1.5 Identify and support a neutral Chair

A transition in Chairs should be envisioned at the four month point, with the current Chairs being ex-officio on the LC for necessary continuity, in addition to possible additional CIO and VPR appointees.

For avoidance of conflict of interest, and to act as a champion for DI, a single neutral (non-aligned) chair should be identified. This individual would have DI as his/her primary or dominant responsibility, not deal with DI "off the edge of the desk".

4.1.6 Form policy Working Groups not covered by pilots
Two key policy Working Groups (WG) should be formed ASAP – areas where there are no planned or existing pilots.

1. WG mandated to refine the policy framework and articulate key strategies/actions for delivery of the goals.
2. WG mandated to take a macro perspective and review the overall structure of and funding mechanisms for national DI in Canada. Among other things, it would examine the relationships, appropriateness of funding duration, cost sharing and optimal approaches to integration and connectedness among computational, network and data components of the DI ecosystem, proposing and evaluating alternatives.

4.1.7 Ensure integration with existing initiatives
The LC should ensure it maintains an active watching brief on existing DI elements to seek synergies among them and to avoid any overlaps. Some consultation and planning is now being done independently by e.g., CANARIE and Compute Canada, but is focused on specific mandates. Others (e.g. the CARL initiative, the work by CRKN, the work of the various RDC subcommittees) are expected to respond directly to recommendations emanating from the Summit.

4.1.8 Develop accountability mechanisms
Demonstrate accountability to the community. Overall, this should be principles-based (which can be built around principles found in the Policy Framework).

4.1.9 Develop an LC Charter
Within the next four months, prepare a new Charter for the work of the LC over the next two years. The Charter would address features found in the Roadmap (e.g., governance, accountability, communications, etc.) Membership would then be adjusted (form to follow function) accordingly.

4.1.10 Obtain resources for an LC Secretariat
Immediately seek resources for a light LC Secretariat, initially of 2 years duration. This would ideally be a combination of in-kind support (as CUCCIO has provided over the past 18 months for administrative tasks) and contractual support for managing the web, supporting the meetings of the LC and its Executive Committee, and planning for the next Summit. Existing LC members may "pass the hat" to support this function, but other possibilities include something akin to a membership fee for inclusion.

4.1.11 Develop and support pilots
The LC should immediately support the development of pilot projects aligned with the Policy Framework and as suggested above and in the Roadmap (e.g., establishing Terms of Reference, membership, renewal, progress reporting, etc.). Resources (as needed) for the pilot projects should be sought, once their requirements are identified. The TC3+ would appear to be an appropriate source of support for pilots. Reports on progress would be made at Summit 2015.

4.1.12 Plan for Summit 2015

Begin planning for Summit 2015 immediately. That Summit should involve discussion of the outcomes of the pilots and thematic discussions as held on Day 2 of Summit 2014.

4.2 Develop Federal DI Policy and Strategy

A key finding of both the international environmental scan and Summit 2014 is that policy can help drive and foster DI design, implementation, and operation. The discussion below repeats some points made above, but puts the policy elements together for ease of reference.

4.2.1 Develop an overarching DI federal policy

There is a need for a national vision and policy on an integrated DI that encompasses the full spectrum of DI services, wherever and however they are delivered. This will also reinforce the TC3+ actions. This policy would frame strategy and tactics, guiding the actions of all players. Responsibility – Industry Canada; in consultation with the LC.

4.2.2 Identify and engage champions

In addition to the TC3+ champion, DI needs a federal ministerial level champion. Ideally there would also be champions at provincial levels.

University Presidents/Principals need to lead the articulation of the importance of a strong DI for research, education, and innovation. They should issue a declaration of Intent – one that is value based, with key principals, which may be developed in association with development of the value propositions.

4.2.3 Develop value propositions

There needs to be a compelling articulation of the DI value proposition for academia, industry, educators, healthcare, and the public.

A tight policy paper is required to make the case for the necessity of federal and provincial funding of a shared, national digital infrastructure as an accessible platform facilitating economic advancement and public good. This would be framed much like the historic definition of public works – bridges, ports, roads, etc. – and with similar market failure arguments likely being central, but with explicit recognition of data, professional expertise and software as integral parts of the modern DI.

Responsibility – small Working Group to draft for LC approval. This could also be addressed by the DI Trust pilot (see section 4.3).

4.3 Develop and Support Pilots/Demonstration Projects

There was broad consensus on the value of developing and implementing some pilots in a rapid time frame to help understand what works and to advance the agenda, even in the absence of a larger policy framework. These would also serve as useful vehicles for leveraging and coordinating the expertise

within institutions and organizations. The LC-DI should receive reports on the progress of these projects on a regular basis and disseminate information on their progress.

The following are the major projects that were proposed or agreed as key stepping stones by the participants at DI Summit 2014 and a means of demonstrating the commitment of the stakeholders to taking both individual and collective responsibility. Given their importance, these are outlined in some detail, capturing developments following the Summit.

4.3.1 Provide a catalogue of existing services
Create a catalogue of services and infrastructure. Many researchers and stakeholders are unaware of the services and capabilities that exist. This initiative is a "low-hanging fruit" and involves developing a catalogue of existing Canadian infrastructure (including services, software, HR and HQP, etc.) relevant to DI. The catalogue would then serve as a basis for identifying gaps in services and/or communities currently not served well. One of the RDC Subcommittees has already developed an inventory of some of the data management infrastructure existing in Canada that could be incorporated into this larger initiative.

Responsibility – CANARIE, Compute Canada and RDC to compare notes on existing inventories and coordinate development of an integrated catalogue under the leadership of the RDC Infrastructure Committee.

4.3.2 Create a developmental and demonstration network
CARL and CRKN had, prior to the Summit, already begun working towards the development of a _pan-Canadian Research Data Management Network_ that is focused on:
- Education and expertise, e.g.:
  - Distributed expertise;
  - DCC like;
  - Diverse target audiences.
- Services, e.g.:
  - Data management plans and planning;
  - Access to data; referral to expert advice;
  - Services for data visualization and modelling;
  - Data rescue.
- Tools and technology, e.g.:
  - Focus on librarian's expertise in management of content;
  - dissemination, discovery, preservation, and access;
  - Offer preservation platform; disseminate preservation standards;
  - Requirements for data integrity.

Project coordinator – Martha Whitehead, Queen's.

CARL has taken the lead on developing this initiative and is the driving force; CRKN is participating to bring in the libraries outside the narrower CARL membership. Initial meetings have had very good buy-

in. Participants at the Summit noted the alignment with the needs of the community in exercising data stewardship and that it was an important element of compliance with the forthcoming TC3+ policy. In addition to development of a supportive and complementary pilot (see 4.3.3), CUCCIO expressed its support for the initiative. CANARIE could also be a player in underpinning this initiative.

The project will require funding. It has seed funding from CARL; it would be valuable to have some TC3+ support.

4.3.3 Create an RDM Pilot

This pilot concept, developed at DI Summit 2014, is designed to identify the specific RDM protocols and technology requirements, using three pilot domains (Astronomy, Social Sciences and Medical Genomics), and building on the foundation of existing expertise and the RDC network. One key element is to explore what elements are transferrable – from one discipline to another; another is to examine data federation capability. Ultimately the objective is to migrate experiences and insights to other disciplines. RDC will lead the project, with CARL, CRKN, Compute Canada, CANARIE, CRDCN and Genome Quebec participation. David Schade (CADAC), Dominic Lam (IBM), Mark Leggott (UPEI), Byron Spencer (CRDCN), will contribute additional middleware and expertise. Pilot coordinator – Walter Stewart, RDC.

Set #1 Goals for RDM pilot:
• Supporting data management plans and solutions to support policies and that have the capacity to be scaled.
  o Timing – need services in place within 2 years.
  o Staged approach – what can be done in 6 months, 1 year, 2 years.
• Develop Common metadata record, map a standard, mint DOIs, micro-services with a common framework to support, Research Data alliance link to develop in direction world is moving in.
• Need predictable data flow and predictable metadata flows. There are challenges across disciplines.
• Data intensive research.
• Longer term archiving.
This will require:
• A requirements document.
• Timing, scope projections.
• Work packages (to come out of the discussion).

Set #2 Goals for RDM pilot:
• The overriding objective of the initiative(s) should be directed at shared data and content.
• The initiative(s) should be able to be accomplished over the next 3-4 months.
• The initiative(s) should build on what already exists and promote the re-use, adoption or adaptation of existing services and tools. Development of new tools or services should be a last resort.
• The initiative(s) should promote a service-oriented or user-oriented approach, hence the need to integrate researchers in the initiative(s).

The initiative should be open to different architectures to ensure that all fields of inquiry can be served.

Initial project scoping
1. RDC (and CRKN, CARL, ...) to lead the project by providing data and metadata management standards and policy compliance guidance.
2. Full participation from Compute Canada (end user researcher interface and system resource & expertise) and CANARIE (network resource, middleware - asset from the NEP projects and pilot funding). End user researcher participation in identifying their data requirement to the developer is critical.
3. Pam Bjornson (NRC) to help find funding.
4. Need a full time project executive to coordinate and enforce execution.
5. Jim Ghadbane (CANARIE) suggested that the project be started based on research data processing workflow - first task of the project team is to define this workflow. Based on the workflow, we can identify existing elements that can be reused and gaps needed to be developed. We can also set development priorities in order to achieve short term objectives, e.g. end user demonstration that shows benefits in research and discovery in one year.
6. Develop multiple data depositories (initially target 3 specific research domains: Astronomy, Social Sciences & Medical Genomic) to identify both common and specific data management and technology requirement.
7. Develop data federation capability to allow data access for researchers who need data from outside of their own discipline.
8. David Schade (CANFAR) to provide middleware development knowledge and help determine middleware reuse.
9. Dugan O'Neil and Mike Bauer (Compute Canada) to engage researchers who can take advantage of the proposed DI pilot.
10. Genome Quebec (Guillaume Bourque, McGill) to provide input for developing data intensive processing environment.
11. Byron Spencer (McMaster and CRDCN) and Mark Leggott (UPEI) to provide common (and the 20% specific) data requirement from the social science research community.
12. Martha Whitehead (Queen's) and Walter Stewart (RDC) to provide their interpretation of the upcoming TC3+ policy for advancing digital scholarship to assist data management development in compliance with it.

4.3.4 Coordinate with the CRKN Integrated Digital Scholarship Ecosystem (ISDE) project

Another initiative that was launched during the development of DI Summit 2014 is the Integrated Digital Scholarship Ecosystem (IDSE) – enabling the next generation of academic library collaboration to further seamless access, diversity of content, and linked infrastructures. This effort aims to advance academic research in Canada by understanding the complexity of the digital landscape and by seeking opportunities to align key stakeholders and providers within the Canadian academic library community around a series of shared objectives. The ecosystem is seen as combining capabilities and infrastructure

of the academic library community beyond content to seamlessly harness the work of diverse organizations that contribute to digital scholarship.

With its focus on mapping the various stakeholders and associated activities involved in digital scholarship within the University library community and identifying the resulting gaps, challenges, and opportunities, CRKN seeks to advance initiatives which strengthen the landscape and present shared solutions for common problems. IDSE is defined as:

- a map of the existing state of digital scholarship in Canada;
- a lens through which to foster collaboration and coordination;
- a platform to implement and sustain key services, programs, or projects;
- a results-oriented initiative that serves faculty, students, and staff at research institutions as well as supporting inquiry-based research by all Canadians; and
- an ongoing and evolving process that must be flexible and agile in order to respond to the changing nature of the digital environment and scholarly directions.

Conduct of the project will be linked with the work of the Leadership Council and of the other pilot projects underway. Mike Ridley is the project manager. A survey instrument has recently been finalized for use in the mapping exercise.

4.3.5 Develop and Support an RDM Metrics Pilot

Another pilot developed and agreed at DI Summit 2014 is on DM metrics. Under auspices of RDC it was agreed to create a multi- disciplinary multi-sector Working Group including Geoffrey Rockwell (lead), Catherine Middleton (co-lead), David Baker (CASRAI and RDC) and Eleanor Fast (Federation). The involvement of the GRAND NCE will be sought. Others are to be added as appropriate.

Goals
- Develop some templates for "measuring" DM performance at various levels;
- Inform that development process up-front with an international scan;
- Benefit from an arms-length body.

Process
- The initial group would first develop *a charter* for 2-year project that brings 'multi-reps' to identify and report back to the larger community on how data management can be assessed;
- Wiki and/or sandpit type of approach;
- Learn from Genome Canada's experiences with DMPs;
- Benefit of a framework or logic model ;
- Talk further with Chad Gaffield to fully understand what is needed and most useful.

Project leads – Geoffrey Rockwell and Catherine Middleton.
Participants – David Baker CASRAI and RDC, Eleanor Fast, HSSF, others to be approached.

Two levels of approach
1. *6 month horizon* - immediate issues – how to assess DMPs:

- At the front end (in peer review) – the extent to which the DMP is appropriate for the proposed project and the type of data generated.
- Following completion of the project – how would one assess successful implementation of DMPs and effective stewardship of data?
  Chad Gaffield would like this within 6 months; he would take any outcomes back to the TC3+.

2. *Two year horizon* – a more in depth analysis of the DI dimensions of research performance – e.g. responding to the following questions:
   - How effectively is the Canadian research community using DI?
   - What are we getting from it that could not be done otherwise?
   - What are effective means of measuring the outcomes and impacts of digitally enabled research (e.g. for peer review of grants and T&P)?
   - How do we stack up internationally?

4.3.6 Investigate a "DI Trust"

Emerging from discussions at DI Summit 2014 on private sector–academic–government common interests in DI, was a proposal for a "DI Trust". This pilot will engage a small group of senior leaders to clarify possibilities and take action on fostering collaboration and concerted action among sectors.

Underpinning the background discussion for the creation of this group was an agreement that it would be an enormous advantage for Canada to have a cohesive strategy that treats all digital resources as a national asset; a strategy that takes advantage of all current resources and spending; a collaboration focused not only on the physical infrastructure, but one that also coordinates our national investment in knowledge generation and human capital. The first steps entail addressing:
- how little we knew about what each of government, academia and business are doing right now, today, and how seldom we consider the competitive advantage that could be created through more knowledge – and more collaboration; and
- making more effective use of what is already in place through collaboration rather than asking governments for more money.

Don Aldridge and Don Hathaway have agreed to take leading roles in launching the 'next steps'.