

Tutkimusaineistojen tiedonhallinta

Tutkimushankkeiden
suunnitteluseminaari

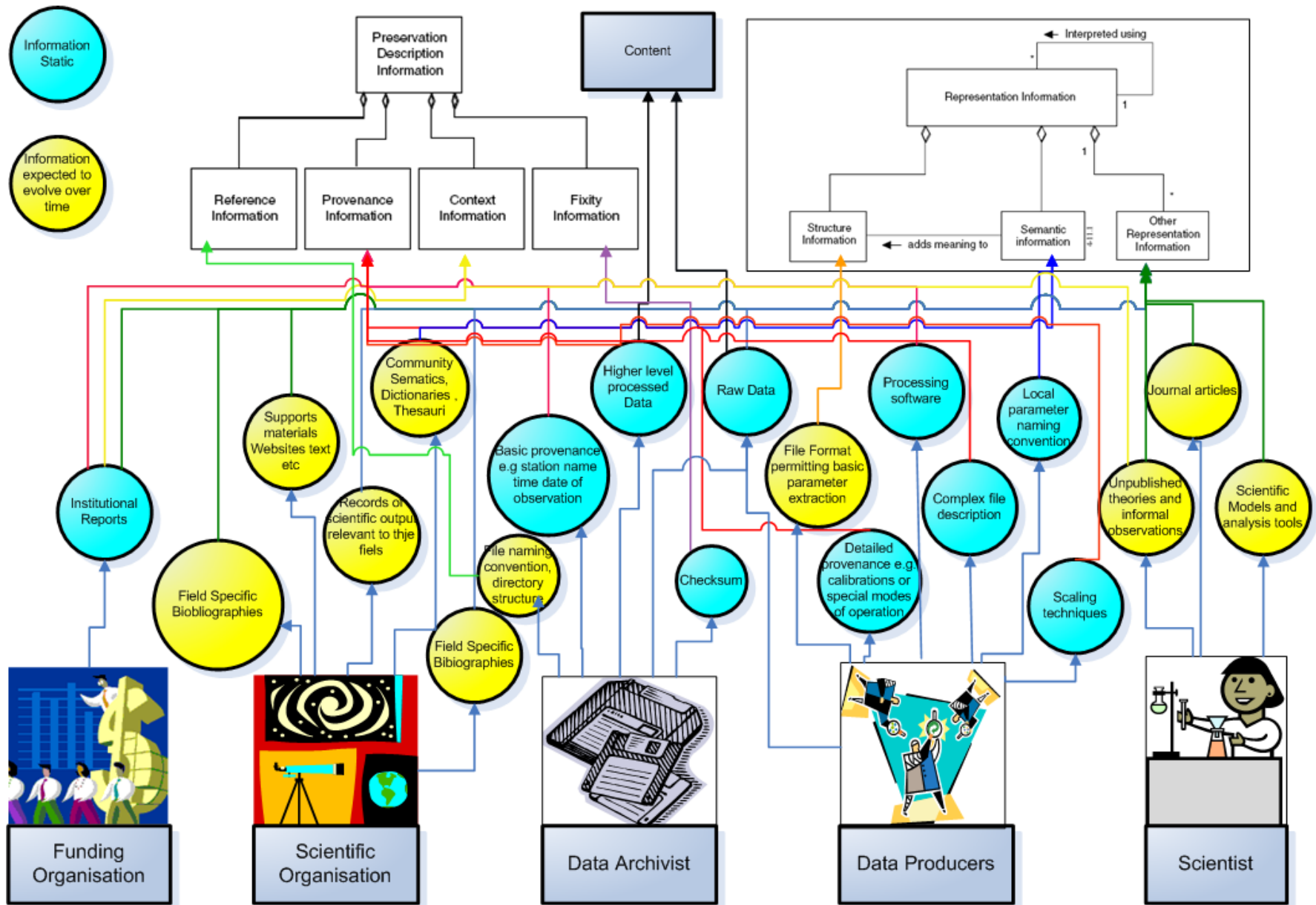
9.10.2009

Jussi Nuorteva



KANSALLISARKISTO
RIKSARKIVET

Managing scientific information



Data scientist

"The Solar Oscillations Investigation (SOI) is conducted by a team of researchers and their associated staff with interests and expertise covering a wide range of the fields of helioseismology and solar physics. The team consists, in the first place, of the Investigator Team consisting of the [PI](#) and the [Co-Investigators](#) selected by NASA and ESA to build and operate the Michelson Doppler Imager (MDI) and to process and analyze the data therefrom."



Dr. Richard S. Bogart

Data scientist

The Solar Oscillations Investigation (SOI)
Stanford University, CA, USA



KANSALLISARKISTO
RIKSARKIVET

Kansainvälinen kehitys

- **Revolutionizing Science and Engineering through Cyberinfrastructures (NSF Report 2003).**

"A new age has dawned in science and engineering research, pushed by continuing progress in computing, information and communication technology, and pulled by the expanding complexity of scope and scale of today's challenges.

Scientists in many disciplines have begun revolutionizing their fields by using computers, digital data, and networks to replace and extend their traditional efforts. The calculations that can be performed and the information that can be archived and used are exploding.

The emerging vision is to use cyberinfrastructure to build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools and instruments and that operate at unprecedented levels of computational, storage, and data transfer capacity."



National Science Foundation
WHERE DISCOVERIES BEGIN

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT | FastLane

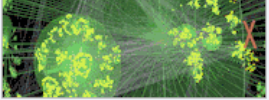
• Find Funding • Recent Funding • Due Dates • Preparing Proposals • About Funding

 **National Science Foundation**
OFFICE OF
Cyberinfrastructure

SEARCH
NSF Web Site

OCI Home | OCI Funding | OCI Awards | OCI Discoveries | OCI News | About OCI

Office of Cyberinfrastructure (OCI)



OCI Home
About OCI
Funding Opportunities
Awards
News
Events
Discoveries
Publications
Advisory Committee
Career Opportunities
See Additional OCI Resources
View OCI Staff
Search OCI Staff

Proposals and Awards
Proposal and Award Policies and Procedures Guide
Introduction
Proposal Preparation and Submission
• Grant Proposal Guide
• Grants.gov Application Guide
Award Administration
• Award Administration Guide

Press Release 06-098
Collaboration Will Preserve Digital Research Data

The National Archives and the San Diego Supercomputer Center sign landmark agreement



The National Archives and San Diego Supercomputer Center will preserve national digital data.
[Credit and Larger Version](#)

June 29, 2006

The National Archives and Records Administration (NARA) and the San Diego Supercomputer Center (SDSC), with concurrence from the National Science Foundation (NSF), have signed a landmark Memorandum of Understanding (MOU) that provides an avenue for preserving some of the nation's most valuable digital research collections. This collaboration marks the first time NARA has established an affiliated relationship for preserving digital data with an academic institution.

In accordance with NARA's mission, the agreement provides for governance of historically valuable federal, presidential and congressional electronic records, as well as other scientific, engineering and research-related digital data collections. The SDSC, an organized research unit of the University of California-San Diego, is a leader in cyberinfrastructure with a national data repository featuring unique large-scale data storage, preservation, management and analysis facilities.

"The National Archives and the San Diego Supercomputer Center sign landmark agreement

... The partnering of NARA, SDSC and NSF comes at a time when the nation's scientists and engineers are seeking to increase U.S. competitiveness and leadership--and when the massive amounts of digital data being generated by researchers, educators and practitioners are demanding new and innovative strategies for digital preservation"



KANSALLISARKISTO
RIKSARKIVET

Archiving data for research

"Digital information compiled for research purposes is playing an increasingly important role in today's knowledge economy. In many ways, data is the fuel driving innovation and our capacity to address complex social and economic problems.

"...The core mission of a research data-archive is not to preserve the recorded memory of a group, organization or nation, but to provide a vital service to the research community."



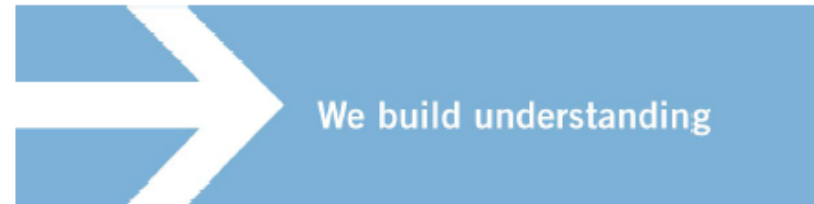
Social Sciences and Humanities
Research Council of Canada

National Archives
of Canada

Conseil de recherches en
sciences humaines du Canada

Archives nationales
du Canada

Canada



Final Report

National Data Archive Consultation

Building Infrastructure for Access to and
Preservation of Research Data

Submitted by the NDAC Working Group to
the Social Sciences and Humanities Research
Council of Canada and the National Archivist
of Canada

June 2002

www.sshrc.ca/web/about/publications/da_finalreport_e.pdf



KANSALLISARKISTO
RIKSARKIVET

• OECD 2004

"The rapid development in computing technology and the Internet have opened up new applications for the basic *sources* of research — the base material of *research data* — which has given a major impetus to scientific work in recent years.

Databases are rapidly becoming an essential part of the infrastructure of the global science system."

OECD Principles and
Guidelines for Access
to Research Data from
Public Funding



www.oecd.org/dataoecd/9/61/38500813.pdf



KANSALLISARKISTO
RIKSARKIVET

OECD Principles (2004)

- "Throughout OECD member countries, continuously growing quantities of data are collected by publicly funded researchers and research institutions. This rapidly expanding body of research data represents both a massive investment of public funds and a potential source of the knowledge needed to address the myriad challenges facing humanity."
- "Access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators."
- "Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs."



KANSALLISARKISTO
RIKSARKIVET



OECD:n datasuosituksen toimeenpano Suomessa

SAMI BORG JA ARJA KUULA

**Julkisrahoitteisen tutkimusdatan
avoin saatavuus ja elinkaari**

VALMISTELURAPORTTI OECD:N DATASUOSITUKSEN
TOIMEENPANOMAHDOLLISUUKSISTA SUOMESSA

Open Access – avoin saatavuus

- Open access-liikkeen tavoitteena oli edistää tiedon avointa saatavuutta
- Taustalla tieteellisen julkaisemisen kustannusten voimakas nousu
- Jean-Francoise Lyotard, *La conditionne postmoderne: rapport sur le savoir* (1979)
- 1991 Linux (Linus Torvalds, Finland) – avoimeen lähdekoodiin perustuva käyttöjärjestelmä
- 1991 Arxiv (www.arxiv.org) Paul Ginspargin Cornell University'ssa kehittämä julkaisukanava ja arkisto. Nykyisin noin 500,000 sähköistä julkaisua matematiikan ja fysiikan alalta. 2002 Grigori Perelman (Venäjä), sai Fieldsin mitalin siellä julkaisemiensa tutkimusten ansiosta 2006.



KANSALLISARKISTO
RIKSARKIVET



**COUNCIL OF
THE EUROPEAN UNION**



Council Conclusions on scientific information in the digital age: access, dissemination and preservation

*2832nd COMPETITIVENESS (Internal market, Industry and Research) Council meeting
Brussels, 22 and 23 November 2007*

- reinforce national strategies and structures for access to and preservation and dissemination of scientific information, tackling organisational, legal, technical and financial issues;
- enhance the co-ordination between Member States, large research institutions and funding bodies on access, preservation and dissemination policies and practices;
- maximise access for researchers and students to scientific publications, in particular by improving public procurement practices in relation to scientific information; this could include exchanging information on these practices and increasing the transparency of the contractual terms of "big deals", and exploring the possibilities for funding bodies, research institutions and scientific publishers from different Member States to work together in order to achieve economies of scale and efficient use of public funds by demand aggregation;
- ensure the long term preservation of scientific information - including publications and data - and pay due attention to scientific information in national information preservation strategies;

Euroopan unioni

- **Green Paper – The European Research Area: New Perspectives (2007)**

”2. Maailman huippuluokan tutkimusinfrastruktuuri.

Merkittäviä infrastruktuureja tulisi luoda ja käyttää Euroopan yhteishankkeina. Niiden pitäisi olla koko Euroopan ja maailman tutkimusryhmien käytössä, ja Euroopassa työskentelevien tutkijoiden pitäisi voida käyttää kansainvälisiä infrastruktuureja ja laitteistoja muualla maailmassa.

Infrastruktuurien tulisi olla integroituja ja verkotettuja Euroopassa ja maailmanlaajuisesti, mihin päästään samanaikaisesti kehitettävien uuden sukupolven sähköisten viestintäinfrastruktuurien avulla.”



KANSALLISARKISTO
RIKSARKIVET

Euroopan unioni 2

- **Green Paper – The European Research Area: New Perspectives (2007)**

”Monet ehdotetuista infrastruktuurihankkeista ovat niin isoja ja laajakantoisia, että ne edellyttävät maailmanlaajuisia yhteistyötä.

Näillä sähköisillä infrastruktuureilla on olennainen merkitys osien yhteentoiminnan kannalta, sillä niiden avulla voidaan poistaa perinteiset ajalliset, maantieteelliset ja tieteenalojen ja instituutioiden rajojen asettamat esteet.

Tämän vuoksi välttämättömänä edellytyksenä on huolehtia Euroopan tiede- ja teknologiainfrastruktuurien ja uusien sukupolvien sähköisten verkkojen keskenään johdonmukaisesta suunnittelusta, samanaikaisesta kehittämisestä ja yhteentoimivuudesta.”



KANSALLISARKISTO
RIKSARKIVET

Euroopan unioni 3

- **Komissaari *Viviane Reading* 16.2.2007:**

"The use and re-use of scientific data through applications such as data-mining, as well as possibilities to combine journal articles with the underlying research data is gaining importance in many scientific disciplines. And it is generally believed that this kind of application has an enormous potential for the future.

Indeed, we can already see a trend towards a **continuum of the scientific information space, from data to publications.**

In parallel, **scientific institutions and research communities are setting up digital repositories** in which they collect the output of their researchers, sometimes in combination with the experimental data.

Some funding bodies have started requiring from their researchers that they deposit their research output in open repositories, accessible to all."



KANSALLISARKISTO
RIKSARKIVET

ICT Infrastructures for eScience

- **EU Commission communication 5.3.2009, COM(2009) 108 final**

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF>

- "This Communication *highlights* the strategic role of ICT infrastructures as a crucial asset underpinning European research and innovation policies, and *calls* on Member States and the scientific communities, in cooperation with the European Commission, for a reinforced and coordinated effort to foster world-class ICT infrastructures, also known as *e-Infrastructures*, to pave the way for the scientific discoveries of the 21st century."



KANSALLISARKISTO
RIKSARKIVET

Kehittämislinjoja Suomessa

- Suomen eScience-ohjelma (Opetusministeriön työryhmämuistiota ja selvityksiä 2007:7)
- Laskennallisen tieteen kehittäminen Suomessa (Opetusministeriön työryhmämuistiota ja selvityksiä 2007:23)
- Kansallisen tason tutkimusinfrastruktuurit-tiekartta
- Tieteen tietotekniikan keskus CSC - tutkimustiedon hallinnan kehittäminen ja osallistuminen eurooppalaisiin ja kansainvälisiin grid-ohjelmiin



KANSALLISARKISTO
RIKSARKIVET

Kansallisen tason tutkimusinfrastruktuurit

- *Kansallisen tason tutkimusinfrastruktuurit. Nykytila ja tiekartta* (Opetusministeriön julkaisuja 2009:1)
- Tällä hetkellä kiireellinen tehtävä on saattaa tutkimuksen kannalta tärkeät aineistot digitaaliseen muotoon ja turvata alkuperäisaineistojen säilyminen tuleville sukupolville tehokasta hyödyntämistä varten.
- **Suositus 1.** Kansallisten rekistereiden käytettävyyttä ja aineistojen saatavuutta tulee parantaa ja käyttäjän kustannuksia alentaa, tarvittaessa ao. lainsäädäntöä muuttamalla. Suomessa kerätyt arvokkaat aineistot tulee saattaa laajempaan kansainväliseen käyttöön lisäämällä aineistojen digitointia, ja yhdenmukaistamalla aineistojen keruu vastaamaan kansainvälisiä standardeja.
- **Suositus 2.** Maassamme tarvitaan yhteinen näkemys siitä, millainen e-infrastruktuuri palvelee parhaiten korkeatasoista tutkimusta.



KANSALLISARKISTO
RIKSARKIVET



Kansallinen digitaalinen kirjasto

- *Hankkeessa parannetaan kirjastojen, arkistojen ja museoiden sähköisten aineistojen saatavuutta ja pitkäaikaissäilytystä.*
- Arkistot, kirjastot ja museot ovat keskeinen osa humanistisen ja yhteiskuntatieteellisen tutkimuksen infrastruktuuria
- Digitoidun aineiston luomisessa käytetyn tekniikan ja siihen liitetyn metadatan tulee vastata käyttötarpeita ja pitkäaikaisen säilyttämisen vaatimuksia niin, ettei aineistoa tarvitse digitoida myöhemmin uudestaan.
- KDK on osa Euroopan digitaalista kirjastoa (Europeana) <http://www.europeana.eu/portal/>



KANSALLISARKISTO
RIKSARKIVET



SUOMEN AKATEMIA

- *Akatemia edellyttää hakijoilta selvitystä siitä, miten hankkeen tutkimusaineistot tullaan hankkimaan, miten niitä käytetään ja säilytetään ja miten mahdollistetaan niiden myöhempi käyttö. Tämä tiedonhallintasuunnitelma esitetään tutkimussuunnitelmassa.*
- *Akatemia suosittelee, että hankkeet luovuttavat kokoamansa yhteiskuntatieteellisen aineiston Yhteiskuntatieteellisen tietoarkiston www.fsd.uta.fi käyttöön.*
- Tiedonhallintasuunnitelma on edellytetty hankehakemuksilta syyskuun 2008 hausta lukien



- Kokonaiskustannusmalli (Full cost model) otettiin käyttöön Suomen Akatemiassa, yliopistoissa ja valtion tutkimuslaitoksissa 1.1.2009 lukien – todellisten kustannusten määrittely

Mikä on tiedonhallintasuunnitelma?

- **Tietoaineistojen hankkiminen tutkimushankkeelle** (mitä, mistä, tiedonhallinnan tarvitsemat resurssit)
- **Tiedonhallintaympäristö** (käytettävät ohjelmat ja laitteet, Virtual Research Community – keillä on pääsy tietoon, tietoarkkitehtuuri, kuka ylläpitää, informaattikko?, data scientist?)
- **Tiedon tuottaminen** (datan tuottamisessa ja hallinnassa käytettävät kansainväliset standardit ja yhteistyö, pysyvästi säilytettävän datan määrittely ja säilyttämisestä päättäminen, säilyttäjän kanssa tehtävä yhteistyö, FinnOA)
- **Julkaiseminen** (sähköiset julkaisut, linkittäminen tutkimusaineistoihin?, open access vai kaupallinen kustantaja, sähköisten julkaisujen ylläpito ja pysyvä säilytys)



KANSALLISARKISTO
RIKSARKIVET

FSD – Yhteiskuntatieteellinen tietarkisto

- *Tutkimusaineistojen tiedonhallinnan käsikirja* (2009): www.fsd.uta.fi/tiedonhallinta/index.html
- Tiedonhallinnan periaatteet yhteiskuntatieteissä
- Osa hyvää tieteellistä käytäntöä ja tutkimuksen eettistä toimintatapaa
- Tutkimusaineistojen avoin saatavuus ja tekijänoikeudet entistä keskeisempiä haasteita
- Kvantitatiiviset ja kvalitatiiviset aineistot
- Aineistojen säilytys ja käytettävyys – metatietomallit
- Sopii hyvin muillekin tutkimusaloille!



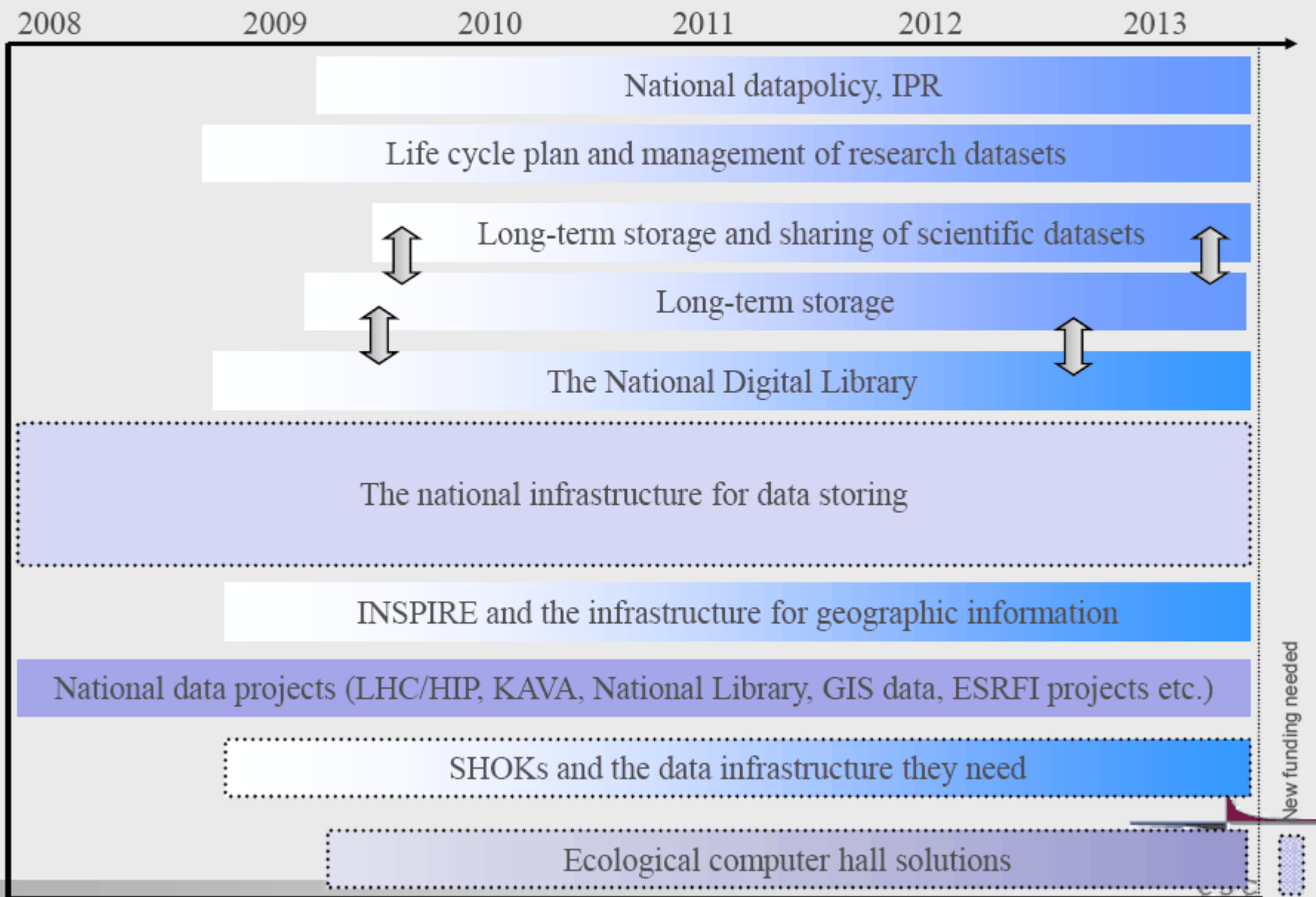
KANSALLISARKISTO
RIKSARKIVET



C S C

CSC — IT Center for Science Ltd





Research e-infrastructure in Finland

Structural reforms

- access to national e-infra for state research institutes
- access to national e-infra for industry and SHOKs
- access to international e-infra
- new international e-infra
- joint development projects
- contract arrangements between e-infra producers and users
- control structures and monitoring

Competence development

- EU level e-infra competence center
 - collaboration
 - computational science competence
 - software development competence
 - effective knowledge transfer
- widely accessible expert services
- e-infra consultation services
- support for communities and competence networks

Shared content

- access to digital repositories
- open standards and interfaces
- management of life cycle and metadata
- scientific applications
- collaborative tools

ICT solutions

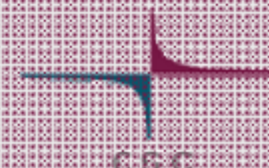
- machine room facilities
- computer networks
- computing capacity
- data management
- middleware and user interfaces
- user authentication
- IT architecture specification and implementation
- grid technology and Finnish Grid Initiative
- software development

Universities and polytechnics

State research institutes

Other expert organizations

EU and the global research system



Tutkimuksen teon käytäntöjä on uudistettava!

- Digitaaliset aineistot mahdollistavat tiedon nopean käytön ilman paikan asettamia rajoitteita
- Viittauskäytäntöjä on kehitettävä niin, että sähköisestä julkaisusta on suora linkki digitoituun asiakirjaan tai päättelyssä käytettyyn tutkimusdataan
- Tutkijoiden on tehtävä entistä tiiviimmin yhteistyötä tutkimuksen sähköisen infrastruktuurin kehittämiseksi
- Kansainvälisessä yhteistyössä on nostettava esiin kysymys yhteisistä aineistoista ja viittauskäytännöistä



KANSALLISARKISTO
RIKSARKIVET

Mitä hyötyä tästä kaikesta on?

- Tehokkaampi suunnittelu, sujuvampi tutkimusprosessi
- Parempi aineistohallinta – tutkimuksen validiteetin ja reliabiliteetin lisääntyminen
- Kansainvälisen yhteistyön pitkäjänteisempi suunnittelu – verkottuneen toimintatavan kehittäminen
- Sähköisen tiedonhallinnan parempi metodinen osaaminen
- Parempi kyky hyödyntää tarjolla olevaa tutkimusdataa uuden tiedon tuottamisessa
- Kansainvälisesti korkeatasoinen tutkimus!



KANSALLISARKISTO
RIKSARKIVET