



La red de repositorios institucionales del Consorcio Madroño ante el “diluvio de los DATA”

Juan Corrales Correyero
Alicia López Medina



Madroño

www.consorciomadrono.es



ACERCA DE

NOTICIAS Y EVENTOS

DOCUMENTOS

RECURSOS ELECTRÓNICOS

ACTIVIDADES



INTRANET



MAPA WEB



AYUDA

Consortio de Universidades de la Comunidad de Madrid y de la UNED para la cooperación bibliotecaria.



UNIVERSIDAD DE ALCALÁ



UNIVERSIDAD COMPLUTENSE MADRID



UNIVERSIDAD AUTÓNOMA DE MADRID



UNIVERSIDAD POLITÉCNICA DE MADRID



UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA



UNIVERSIDAD REY JUAN CARLOS



UNIVERSIDAD CARLOS III DE MADRID

Plataforma digital de acceso abierto a la producción científica generada en la Comunidad de Madrid

- La necesidad de una estrategia para la gestión, organización y preservación de la creciente producción científica y académica en formato digital de nuestras Instituciones.
- La crisis del modelo de comunicación científica
- Las nuevas formas de investigar en el entorno digital: en red, en colaboración, con recursos compartidos (computacionales, almacenamiento, software, instrumentos y datos)

Acceso abierto para las
personas

“Open Acces”

“Open Data”

Acceso abierto para las
máquinas

**Interoperabilidad
(protocolos y
estándares abiertos)**

“e-ciencia” I: 2006-2008 (fase “Open Access”)

- La necesidad de una estrategia para la gestión, organización y preservación de la creciente producción científica y académica en formato digital de nuestras Instituciones: **publicaciones científicas**
- Nuevos modelos de comunicación científica: **OA, OAI**
- Las nuevas formas de investigar en el entorno digital: en red, en colaboración, con recursos compartidos (computacionales, almacenamiento, software, instrumentos y datos)

“e-ciencia” I: 2006-2008

- **Red de repositorios institucionales abiertos:** las universidades del Consorcio Madroño y el CSIC. La red está abierta a la participación de cualquier organización de investigación.
55.340 documentos
- Además, hemos creado un repositorio independiente para albergar los contenidos de instituciones que no dispongan de su propio repositorio y para investigadores individuales.
- **Contenidos:** artículos, tesis, conferencias, documentos de trabajo, capítulos de libros, libros, vídeos, imágenes
 - texto completo
 - 95% acceso abierto (algún embargo)
 - científicos
 - revisados/no revisados
- **Estándares:** todos los repositorios son conformes con el protocolo OAI-PMH. Dublin Core. Estamos implementando las directrices de DRIVER.
- **Software:** e-prints, D-space, Fedora, Digitool
- **Servicios:** portal de búsqueda “e-ciencia”, RSS, búsqueda texto completo, estadísticas, asesoría en cuestiones relacionadas con los derechos de autor.
- **Políticas:** Firma de la Declaración de Berlín (todas las Instituciones participantes), recomendación de la Comunidad de Madrid (CSIC, uc3m, urjc). En la nueva convocatoria de la Comunidad de Madrid el autoarchivo en el repositorio institucional será obligatorio (mandato)
- **Proyectos** nacionales e internacionales: RECOLECTA, DRIVER, ERMACS, Proyecto NEE0 (uc3m)

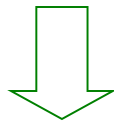
“e-ciencia” II: 2009-2011

•La necesidad de una estrategia para la gestión, organización y preservación de la creciente producción científica y académica en formato digital de nuestras

Instituciones: **DATA**

•La crisis del modelo de comunicación científica: **ORE**

•Las nuevas formas de investigar en el entorno digital: en red, en colaboración, con recursos compartidos (computacionales, almacenamiento, software, instrumentos y datos)



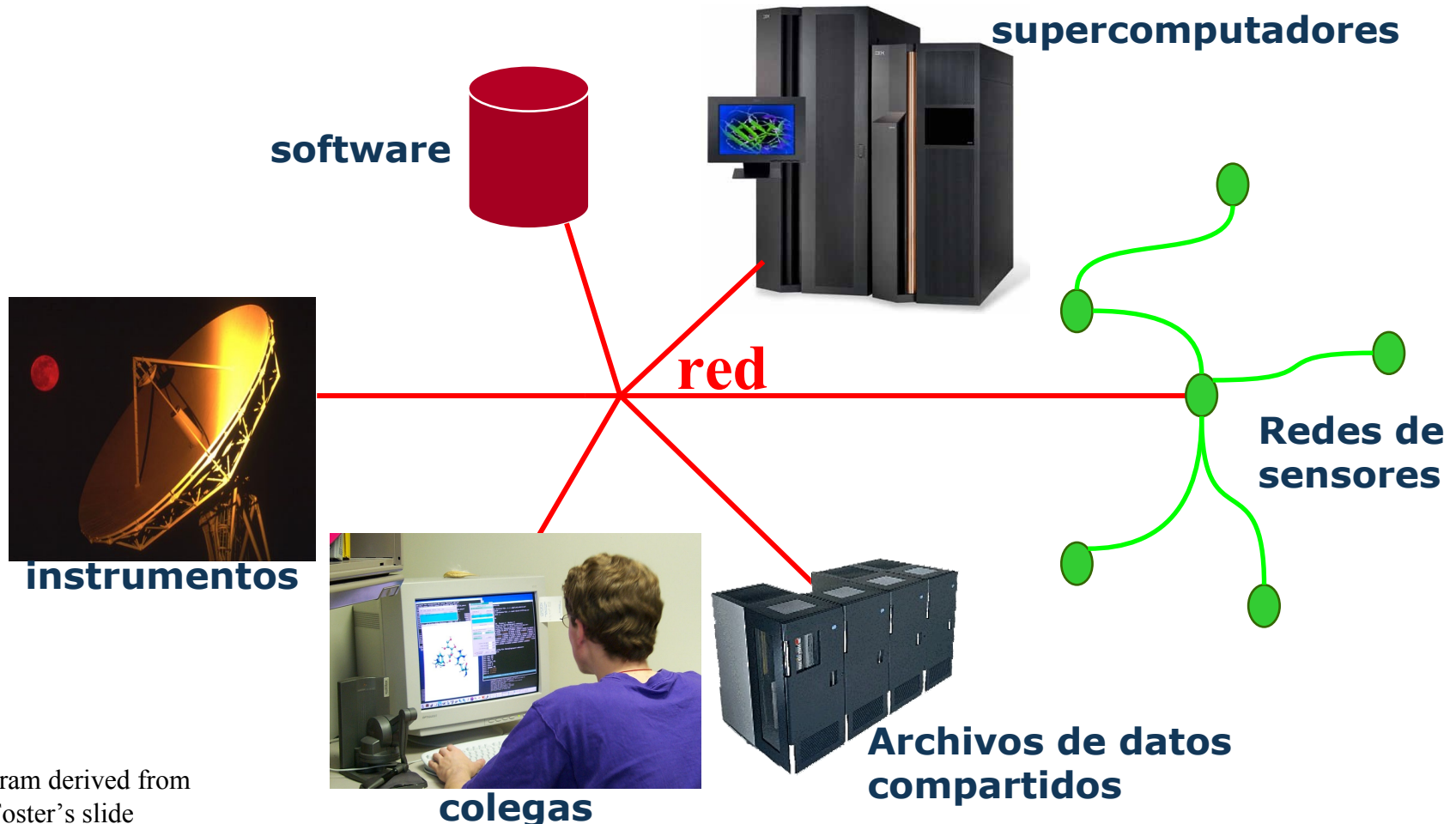
“e-ciencia”

“DATA”

“e-infraestructura”

E- ciencia, e-investigación: Actividad científica mediante el uso colaborativo de recursos (computacionales, de almacenamiento, software, datos) geográficamente distribuidos e interconectados mediante Internet, en Ciencias, Humanidades y Ciencias Sociales.

▪ **de la ciencia empírica, experimental, teórica y computacional... al uso intensivo de DATA... abstracciones... modelos... simulación... e-ciencia...**



El “diluvio de los DATA”

“Raw data is a collection of numbers, characters, images or other outputs from devices to convert physical quantities into symbols, in a very broad sense. Such data is typically further processed by a human or input into a computer, stored and processed there, or transmitted (output) to another human or computer. Raw data is a relative term; data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next”

Wikipedia

El volumen de DATA generados en la “e-ciencia” por los instrumentos científicos superará muy pronto todos los datos científicos generados hasta ahora en toda la historia de la investigación (y lo hará cada día)

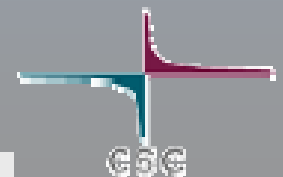
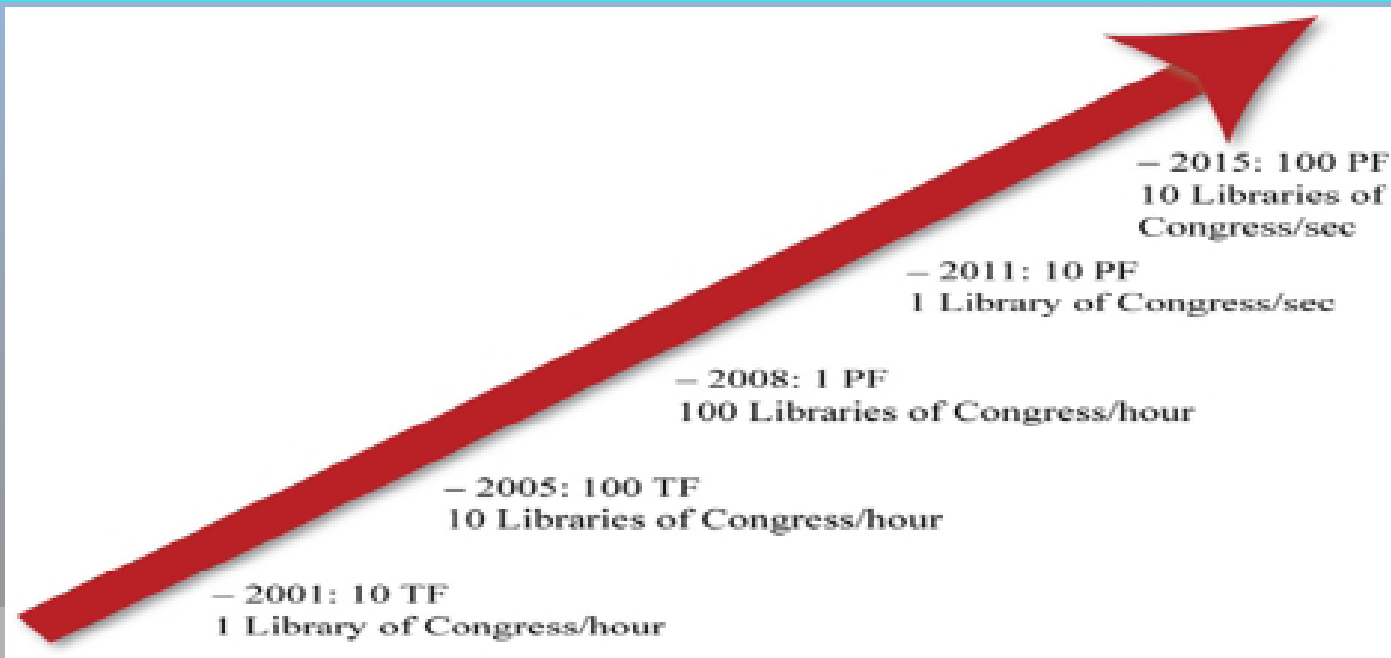
Diferentes en volumen y características, este diluvio de DATA tiene profundas implicaciones en la forma en que se financia y se lleva a cabo la ciencia y afecta a todas las disciplinas científicas, incluidas las Humanidades.

El reto que plantean es complejo y urgente: cómo desarrollar nuevos métodos, estructuras de gestión y tecnologías para gestionar la diversidad, el tamaño y la complejidad de los data actuales y futuros.

Data and information explosion

Petascale computing produces exascale data

1 Gigabyte (1GB) = 1000MB CD album	1 Terabyte (1TB) = 1000GB Word yearly Book production	1 Petabyte (1PB) = 1000TB One LHC- experiment yearly data production	1 Exabyte (1EB) = 1000 PB World yearly information production
--	--	---	---



El acelerador de partículas del CERN



- **GRID:** tecnología que permite compartir una serie de recursos en la red de manera uniforme, segura, transparente, eficiente y fiable, ofreciendo un único punto de acceso a un conjunto de recursos distribuidos geográficamente en diferentes dominios de administración.

La enorme cantidad de datos que recogería anualmente el LHC no iba a poder almacenarse de forma local. Para almacenar los datos generados en un solo año serían necesarios 56 millones de CDs, los cuales en caso de apilarse los unos encima de los otros, alcanzarían una altura de 64 kilómetros.



Desde hace semanas, apenas ocurre nada en Doñana que no sea registrado por uno de los **cientos de sensores instalados a lo largo de miles de hectáreas de terreno.**

Un sistema de transmisión de datos sin cables y un **potente servidor informático** permiten que esas observaciones se vuelquen en internet cada segundo. Desde cualquier lugar del mundo, un investigador o un ciudadano curioso pueden saber lo que está pasando en Doñana

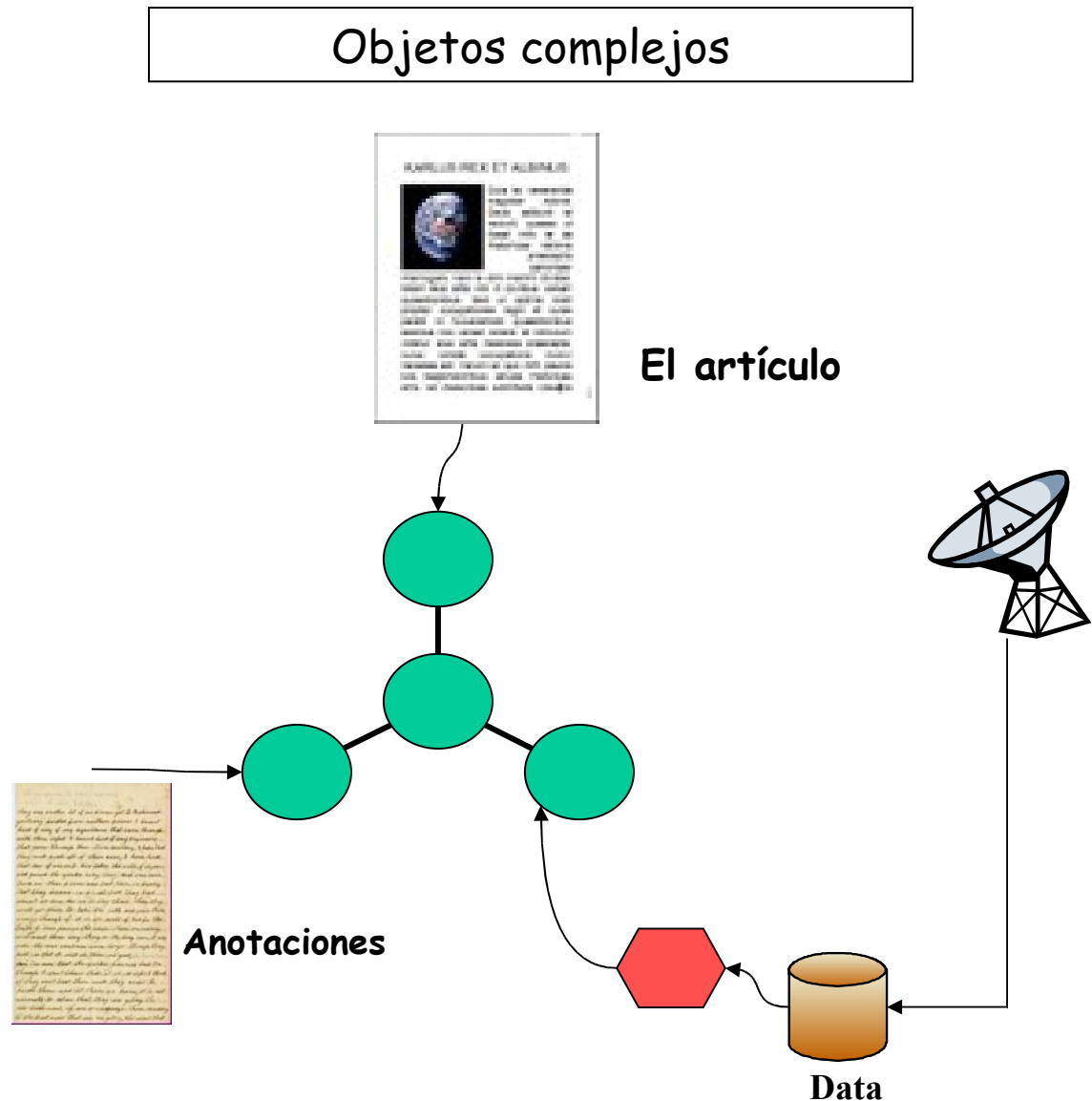
Al mismo tiempo, desde el interior de la marisma, un investigador puede conectarse a internet, bajarse un mapa del terreno, una información que necesite o transmitir los datos que esté anotando en ese momento desde su agenda electrónica. El número de cosas que pueden registrar los sensores es difícil de imaginar. **Son más de 200 parámetros del medio físico y del biológico.**

Se puede conocer la temperatura, la salinidad y la composición química del agua de la marisma y el volumen de CO_2 de la atmósfera, pero también se puede vigilar los nidos mediante alguna de las 16 cámaras que transmiten por internet sus grabaciones, saber cuánta savia circula por un árbol, medir cómo crece el tallo de un arbusto, cuánto ozono troposférico hay...

La unidad de información en la “e-ciencia”

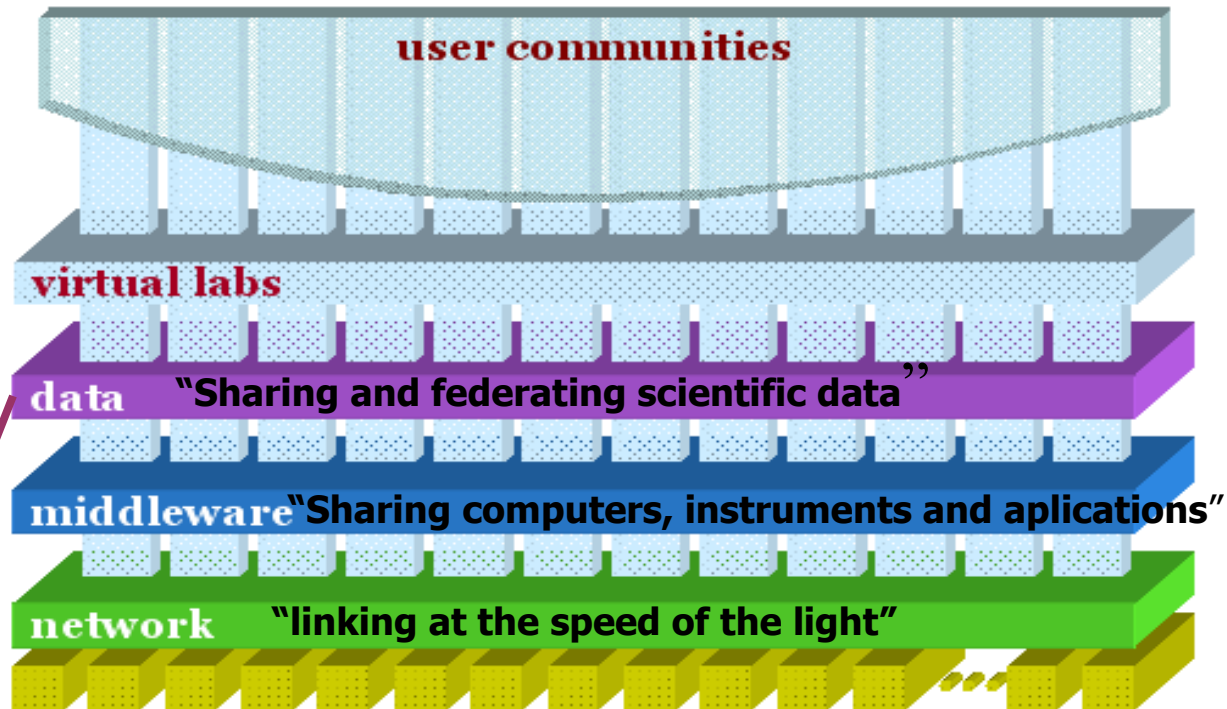
- Documentos
- Texto
- Data
- Objetos de aprendizaje
- Simulaciones
- Imágenes
- Video
- Software
- Análisis automáticos
- Algoritmos matemáticos

Agregaciones



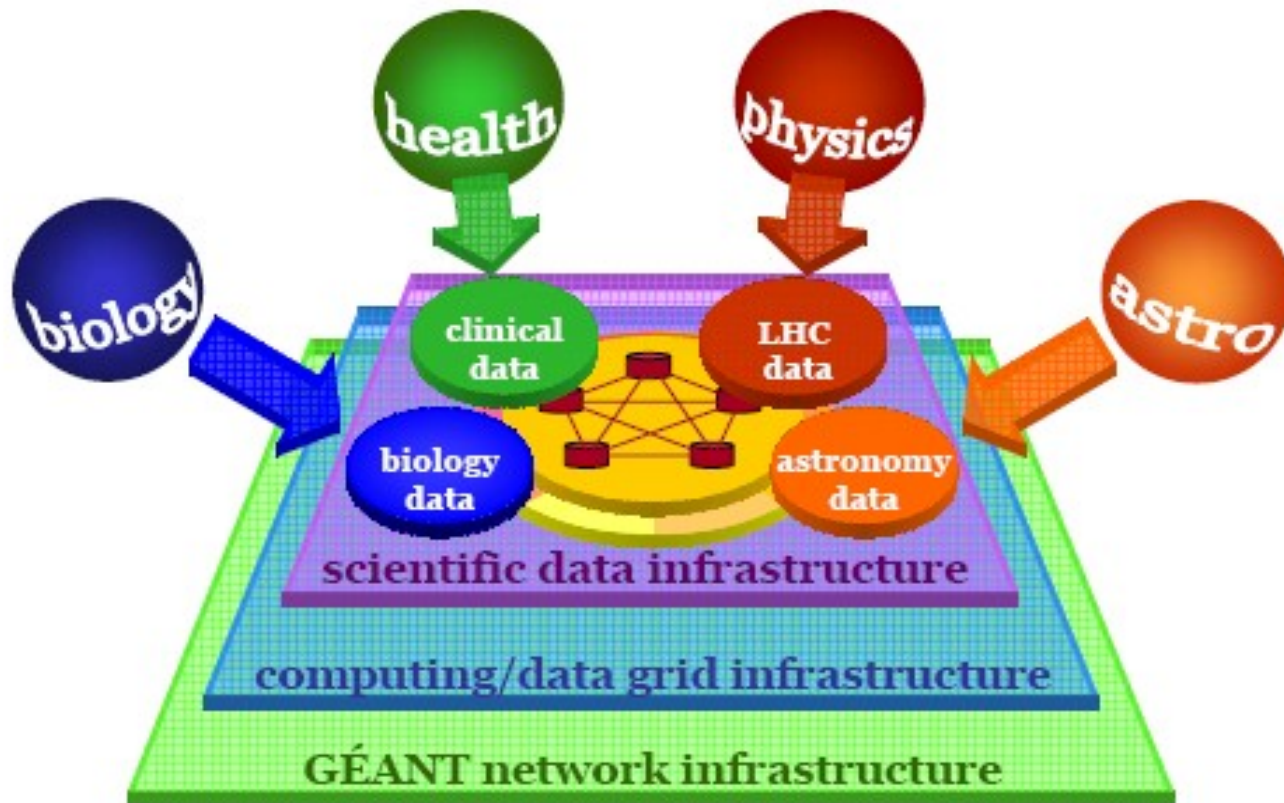


e-Infrastructure - implementation



"data": cualquier información que pueda ser almacenada en formato digital y accedida electrónicamente, incluyendo **DATA**, texto, video, audio, software, modelos y simulaciones, imágenes, etc.

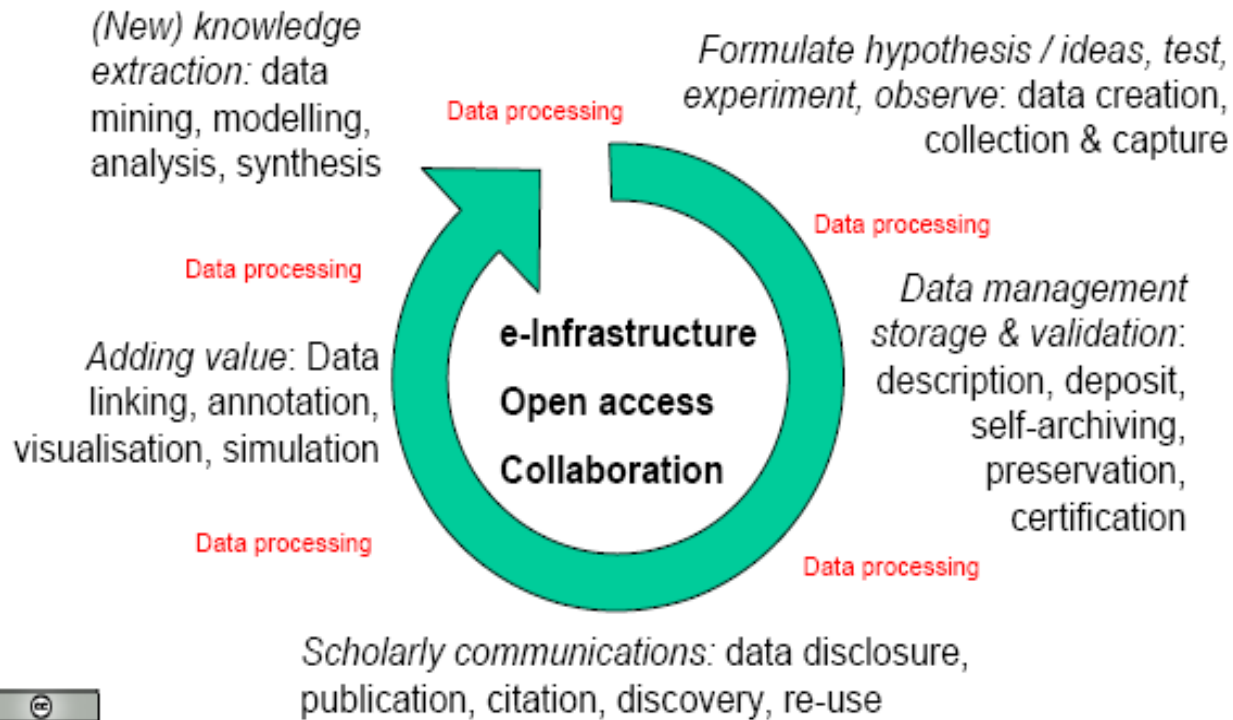
scientific data as an infrastructure



Policy requirements identified by e-IRG and ESFRI:

(availability, preservation/curation, quality, rights of use, interoperability)

(e)-Research Life Cycle view of Data Curation?



This work is licensed under a [Creative Commons License](https://creativecommons.org/licenses/by-sa/2.0/)
Attribution-ShareAlike 2.0

Figure 7 e-Research Life Cycle and data curation, Liz Lyon.

INFRA-2008-1.2.2

Scientific Data Infrastructure

Carlos Morais Pires

European Commission
DG INFSO
GÉANT & e-Infrastructure Unit

Fourth call: deadline on 11 September 2008



INFRA-2008-1.2.2

Scientific Data Infrastructure

simple !

This activity supports the deployment of a broad European multidisciplinary scientific data infrastructure able to be easily federated with other knowledge infrastructures in other parts of the world, building upon the achievements of network and grid infrastructures and opening its benefits to other potential research areas such as e-health, e-learning and others.

- This activity addresses the rapidly increasing use of digital content in research and in the generation and dissemination of scientific and technical knowledge. The increasing availability of primary sources of data in digital form (e.g. experimental raw data, social sciences data) has the potential to shift the balance away from research based on secondary sources (such as publications), thus positioning data as the central element in the scientific process.
- This activity should provide an integrated set of services exploiting the middleware and grid capabilities to federate data in an eco-system of digital resources. These services should enhance the ability of researchers to extract further meaning from masses of data stored in institutional, national or community repositories, by supporting the deployment of standardised mechanisms to store, archive, authenticate, access, transfer, preserve, curate, certify and interpret scientific data.
- Furthermore, the deployed scientific data infrastructure will require adaptation in cultures and new approaches and competences, given the intrinsic relation between data and associated software to read, interpret and process it.





Implementation (2)

INFRA-2009-1.2.3: Scientific Information Repository supporting the European FP7 Research Programme

- Call for proposals to be published on 9 Dec. 2008; deadline 17 March 2009, budget: € 4 million
- Seeks projects to support the establishment and operation of a technical infrastructure of digital repositories to deposit and access scientific articles and data produced under FP7 (incl. ERC)
- Infrastructure to build on existing work on repositories
- Expected impact: Enhancement of access to scientific information produced in the context of FP7; linking of many different repositories as a service and e-infrastructure

Further information: <http://cordis.europa.eu/fp7/ict/e-infrastructure/>

Relevant Work Programme: Research infrastructures 2009

ftp://ftp.cordis.europa.eu/pub/fp7/docs/wp/capacities/infrastructures/n_wp_200901_en.pdf



E-IRG Data Management Task Force



RATIONALE

- Increasing importance of scientific data to support research activities
- Exploding volume of data coming out from many different scientific domains
- Difficulty to access or navigate data
- Global lack of interoperability in data structure rather than hardware infrastructures



008

2

OBJECTIVES

- Address a few basic relevant issues:
 - Volume of data
 - Sustainability and curation
 - Data structure and addressing
 - Standard for meta-data
 - Interoperability

Australian National Data Service

The Australian National Data Service (ANDS) aims to:

- influence national policy in the area of data management in the Australian research community
- inform best practice for the curation of data
- transform the disparate collections of research data around Australia into a cohesive collection of research resources

ANDS is funded by the Australian Commonwealth Government's [Department of Innovation, Industry, Science and Research](#) (DIISR). The funding has been provided through the [National Collaborative Research Infrastructure Strategy](#) (NCRIS) as part of the [Platforms for Collaboration](#) Investment Plan.

NCRIS is providing \$542 million from 2004-05 to 2010-2011 to develop and fund national research infrastructure projects. \$24 million of this will fund the Australian National Data Service.

Government of Canada / Gouvernement du Canada

Canada

Research Data Canada
data-donnees.gc.ca

Franglais Home Contact Us Help Search canada.gc.ca

Home >

About Research Data Canada
Backgrounder
Key Achievements
The Working Group
Members
Members' corner
Background Resources
Documents, Presentations, and Links
National Consultation on Access to Scientific Research Data (NCASRD)
Proactive Disclosure

Research Data Canada

"... urgent need for action to propel Canada into a new and transformational data-intensive paradigm for Canadian research."

David Strong (National Consultation on Access to Scientific Research Data, Final Report)

Research Data Strategy Working Group: Opening new pathways to Canadian research data

News
[Working Group news](#)
[Suggest a news item](#)

Data Resources
[Scientific Data Sets](#)
[Social Science Data Sets](#)
[Tools and Standards for Data Discovery, Analysis and Management](#)
[Make A Suggestion](#)

Task Group Activities
[Task Group Activities](#)

Office of Cyberinfrastructure (OCI)

[OCI Home](#)

[About OCI](#)

[Funding Opportunities](#)

[Awards](#)

[News](#)

[Events](#)

Sustainable Digital Data Preservation and Access Network Partners (DataNet)

CONTACTS

Name	Email	Phone	Room
Lucy Nowell	lnowell@nsf.gov	(703) 292-8970	
Sylvia Spengler	sspenqle@nsf.gov	(703) 292-8930	

PROGRAM GUIDELINES

Solicitation [07-601](#)

DUE DATES

(\$100 millones/10 años)

Full Proposal Target Date : May 15, 2009

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

“Access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators”

La Red

Participantes y
Colaboradores

Eventos

Iniciativas e-Ciencia

Resultados

Red Española de e-Ciencia

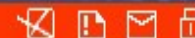
Grid y Supercomputación al Servicio de la Comunidad Científica Española

Áreas de Aplicación

GRID

Supercomputación

Middleware


[Aviso Legal](#)


La Red Nacional de e-Ciencia persigue coordinar e impulsar el desarrollo de la actividad científica en España mediante el uso colaborativo de recursos geográficamente distribuidos e interconectados mediante Internet. En la red participan usuarios y expertos en aplicaciones de diversas disciplinas científicas (biocomputación, imagen médica, química computacional, fusión, meteorología, etc.), investigadores en el ámbito de las TIC y centros proveedores de recursos, quedando así representados todos los actores de la e-Ciencia. La necesidad de la e-Ciencia se fundamenta en la creciente exigencia por parte de los científicos de más recursos de procesamiento y almacenamiento de datos, así como de nuevas formas de trabajo colaborativo que conduzcan a la sociedad del conocimiento. El desarrollo de la e-Ciencia permitirá nuevos modelos de aplicaciones y desplegar middlewares que permitan explotar eficientemente los recursos de la comunidad

Acción Financiada por:



Entidad Coordinadora:



Coordinador Científico:

e-ciencia
wiki
accede

PANGAEA

<http://www.awi.de/index.php?id=3086&L=0>

“e-ciencia” II: 2009-2011

- Frente al reto de la gestión de los DATA generados por los investigadores de nuestras Instituciones
- Frente al reto de la integración de nuestros repositorios en la e-infraestructura de DATA
- Frente al reto de nuestro reciclaje para responder al reto: ¿“data librarians”?

Muchas gracias por vuestra atención

alopezm@pas.uned.es

webmaster@consorciomadrono.net