

The European Approach

Wouter Spek

Berlin, 10 June 2009

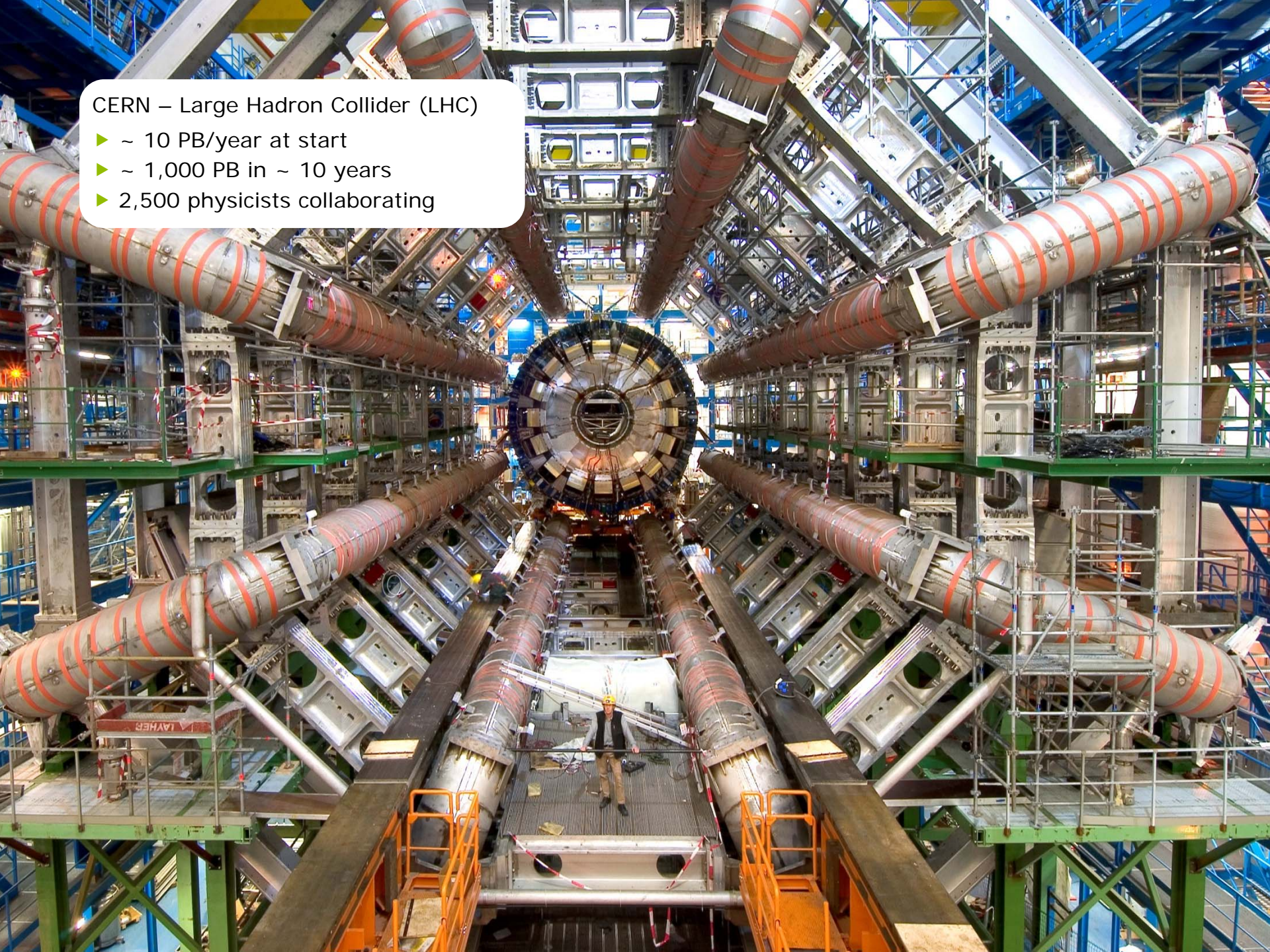
Plinius Major – Plinius Minor



Today vulcanologists still use the writing of Plinius Minor to discuss this eruption of the Vesuvius

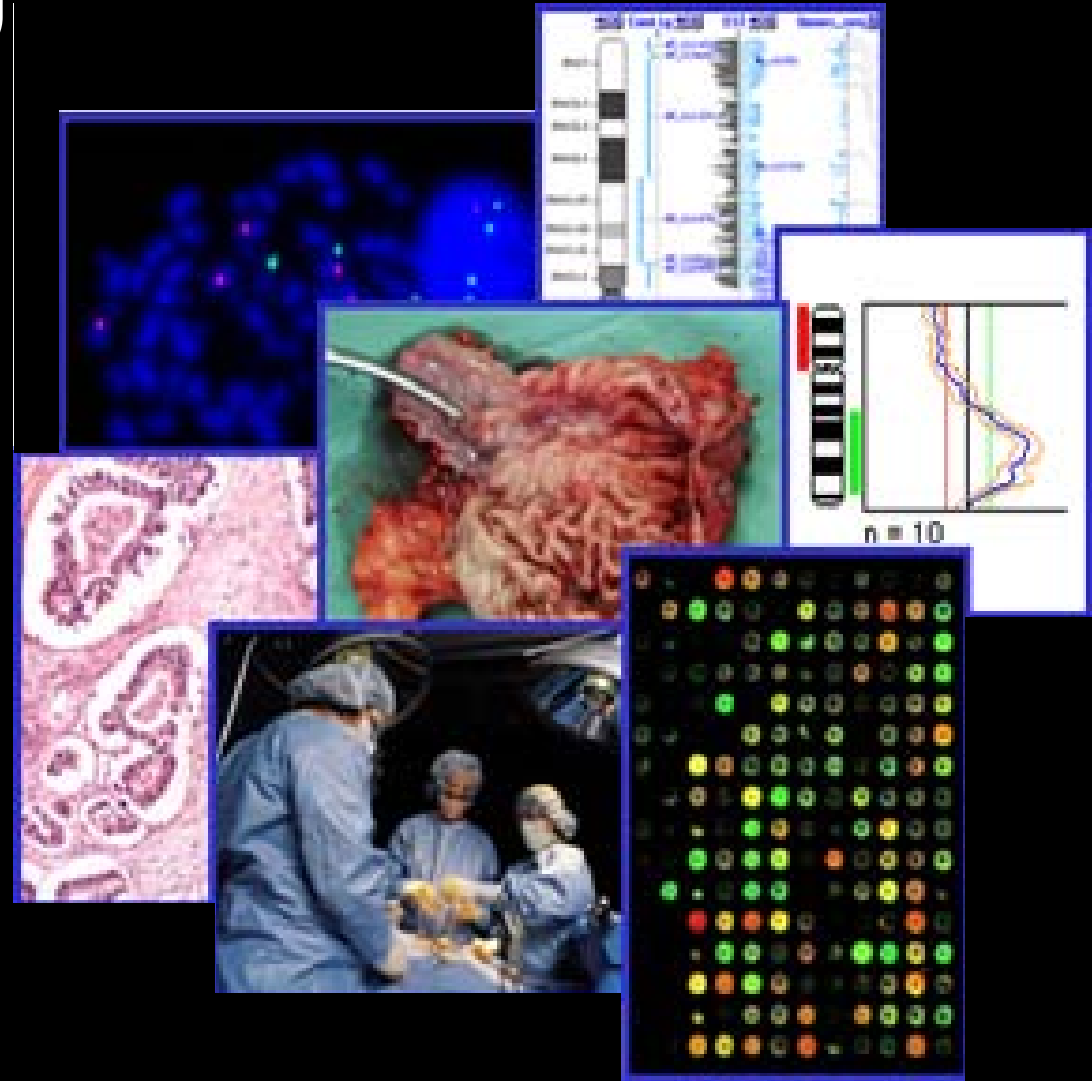
CERN – Large Hadron Collider (LHC)

- ▶ ~ 10 PB/year at start
- ▶ ~ 1,000 PB in ~ 10 years
- ▶ 2,500 physicists collaborating



Terabytes – Data Explosion Science where is it going?

International Cancer Genome Consortium
1,000 genomes / ICGC = ~1 petabyte



Storing and accessing the digital records of science: a serious strategic problem



ESA: > 30% of requests for earth observation data for old data, number is increasing; also ESA offers its capabilities to the tens of repositories with earth observation data to get common solutions

Objectives

- To support the development of a sustainable European digital information infrastructure
- To partner with national governments and European Commission
- To collaborate with European Science Consortia
- To initiate and facilitate strategic alliances between key stakeholders
- To strengthen the role of European parties in worldwide collaboration

Justification

- Science has changed in part because informatics has changed.
- Research is generating rapidly increasing volumes of data...
- ... a diverse range of formats like images, data, publication...
- New knowledge and know-how depend more and more on the on the ability to re-use and share data

So ...


We need to take action!

USB-stick gegevens 12 miljoen Britten verloren

Gepubliceerd door Jochem de Goede op maandag 3 november 2008 07:10
4 reacties - views 1139

Op een parkeerplaats in het Verenigd Koninkrijk is een USB-stick gevonden die aanmeldgegevens bleek te bevatten van 12 miljoen Britten voor een belangrijke overheidssite. De stick behoorde tot Atos Origin, een IT-dienstverlener.

De aanmeldge
Government G
krijgen tot hon
belastingterug
Naast de aanm
op de USB-stic

 Uw advertentie
klik voor meer info

Home | Nieuws | Forum | Columns | Achtergrond | Poll | Dossiers | Nieuwsbrief

Nieuws

USB-stick met NAVC
04-01-2008, 14:34 door Redact
Reacties: 11

Een USB-stick met zeer
defensie-minister van Liban
zeer specifieke informatie c
over de situatie in Afghanist
elite-soldaten.

Wanneer de vertrouwelijke i
voor Zweedse ISAF-soldater

In 2006 liet een kapitein
informatiedrager bevatte tal
defensie-minister Kamp.

TIMESONLINE

NEWS COMMENT BUSINESS MONEY SPORT LIFE & ST
UK NEWS | WORLD NEWS | POLITICS | ENVIRONMENT | WEATHER

Where am I? > Home > News > Politics

From [The Sunday Times](#)
January 25, 2009

Loss of British Council staff data David Miliband

David Leppard

THE foreign secretary, David Miliband, faced embarrassment last night after it emerged that a computer disk containing confidential bank details of up to 2,000 public servants working for the British Council has been lost.

The council — the Foreign Office's cultural arm — admitted that the disk, which held the names, National Insurance numbers, salary and bank account details of staff on its UK payroll, went missing while in transit last month.

The incident is the latest in a string of losses which has led to criticism of the government's handling of sensitive data.

In November a memory stick holding data said to give access to tax and benefit records was lost in a pub car park, forcing officials to suspend the government's gateway website.

HSBC loses customers' data disc

The HSBC banking group has admitted losing a computer disc with the details of 370,000 customers.

The disc was lost four weeks ago after being sent by courier from the bank's life insurance offices in Southampton.

The customers' details included their names, dates of birth, and their levels of insurance cover.

However, there were no addresses said the customers' exposure to p

"We are looking into it and basically an HSBC spokesman.

"The reinsurer we sent it to is doing We will do anything we can to find

"There are no financial details there are no address details or anything

As well as name, date of birth and revealed only the customer's policy customer was a smoker.



The Local SWEDEN'S NEWS IN ENGLISH

News Travel Lifestyle Galleries Dating Museums Blogs ^{NEW} Disc
National Money Politics Sci & Tech Society Education Sport Opinion Pi

Secret Swedish military files left in library

Published: 4 Jan 08 11:08 CET

Dictionary tool Double click on a word to get a translation

A USB-memory stick loaded with secret documents from the Swedish Armed Forces was found in a Stockholm library, reports Aftonbladet.

- **Swedish major facing sack over boozy African sex romp** (1 Jun 09)
- **Thailand to delay Gripen purchase: report** (22 May 09)
- **Spy ship rescues Russian fishermen** (18 May 09)

The secret documents included material related to Nato's ISAF peace keeping force in Afghanistan, as well as an intelligence report from a private American security company analyzing the murder of the Sri Lankan foreign minister and the attempted assassination of Lebanon's defence minister.

"We take this kind of carelessness very seriously. Its primarily a matter of security for our soldiers," said Colonel Bengt Sandström, from the Military Intelligence and Security Service (MUST).

Threats & Opportunities

Case Study: The 1986 Domesday project



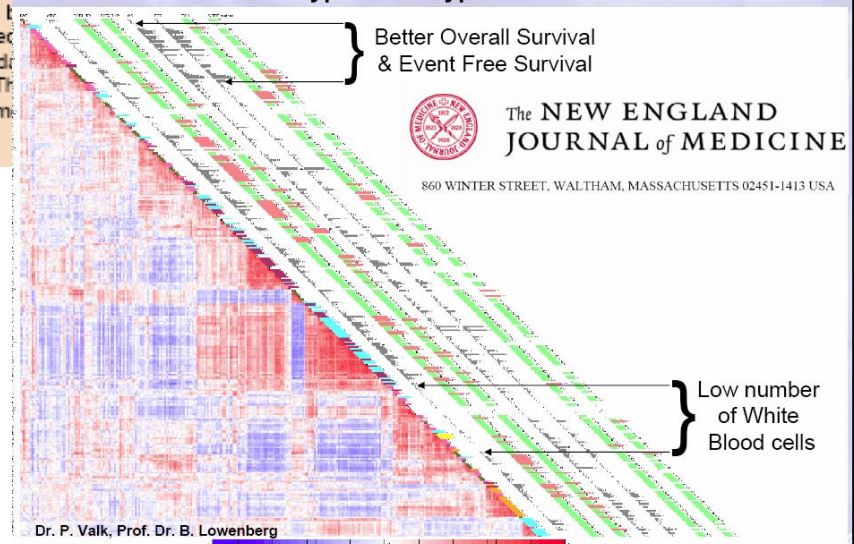
Copyright The National Archives, Catalogue reference: E 31/2/2 f.238a

The original Domesday book, which was produced on vellum in 1086 for William the Conqueror, is still readable today at The National Archives, over 900 years later.

To mark the 900th anniversary of the original Domesday book, the BBC launched the 1986 Domesday project. This consisted of thousands of photographs and maps combined with statistical, written and visual information to produce a picture of Britain in 1086.

The project was recorded onto 30cm laserdiscs and could be viewed with software running on BBC Microcomputers. Less than 20 years later, not only are the 30cm laserdiscs obsolete, but also the necessary hardware. Separate projects were launched by The National Archives and by Leeds University to rescue the data and the software that made up the 1986 Domesday project. This was only achieved thanks to a surviving laserdisc player and months of a year's effort by specialist teams.³

Leukemia Patient Classification - Genotype-Phenotype Correlation -



Courtesy Professor Peter van der Spek – Erasmus Medical Centre

Mind the Gap – Digital Preservation Coalition, 2006

Alliance members



Activities

- Promote strategic agenda for European e-science infrastructure
- Creating a stakeholders forum for interaction
- Facilitate linkage between different scientific communities
- Annual conference: for governments, EU, funding agencies, research organisations, FP7 and ESFRI projects, publishers, national libraries, archives etc.
- PARSE.insight

PARSE.insight



Facts and Figures:

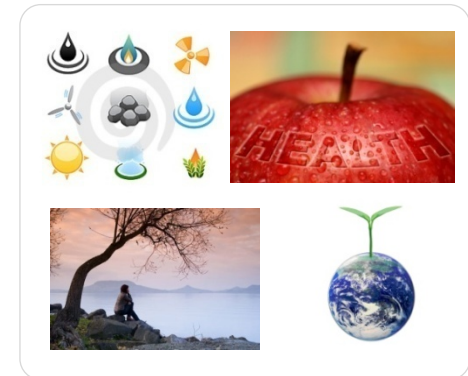
- ▶ Duration: March 2008 – February 2010
- ▶ 9 partners ranging from libraries to universities and large research organisations
- ▶ EC funding: M€1,8, through the FP7 ICT Work Programme
- ▶ Website: www.parse-insight.eu

PARSE.insight

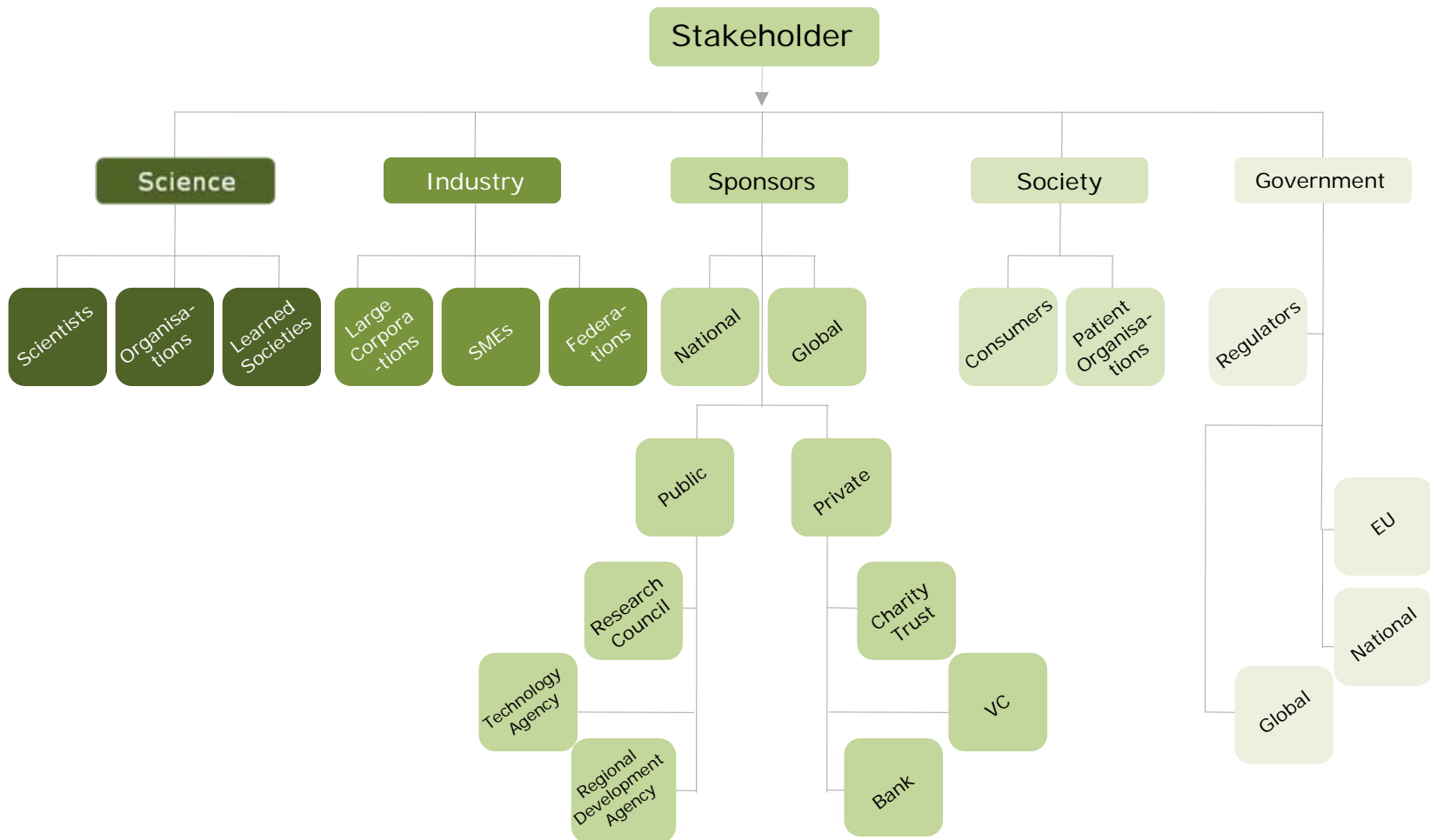


- Insight into issues of Permanent Access to the Records of Science in Europe
- Objectives:
 - ▶ Develop a roadmap and recommendations for building the e-Science infrastructure in order to maintain long-term access and use of scientific digital information in Europe

Grand challenges



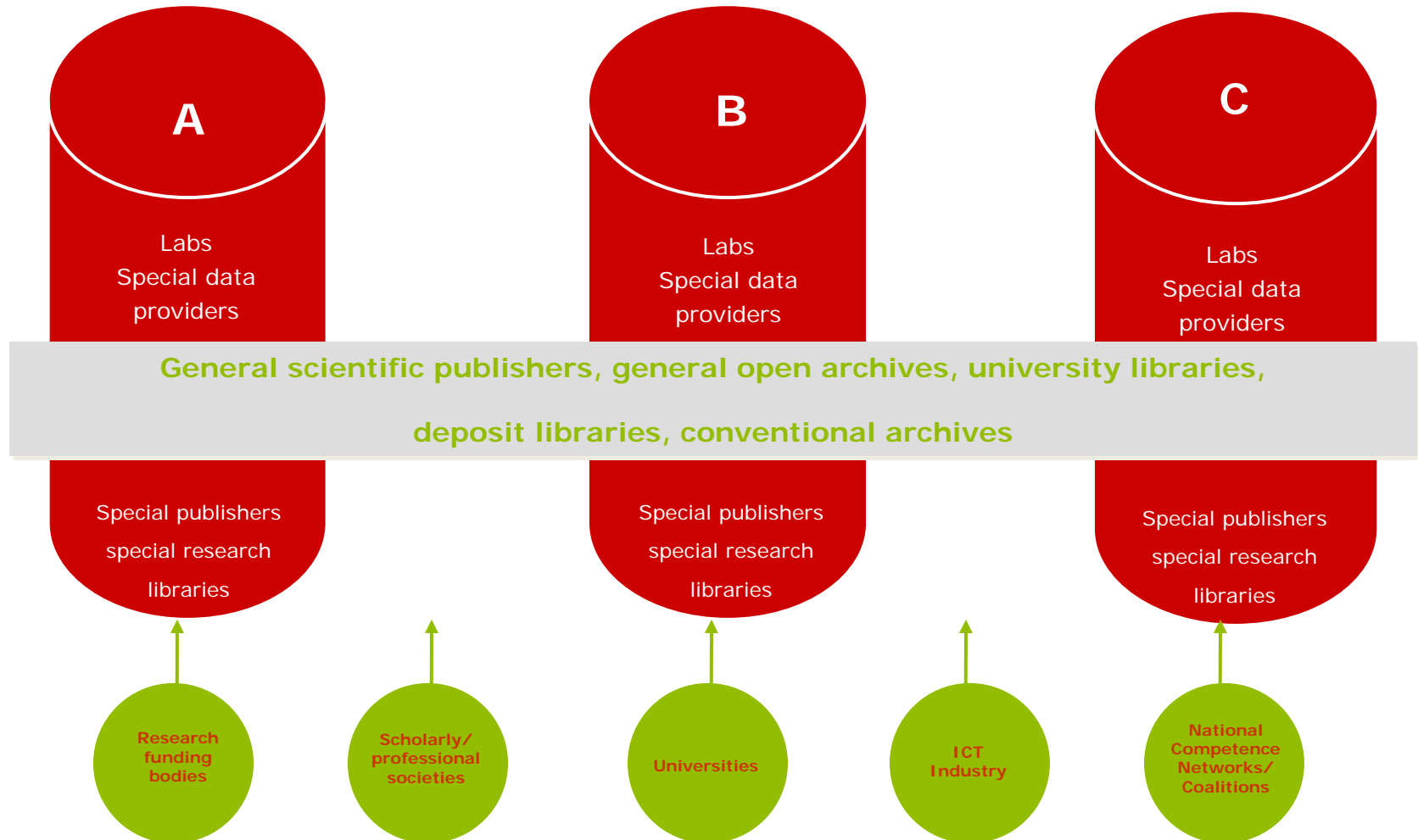
Whose problem is it anyway, who is picking up the check?



Towards permanent access

- Management & Organisation
- Policy & strategy
- Standardisation
- Technical issues
- Legal Framework
- Finance

Management & Organisation



Community example - ELIXIR


ELIXIR
EUROPEAN LIFE SCIENCE INFRASTRUCTURE FOR BIOLOGICAL INFORMATION

www.elixir-europe.org




ELIXIR: DATA FOR LIFE

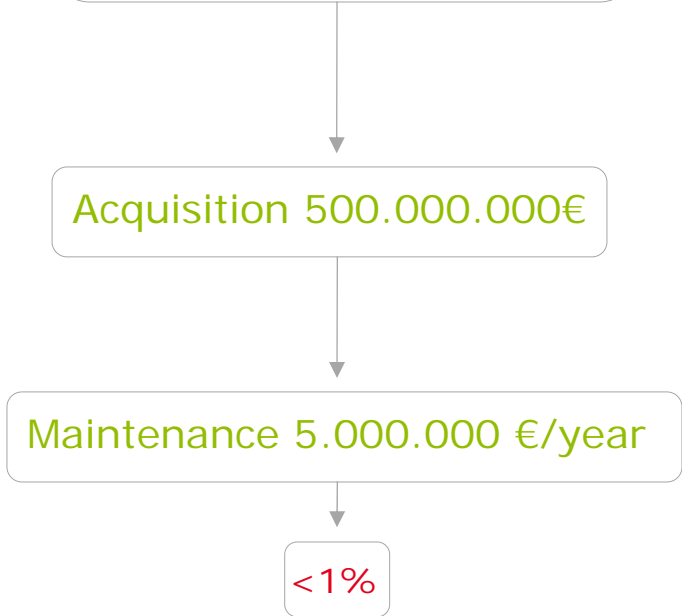
Imagine going to your computer to look up the sequence of a gene that you're working on, but the sequence database has disappeared. Suddenly you realise that the entire EST library that you characterised a few years ago has vanished without trace and you don't have any other record of it.

We need your support to secure the future of Europe's biological data and make sure that this scenario stays in the realms of science fiction.

Sequencing the human genome alone cost 3,000 M€. Compared with the costs of generating sequencing data, the cost of storing all known DNA sequence information in openly accessible databases that researchers everywhere can use is very small – probably less than 1%. **If we don't ensure the future of our databases, it's as if we didn't collect the data in the first place.**

ELIXIR's objective is to secure funding commitments from government agencies, charities, industry and intergovernmental organisations throughout Europe, to strengthen and sustain a world-class infrastructure for the management and integration of information in the life sciences.



Community example - CLARIN



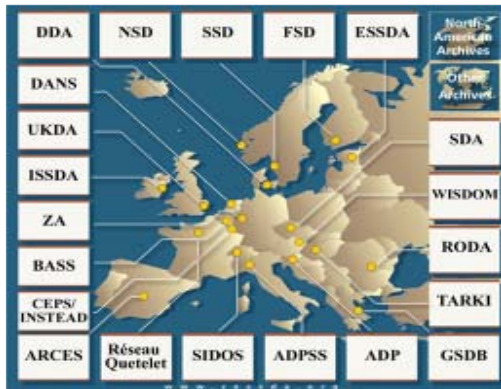
Estimated construction cost	106 M€
First open access foreseen	2008

Common Language Resources and Technology Infrastructure:

- Distributed infrastructure making available language resources and technology to researchers and scholars of all disciplines, in particular the humanities and social sciences
- Harmonises structural and terminological differences
- Based on a Grid-type infrastructure and using Semantic Web technology

www.mpi.nl/clarin

Community example - CESSDA



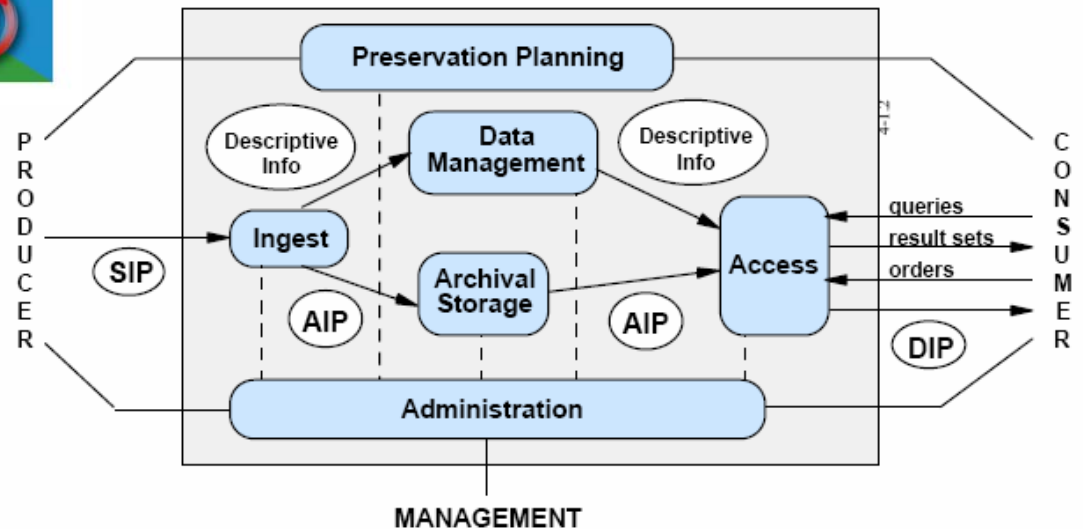
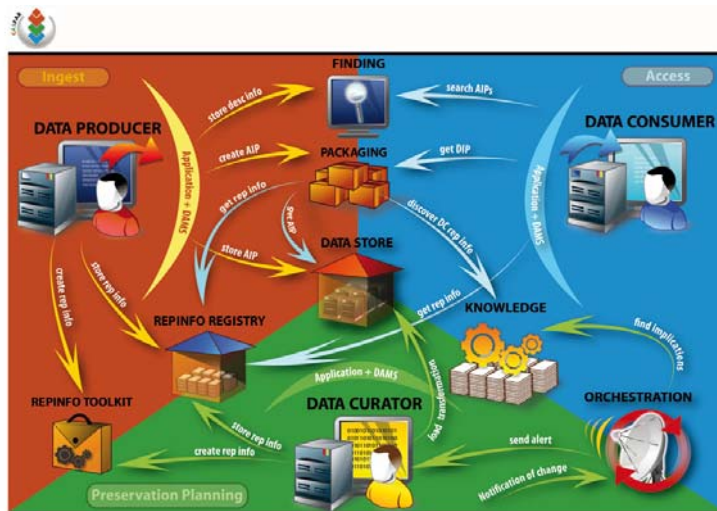
Estimated construction cost	30 M€
First open access foreseen	2008

Council of European Social Science Data Archives:

- Distributed infrastructure providing and facilitating access of researchers to high quality data and supporting their use
- Currently extends across 21 countries in Europe
- Holds some 15,000 data collections
- Provides access to over 20,000 researchers
- Agreements in place with other organisations worldwide

www.nsd.uib.no/cessda

Standardisation & Technical Issues



A European Digital Information Infrastructure to take care of these issues

1. Identify *core physical digital archives/repositories* in several initial communities and among cross-community organisations. Do this for documents and for data
2. These must *OAIS-compliant* to ensure proper archiving, interoperability and long-term preservation
3. Framework for *metadata*, Framework for *persistent identifiers*, and number of *registries*, possibly *other standards*
4. Cost-effective *preservation methods and services* must be available
5. Common framework of *principles and guidelines for management of access and rights* (underlying the technical tools to implement this framework)
6. Create *Financial mechanism* for developing and testing implementation tools, techniques and services, and for strengthening collaboration and training
7.
 - a. *Certification service* providers, accredited according to
 - b. Common *European accreditation mechanism*.

Finance

Innovative Medicine Initiative

Canon Foundation Research Fellowships

7th Framework Programme

FEBS Scholarship

European Investment Bank

EuroCores Life+

European Article 169 Gate2Growth **COST** European Cohesion Funds

Structural EMBO Young Investigator Program **DFG** ESA Research Grant Marie Curie Fellowships

Fund **ESFRI** **BBSC** SPINOZA Award **Global Health Program**

European Investment Fund EraNet **Eureka** **NATO Science Program**

Competitive Innovation Program **EFSA** EuroBioFund

The Welcome Trust **VW-Stiftung** European Regional Development Fund
Risk Sharing Finance Facility

The paradox & problem

- ▶ Lack of funding possibilities
 - ▶ There is no lack of funding possibilities, but there are a lot of good projects in vain seeking funding



The paradox & problem

- ▶ Unmet needs
 - ▶ Few projects are tailored to meet funding requirements, and funders are really looking for high impact opportunities



The paradox & problem

- ▶ Sources of funding
 - ▶ Differences between public and private funders, but also differences between various public funders



The paradox & problem

- ▶ Stakeholders value
 - ▶ Projects are becoming more and more complex, the days of single community approach have passed, stakeholders might have different views on the same project



Keeping the records of science accessible

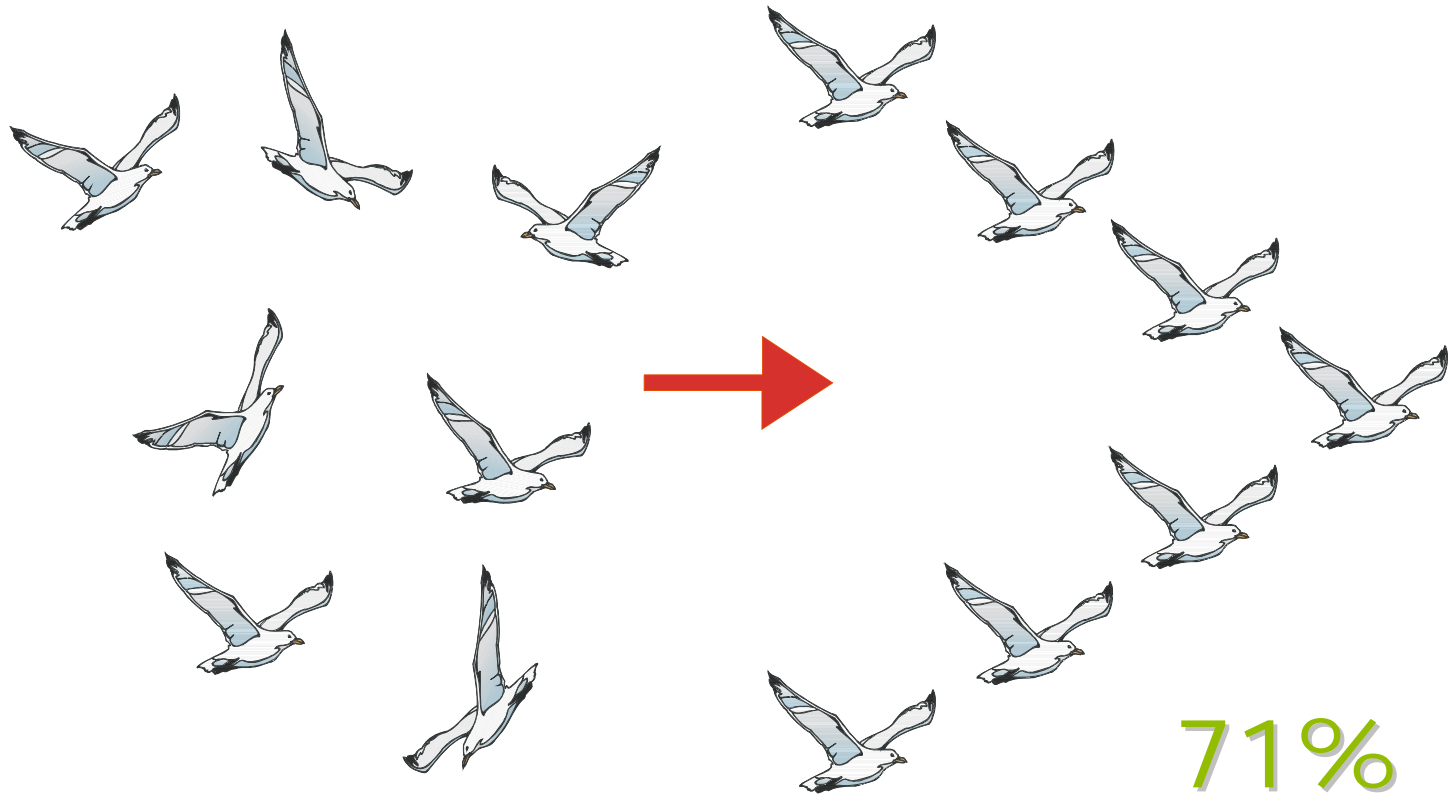
- Leadership
- Science breakthrough /Market opportunity
- Involvement of all relevant stakeholders
- Sense of urgency / Unmet need
- Support from the top



Alliance aims to:

- Establish wide consensus at strategic level among major stakeholders
- Promote creation of main building blocks of the infrastructure
- Work with European partners to ensure that technology, skills, and standards are in place
- Work with public and private funders
- Be key enabling mechanism for national governments and the EU
- Offers platform for effective coordination and sharing information
- Strengthen role of European parties in world-wide efforts

The European Approach



Annual Conference 24 November 2009



- Keeping the records of Science Accessible: towards a sustainable scientific data e-infrastructure!
- Den Haag, The Netherlands

Thank you

For more information, please contact:

Alliance for Permanent Access

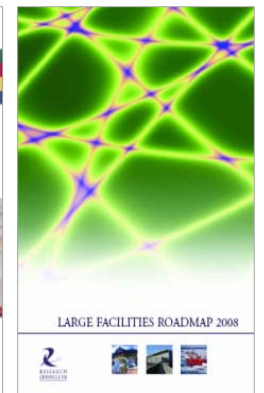
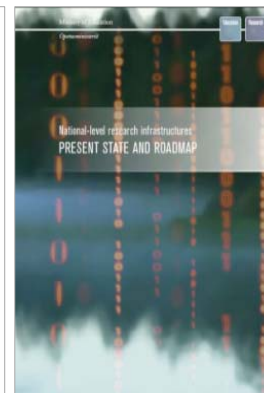
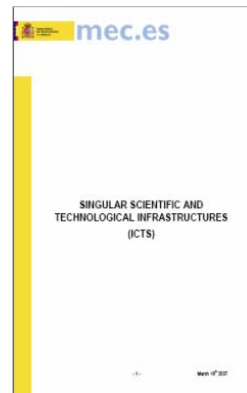
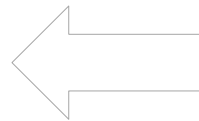
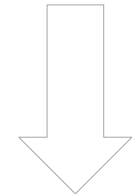
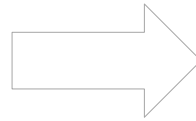
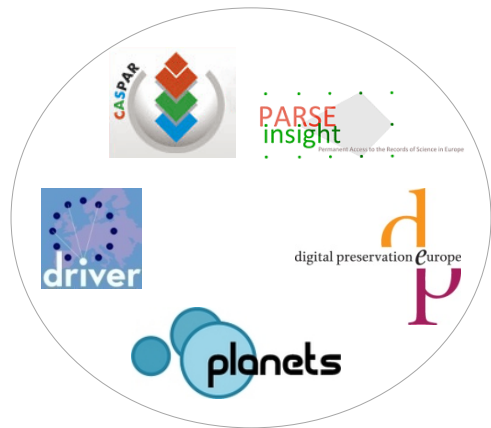
PB 90407

2509 LK The Hague

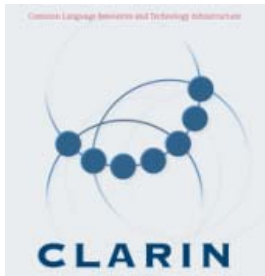
The Netherlands

t: +31 70 3140367

www.alliancepermanentaccess.eu



European landscape



Primary data vs. documents

Primary data	Documents
No copyright, but data policy limitations for commercial use	Copyright
Highly non-standardised	Much more standardised
Need special Representation Information (structure, semantics, software) to be understood and processed, but machine-readable	Human readable when displayed, but not machine-understandable
Digital volume huge	Digital volume modest or even small
Very dispersed at present	In journals or archives with publishers, libraries
Business models that include storage and preservation almost absent	Several models for storage, access and preservation