CHAPTER TWO

THE FIFTH DIMENSION
Chris Greer[1]

INTRODUCTION

The aim of this chapter is to consider a five-dimensional world made possible by cyberinfrastructure and how this notion influences legal frameworks. In discussing this five-dimensional world I will highlight fundamental challenges that hinder this vision, which is a shared vision, not unique to the National Science Foundation, but common to countries throughout the world. I will also consider strategies that could assist in achieving this fifth dimension.

Definitions

Throughout this chapter I use the term 'cyberinfrastructure'. Fran Berman defines cyberinfrastructure as the 'coordinated aggregate of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering.'[2] This definition is particularly appropriate in the context of the fifth-dimension because the definition encompasses not just hardware, software and network fabric but organisations, people and

---

[1] Senior Advisor for Digital Data in the Office of Cyberinfrastructure, National Science Foundation (NSF).

[2] Fran Berman, 'Workshop Concept' (SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences) <http://vis.sdsc.edu/sbe/SBE-CISE_Workshop_Intro.pdf>.

their expertise which makes all of this possible, and may be considered the most integral part of cyberinfrastructure.

I also use the term 'data'. Data refers to items that can be digitised, stored in digital form and accessed electronically. This includes numeric information or text, as well as images, audio, algorithms, software, simulations to name a few.

## HOW IS CYBERINFRASTRUCTURE CHANGING OUR LIVES?

The National Science Foundation believes that:

> The conduct of science and engineering is changing and evolving. This is due, in large part, to the expansion of networked cyberinfrastructure . . .[3]

The fundamental question that should be asked is in what ways are science and engineering changing, and what are the driving forces for those changes?

Prior to the digital age people operated in a world constrained by four dimensions, particularly the three dimensions of place and one dimension of time.

Figure One

---
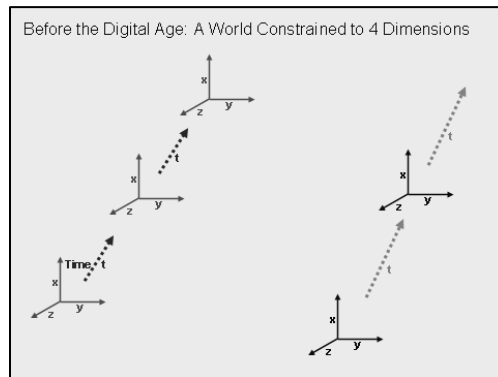
[3] National Science Foundation, *National Science Foundation Strategic Plan 2006–2011*.

Figure One has two trajectories, one of which may have been your trajectory as you prepared for a meeting, the other trajectory may have been the trajectory of the person you were meeting. For both of you to meet you had to agree to interrupt your trajectories, stipulate a time to meet, and then return to your separate trajectories at the end of the meeting (see Figure 2). That is the world that people are accustomed to operating in.

Figure Two
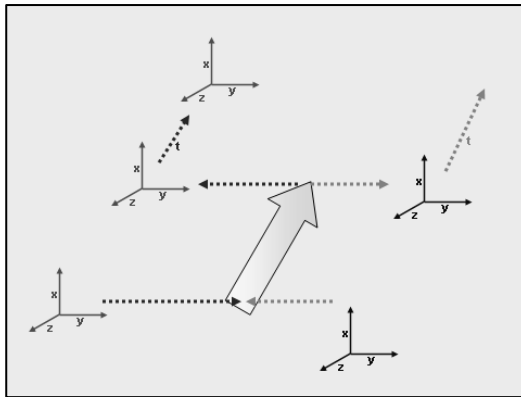
Cyberinfrastructure creates a fifth dimension that is present alongside the existing four dimensions. This fifth dimension provides people with the opportunity to search for information they did not know existed, in places they will never visit, while interacting with unknown people in other places, using instruments they do not own and do not know how to operate in a deep, technical level, but which they have access to because of cyberinfrastructure.

This fifth dimension also allows people to meet in a synchronised mode at an agreed time, or in a meta-synchronised mode where people who are operating at a distance, in their own time zones and context can interact with one another (see Figure Three).
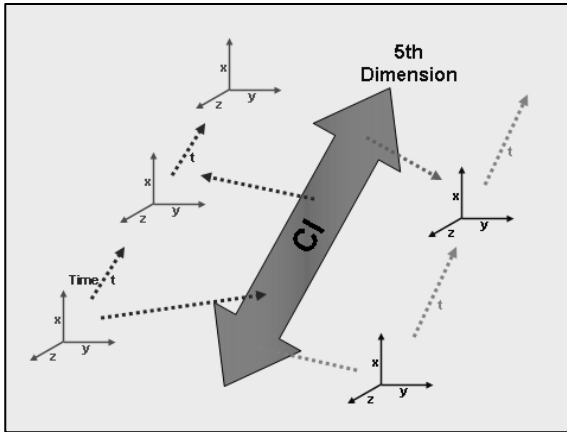
Figure Three

Figure Four is a two-by-two matrix. In a four dimensional world people spend the majority of their time operating in the same-place same-time sector. However other sectors, particularly the different-time different-place sector, have become available for activities involving people-information-facilities and this provides for opportunities that do not exist in the same-place same-time mode. This expansion into other areas of the matrix that is occurring is what is meant by 'operating in a world of five dimensions'.

Opening a fifth dimension through cyberinfrastructure is the defining feature of the digital age. Most people have read Thomas Friedman's *The World is Flat*[4] and are familiar with the idea of dialling up a helpdesk and speaking with someone in India or Malaysia. Software development activities involve teams scattered around the globe so that the development cycle moves with daylight around the globe to become a 24-hour, seven-day a week activity. These concepts are what Thomas Friedman was referring to when he described the world as flat. However this is only part of the picture, because this flat world is also expanding.

---

[4] Thomas Friedman, *The World is Flat: A Brief History of the Twenty-First Century* (2005).

Figure Four
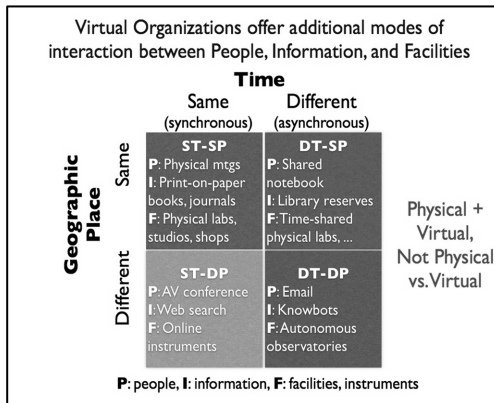
Source: Dr. Daniel E. Atkins

Prior to the advent of network cyberinfrastructure there was no economic space for companies such as Google and Amazon, nor was there a place for the National Virtual Observatory[5] which is vital to astronomy. These new spaces and opportunities, whether economic, scientific or educational, arise because of cyberinfrastructure: the world is getting bigger.

Unlike the four dimensional world in which the driving forces for progress where physical and economic assets, the five dimensional world's primary drivers for progress are information assets and the critical driver for progress in this world is the ability to use information in integrative and innovative ways.

FUNDAMENTAL CHALLENGES OF OUR TIME

Some of the major challenges of our time are enabled by cyber-infrastructure for information integration, such as the ability to ask and answer questions like how and where did life arise on earth? This answer can only be answered using a combination of scientific areas

---

[5] United States National Virtual Observatory <http://www.us-vo.org/>.

such as systematic biology, palaeobiology, biochemistry, metabolic biochemistry, genomics and geochemistry. Likewise the question, what is the biological basis of consciousness? This answer will require integrating information from biology and other sciences.

Some of the large questions of our time will require using information from a wide variety of sources and frameworks together in an integrative way. It is this ability, the ability to integrate information, which is critical to being successful in a five-dimensional world. Individuals, groups and nations that fail to fully embrace this five dimensional world will fall behind.

CHARACTERISTICS OF A FIVE-DIMENSIONAL WORLD

Some of the primary characteristics of a five-dimensional world include: barriers of time and place, which are characteristic of a four-dimensional world, are reduced, information is a primary driver for progress, access to information is available to specialists and non-specialists alike and that the realm of the possible is expanded through new capabilities, resources and mechanisms.

At the National Science Foundation, as in other science and education, engineering and research organisations, the increase of digital outputs is on the rise.

An example of this is from astronomy, where it has becoming increasingly apparent that the dynamics of the universe are an important element for study. The Large Synoptic Survey Telescope is a project that is expected to be realised sometime in the 2010s. [6] The telescope will map the sky every night using a three billion pixel camera, taking a full survey of the sky in under a week. [7] On a clear night this effort will generate 30 terabytes of data. [8]

---

[6] 'Steering the Future of Computing' (23 March 2006)440 *Nature* 383 <http://www.nature.com/nature/journal/v440/n7083/pdf/440383a.pdf>.

[7] 'Steering the Future of Computing' (23 March 2006)440 *Nature* 383 <http://www.nature.com/nature/journal/v440/n7083/pdf/440383a.pdf>; Large Synoptic Survey Telescope <http://www.lsst.org/lsst_home.shtml>.

[8] 'Steering the Future of Computing' (23 March 2006)440 *Nature* 383 <http://www.nature.com/nature/journal/v440/n7083/pdf/440383a.pdf>.

In biology the National Ecological Observatory Network,[9] and in the geosciences and the climate sciences the Global Earth Observation System of Systems[10] are examples of efforts generating datasets analogous to this sort of magnitude. This is also true for society as a whole.

Figure Five[11] is a projection from an International Data Corporation study, which extrapolates the previous work of Michael Laskey of Rutgers University and Hal Varian and Peter Lyman at the University of California at Berkley. The projection shows in exabytes by a year the amount of digital information that is generated globally.

Figure Five
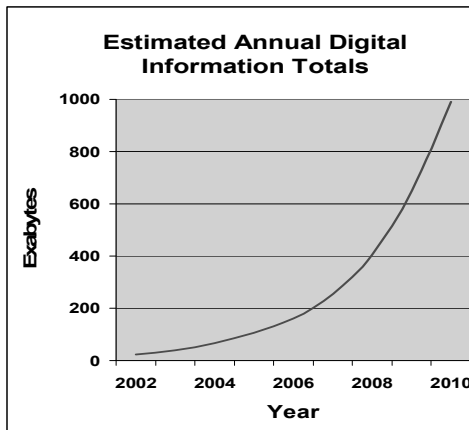
In 2006 a total of 161 exabytes were generated around the world, this is more information then all the documents in the previous 40 000 years human history contained. Printed out in volumes of ones and zeros this would equal a stack of 12 volumes reaching from the surface of the earth to the surface of the sun.

---

[9] National Ecological Observatory Network <http://www.neoninc.org/>.

[10] Global Earth Observation System of Systems <http://www.epa.gov/geoss/>.

[11] International Data Corporation, 'The Expanding Digital Universe' (IDC White Paper sponsored by EMC Corporation, March 2007) <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>; Peter Lyman and Hal R Varian, *How Much Information* (2003) <http://www.sims.berkeley.edu/how-much-info-2003>.

The curve in Figure Five is exponential, which indicates that linear solutions, such as expert curation models, will not be adequate in addressing this growth problem. The curve also illustrates that given the volume of information the vast majority will never be seen by human eyes. The information will be passed, sorted, filtered, analysed and reduced to a level humans can understand. Volume is an important challenge in the world of five dimensions and predicts the need for exponential solutions.

Figure Six[12] is a summary of information technologies over the course of human history from stone, clay and papyrus to paper and now digital forms of storage.

Figure Six

"ONCE IN A HUNDRED GENERATIONS"
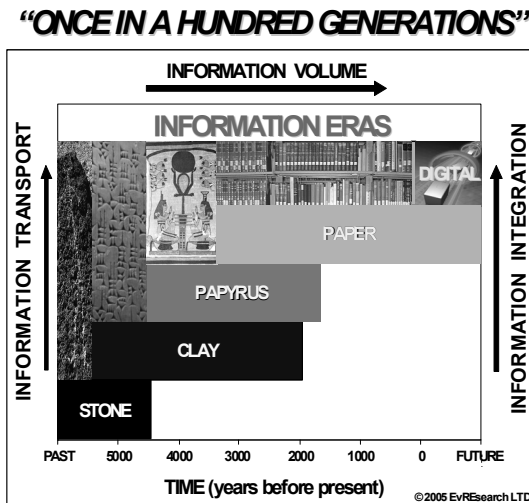
Source: Berkman, P.A. 2008. Once in a hundred generations. In: Halbert, M. and Skinner, K. (eds.). *Strategies for Sustaining Digital Libraries.* Emory University, Atlanta. Pp. 11–21. All rights reserved from EvREsearch

---

[12] Paul Berkman, 'Defining Digital Library Sustainability' (Paper presented at the Sustaining Digital Libraries Symposium, Atlanta, 6 October 2006) <http://www.metascholar.org/events/2006/sdl/viewpaper.php?id=6>.

The progression from different technologies has the advantage of increasing transportability. The volume in which the information can be compacted and the density of information that can be transported is increased and the ability to integrate different types of information improves along the trajectory. These are all positive benefits; however there is an important retrograde projection: fragility.

People can still read a Gutenberg bible printed six centuries ago, but it can be challenging to read magnetic media a decade or two old. Fragility increases along the trajectory and this is a significant challenge for preservation. It predicates a fundamental paradigm change in preservation strategies. An example of the loss of important information is the first electronic mail message. This was sent in 1964 from either MIT, the Carnegie Institute or Cambridge University, however the message does not survive and there is no record to determine which group sent the first email.[13] A less fortuitous example of loss is NASA losing more than 13 000 original tapes of the Apollo moon missions.[14]

A survey completed in 2006 by the United States National Library of Medicine found that of the 6 054 articles in 214 journal issues published in 2006, in the biomedical arena, 10% of the articles have linked digital information or supplementary digital information. What occurs to these links over time? Carmine Sellitto completed a study in 2004 (see Figure Seven) which showed that after just one year 10% of the links are broken.[15] The half-life of links in the study was approximately four and a half years.

---

[13] Report of the Task Force on Archiving of Digital Information (commissioned by The Commission on Preservation and Access and The Research Libraries Group, May 1996), 3 <http://www.digitalpreservation.gov/pdf/waters_garrett_final-report.pdf>.

[14] Seth Borenstein, 'NASA Plans New Search for Missing Moon Tapes' *Houston Chronicle* (Houston) 15 August 2006 <http://www.chron.com/disp/story.mpl/front/4116978.html>.

[15] Carmine Sellitto, 'A Study of Missing Web-cites in Scholarly Articles: Towards and Evaluation Framework' (2004) 30 *Journal of Information Science* 484 <http://jis.sagepub.com/cgi/reprint/30/6/484>.
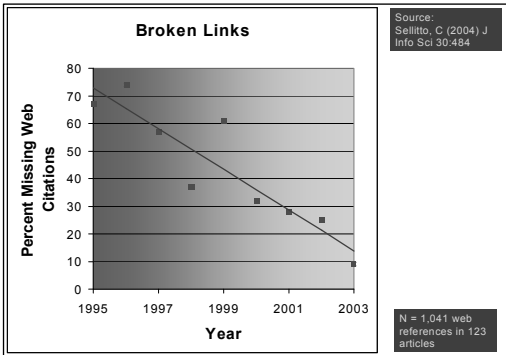
Figure Seven

Figure Eight[16] is a compendium of similar studies.   In 2002 legal citations were analysed and at the time were found to have a half life of less than one and a half years. The loss of information that has been published in the formal publication realm is significant and systematic and ranges in the analysis from one and a half years to four and a half years.

Figure Eight

Study	Resource type	Resource half-life
Koehler (1999 and 2002)	Random Web pages	2.0 years
Nelson and Allen (2002)	Digital Library Object	24.5 years
Harter and Kim (1996)	Scholarly Article Citations	1.5 years
Rumsey (2002)	Legal Citations	1.4 years
Markwell and Brooks (2002)	Biological Science Education Resources	4.6 years
Spinellis (2003)	Computer Science Citations	4.0 years

Source: Koehler W. (2004) Information Research, 9 (2), 174

---

[16] W Koehler, 'A Longitudinal Study of Web Pages Continued: A Consideration of Document Persistence' (2004) 9 (2) *Information Research* 174 <http://informationr.net/ir/9-2/paper174.html>.

There is an exception to the half life of citations as demonstrated in the studies. Digital objects that have been preserved in a formal digital repository, such as the Stanford Linear Accelerator Repository, the Harvard Digital Library and PubMed Central, have a half life of nearly 25 years. It could be argued that given the nature of these objects, this might be closer to the proper half life of digital objects. While some information should not be kept indefinitely, other information should be kept for a longer period of time. This illustrates the role of formal digital preservation organisations and their importance in the five dimensional world.

DATA PRESERVATION AND ACCESS: A SHARED VISION

A task force report issued in 1996 raises the challenge to 'commit ourselves [as a society] technically, legally, economically, and organizationally to the full dimension of the task of preservation and access.' [17] This is a fundamental challenge and progress has been made in globally recognising the nature of the challenge.

Organisation for Economic Co-operation and Development

The Organisation for Economic Co-operation and Development (OECD) believes that the issue of preservation of and access to research data is 'a matter of sound stewardship of public resources.' [18]

Digital Repository Infrastructure Vision for European Research

The Digital Repository Infrastructure Vision for European Research (DRIVER) project is of the opinion that 'any form of scientific-content resource . . . should be freely accessible through simple Internet-based infrastructures.'[19]

---

[17] Commission on Preservation and Access and the Research Libraries Group, *Report on the Task Force on Archiving of Digital Information* (1996).

[18] Organization for Economic Co-operation and Development, *Promoting Access to Public Research Data for Scientific, Economic and Social Development*.

[19] Digital Repository Infrastructure Vision for European Research (DRIVER) <www.driver-repository.eu>.

Canada

The National Consultation to Scientific Research Data (NCASRD) acknowledges the importance of a robust infrastructure framework for digital preservation and access and proposes the 'establishment of a dedicated national infrastructure . . . to assume overall leadership in the development and execution of a strategic plan [for digital data].'[20]

New Zealand

Creating Digital New Zealand: The Draft New Zealand Digital Content Strategy emphasises the importance of preserving the digital products of the current culture for future generations and 'providing the mechanisms to make it quick and easy. . . to find, share, access, use and re-purpose content.'[21]

Australia

The National Library of Australia's Preserving Access to Digital Information (PADI) initiative aims to 'ensure that digital information is managed with appropriate consideration for preservation and future access.'[22]

ACHIEVING THE VISION

The National Science Foundation has a vision in which 'science and engineering digital data are routinely deposited in well-documented form, regularly and easily consulted and analyzed . . . and openly accessible' whilst being reliably preserved.[23]

---

[20] The National Research Council Canada, *Final Report of the National Consultation on Access to Scientific Research Data* (2005) 3 <http://ncasrd-cnadrs.scitech.gc.ca/NCASRDReport_e. pdf>.

[21] National Library, *Creating Digital New Zealand: The Draft New Zealand Digital Content Strategy Discussion Document* (2006) 7 <http://www.digitalstrategy.govt.nz/upload/Main %20Sections/Content/NZ%20Digital%20Content%20Strategy%20Discussion%20Document. pdf>.

[22] Leanne Brandis and Jan Lyall, 'PADI: Preserving Access to Australian Information and Cultural Heritage in Digital Form' (Paper presented at the VALA Conference, Melbourne 28– 30 January 1998) <http://www.nla.gov.au/nla/staffpaper/lyall3.html>.

[23] National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery.*

There are three parts to achieving this vision. The first is that science and engineering data should be 'routinely deposited in well-documented form'.[24] This is not a technology challenge, because the technology exists. Instead this is a cultural change for incentives and motivations to deposit and provide documentation for data.

Secondly that data should be 'regularly and easily consulted and analyzed' by specialists and non-specialists. [25] There are deep research and technology challenges to providing information accessibility for those who are not highly specialised in the field of that particular collection.

Thirdly, data should be 'openly accessible while suitably protected, and reliably preserved'.[26]

In order to achieve this vision, it is necessary to have an infrastructure framework: a framework of repositories, libraries and reliable preservation organisations to provide for this function. In order to meet this vision the National Science Foundation has set itself two goals. Firstly to catalyse the development of a system of science and engineering data collections that is open, extensible and evolvable. While the National Science Foundation cannot meet the digital preservation needs of society as a whole, it can play an important role in demonstrating this ability and establishing more appropriate methodologies and capabilities.

Secondly the National Science Foundation will need to develop new tools and servers to enable this infrastructure framework.

Figure Nine is a schematic of what the National Science Foundation envisions. At the centre of the schematic are the users using the infrastructure. The nodes represent individual repositories or digital libraries and the edges between the nodes represent the links between them.

---

[24] National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery.*

[25] National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery.*

[26] National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery.*

Figure Nine
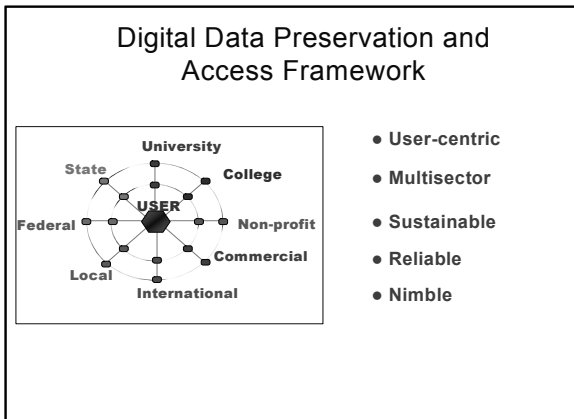
This schematic has several important features. Firstly the schematic is centred around the user and not around the infrastructure itself. The schematic occurs in sectors including federal, state, university, not-for-profit, commercial and international. The schematic should also be sustainable.

The schematic could be a schematic for the existing system of libraries preserving print information. There are international, national, local and university libraries which all have a different, but related, set of roles in preserving print information. These libraries also have a variety of business models through which they draw their funds from a variety of different sources in society.

The net result is a system which is robust and resistant to change in any one sector or catastrophic loss. This is the type of multi-sector sustainable framework that the National Science Foundation envisions. The framework should be reliable and the metrics for reliability and the technologies for reliability are important and are still being developed. The framework will also have to be nimble, because it operates in a swift current of constant technology change.
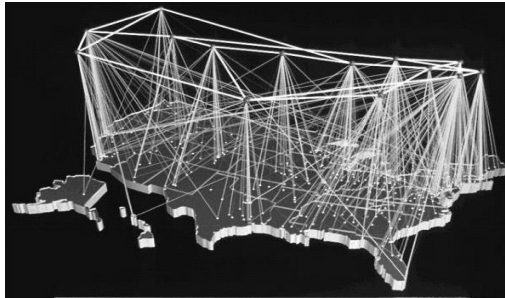
In summary the National Science Foundation's strategic plan includes promoting a change in culture, developing the preservation framework

and supporting the new generation of tools, services and capabilities that this framework will require.

Figure Ten is a graphic from the National Science Foundation summarising traffic on the NSFNet in September 1991. The NSFNet was arguably one of the best infrastructure investments the National Science Foundation made. NSFNet was the consolidation of two precursors, the ARPANet (the Defence Agency Network) and CSNet (the Computer Science Net) which were consolidated in order to provide access to the newly launched supercomputer centres in the United States.

Figure Ten

NSFNet Traffic– September 1991

Source: Visualization prepared by NCSA using data provided by Merit.

NSFNet was intended as an academic network. When it was launched in 1986 the National Science Foundation made the then outlandish claim that in five years time the NSFNet would connect up to 200 academic institutions with 10 000 users, which at the time seemed to be an immense goal. But by the end of 1992, when the NSFNet T1 net was decommissioned in favour of the T3 network there were one thousand institutions connected to the NSFNet with 10 million users.

Opening NSFNet to everybody resulted in the growth of the Internet. The Internet created connectivity, the ability to connect one machine to another without necessarily having to know in advance where that machine was located. It was that simple power of connectivity that

drove the emergence of the infrastructure and the opening of the fifth dimension.

The next driving force of this type is information integration, which requires reliable preservation. To bring about information integration, the National Science Foundation may have to start out relatively small with an initial datanet that is fairly simple and link together a small number of repositories. This will demonstrate the power of access to a wide variety of information and the ability to integrate that information. Then, if the force of integration is analogous to the force of connection, it is possible that after a short period of time, the datanet will grow exponentially.

There are two entities, or types of organisations, that are critical in this next stage. The first are the universities. Andre Oosterlinck states that the traditional function of the university is to create knowledge through research, disseminate knowledge through teaching and public outreach and preserve knowledge through the library systems of the university.

> Ever since their inception, universities have been occupied with the fundamental elements of what we now call 'knowledge management', i.e. the creation, collection, preservation and dissemination of knowledge.[27]

This responsibility is reflected in the mission statement of the University of California.

> The distinctive mission of the University is to serve society as a center of higher learning, providing long-term societal benefits through transmitting advanced knowledge, discovering new knowledge and functioning as an active working repository of organized knowledge.[28]

The universities are in a unique position. The mission of universities is consistent with the affirmation of the shared vision mentioned above. Some universities and libraries are amongst the oldest organisations in the world. The universities are organisations that have substantial information technology capabilities and faculties that generate digital

---

[27] Andre Oosterlinck, *Knowledge Management in Post-Secondary Education: Universities.(2002)*

[28] University of California, *Mission Statement.*

data and computer science breakthroughs, which are cyberinfrastructure advances that are critical to the evolution of the datanet concept.

Similarly the academic libraries have an important role to play. The Association of Research Libraries[29] is a group of 123 North American academic libraries whose mission is the preservation of digital assets.

> It is to the research library community that others will look for the preservation of . . . digital assets, as they have looked to us in the past for reliable, long-term access to the 'traditional' resources and products of research and scholarship.[30]

The University of Queensland library envisions a similar role in providing a link between people and information:

> The University of Queensland Library's mission is to link people with information, enabling the University of Queensland to achieve excellence in teaching, learning, research, and community service. [31]

I-Centre

However the current structure of the university and the university library is not optimal for the access to and preservation of digital information, and a new type of organisation is necessary. For the purposes of discussion this organisation will be called an 'I-Centre'.

It is necessary for the I-Centre to be risk-averse. It must have a timeline for reliable preservation of digital content that stretches into centuries while anticipating how people in the future will use the information that has been preserved. The expertise necessary for developing the I-Centre and its risk-averse capabilities lies in the library and archival sciences.

At the same time the hardware, software and people at the I-Center will be changing. Therefore the organisation has to be risk-capable, it has to be able to operate within a swift current of constant technology change and a steady exponential increase in the expectations of user who will

---

[29] Association of Research Libraries <http://www.arl.org/>.

[30] Association of Research Libraries, *ARL Strategic Plan 2005–2009* <http://www.arl.org/arl/governance/stratplan.shtml>.

[31] Keith Webster, University Librarian and Director of Learning Services.

want more from the cyberinfrastructure than what it is attempting to deliver.

For this reason, the organisation must have capabilities in computer science and computational science to anticipate the next generation of technologies, identify risks associated with those new technologies and plan reliable migration to the new technologies. This will be a constant occurrence through the life of the I-Centre.

Finally the user must be able to understand and access the information. This will require domain expertise necessary for understanding the deep contextual information associated with the information being preserved. An understanding of how the information will be used in the community will also be necessary, and this will require significant expertise in the respected domains.

Figure Eleven
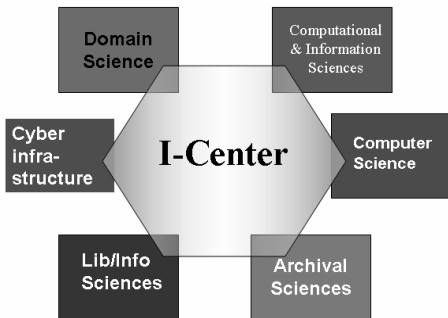
The I-Centre is an organisation that for the most part does not exist. A change may be required in the nature of digital preservation organisations, which will require new partnerships that are currently not present.

AN EFFECTIVE LEGAL FRAMEWORK

The vision of this conference, for an effective legal framework raises some fundamental issues.

The end result of the framework is information integration, everything else is the means towards achieving this. The goal of this cyber-

infrastructure framework is the ability to find, understand, access, use and re-use information. Any legal framework that inhibits or prevents information integration will inhibit the progress of those who operate under that framework.

The foundation of the framework has to be reliable digital preservation and access. If information is constantly lost, not accessible, or moving and changing in significant ways, the ability to effectively use the information over time is significantly decreased.

There are many types of data, for instance data that are public goods and data that are commercial commodities. An effective legal framework must recognise and support the various types of data. The legal framework should not focus on one single category of data, or a finite set of categories into which the data types can fit over time. Rather the framework should recognise the many different types of data being produced and the many different uses and needs for that data.

A world of five dimensions is inherently international, not national in character. In science the closest alignments between individuals are within disciplines, not within geographical regions. People operating in a world of five dimensions will be operating in an international framework. While this is a given for the five dimensional world, it must be an essential part of an effective legal framework.

It must be recognised that the fifth dimension does not arise automatically; it is built by individuals and organisations. The framework should enable individuals and institutions to pursue their innovative approaches to the infrastructure of the future.

Finally, there is a constant change in technologies, users' needs and expectations and opportunities. The legal framework must be built on the assumption that a static framework is dangerous and will almost certainly break immediately. The ability to accommodate a continuing change in the technologies is critical, and failing to do this will put the system at risk. The technology in this area will always improve and the legal framework should anticipate continuing change in this landscape.

CONCLUSION

The Office of Cyberinfrastructure is currently working on the technology challenges and opportunities that exist in creating this fifth dimension. In doing so it has been recognised that it is important to

have an adequate and robust legal framework to enable the technology innovations that are necessary. The legal work being done in this area is just as critical to the future of the five dimensional world as the technology that is being created.