

LA GESTION DE DATOS CIENTIFICOS EN REPOSITORIOS DIGITALES

SEMINARIO SOBRE DATASETS - Consorcio Madroño

17 Noviembre 2008

Luis Martinez Uribe

Luis.Martinez-Uribe@oerc.ox.ac.uk

Digital Repositories Research Coordinator



SEMINARIO SOBRE DATASETS



**UNIVERSITY OF
OXFORD**

Sumario

- Antecedentes
- ¿Qué son los datos científicos y qué es *data management/curation*?
- Beneficios y desafíos
- Repositorios digitales
- Actividades en la Universidad de Oxford



SEMINARIO SOBRE DATASETS



Antecedentes

- Nuevas formas y métodos de investigación gracias a la aparición de herramientas e infraestructuras
- *Data deluge (Hey & Threfethen 2003)*
- *“More digital data will be produce in the next 5 years than in whole human history” (Australian DEST)*
- *Data management* es un componente crucial de la e-Infraestructura



“From Data to Wisdom: Pathways to Successful Data Management for Australian Science ”



SEMINARIO SOBRE DATASETS





¿Qué son los datos científicos?

<http://www.flickr.com/photos/iscjorgegarcia/2359144636/>*

Tipología de datos

National Science Foundation

(2005) “Long Live Digital Data Collections”

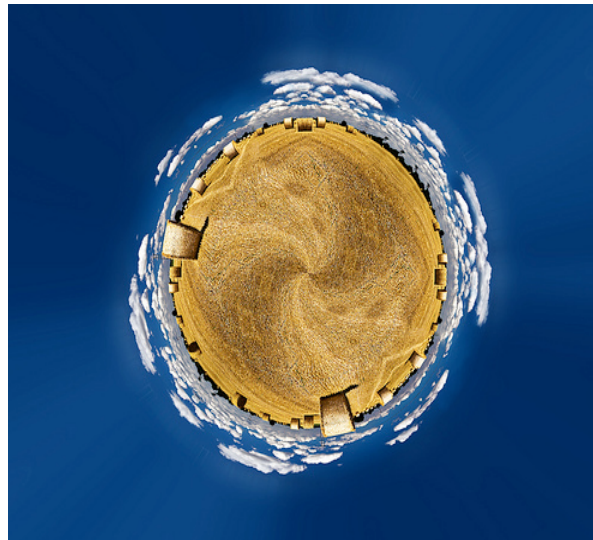


DATOS OBSERVACIONALES



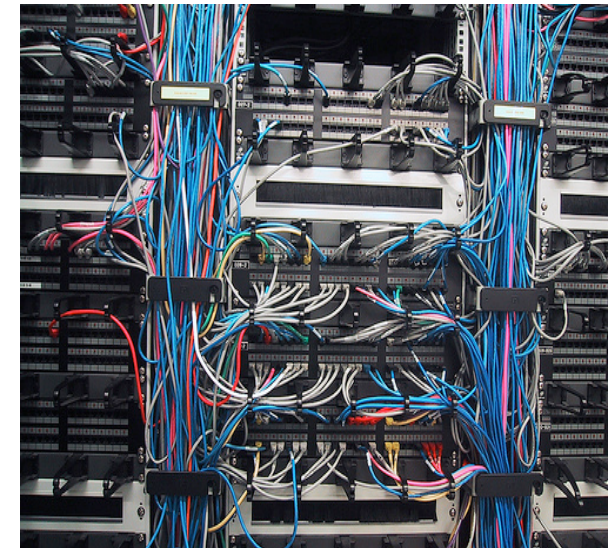
<http://www.flickr.com/photos/autowitch/5702193/>

DATOS EXPERIMENTALES



<http://www.flickr.com/photos/calixtojunior/2311207816/>

DATOS COMPUTACIONALES



<http://www.flickr.com/photos/nebneb/64058504/>



SEMINARIO SOBRE DATASETS



Tipología de datos

The Research Information Network (RIN)
(2007) “Stewardship of digital research data”

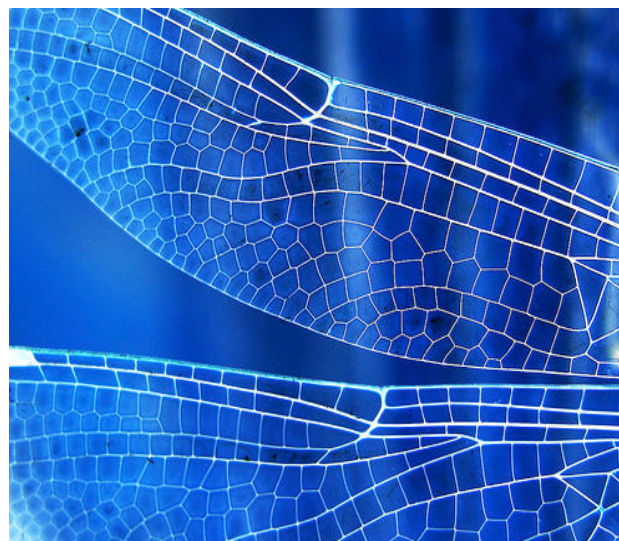


DATOS DERIVADOS



DATOS

- OBSERVACIONALES
- EXPERIMENTALES
- COMPUTACIONALES



<http://www.flickr.com/photos/wester/37660225/>

DATOS CANONICOS



<http://www.flickr.com/photos/wheatfields/2073336603/>



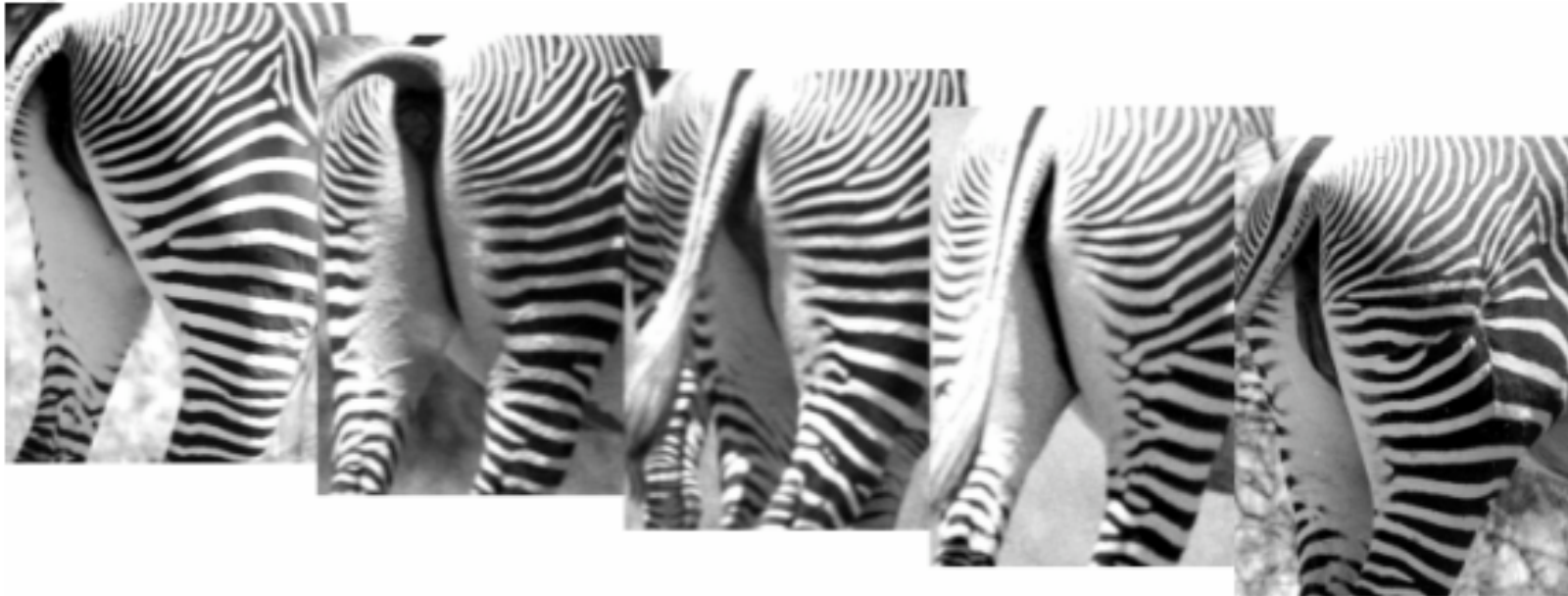
SEMINARIO SOBRE DATASETS





http://www.flickr.com/photos/piet_musterd/2231850447/

Ejemplos



From Dr David Shotton presentation

© Center for Whale Research / www.whaleresearch.com



Madroño

SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD



<http://www.flickr.com/photos/djmccrady/1883226927/>

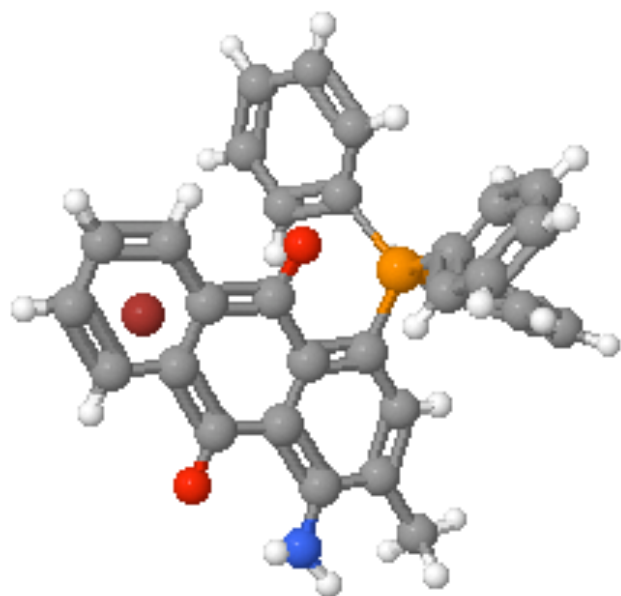


<http://www.flickr.com/photos/oliastro/2987657532/>



SEMINARIO SOBRE DATASETS





<http://ecrystals.chem.soton.ac.uk/604/>

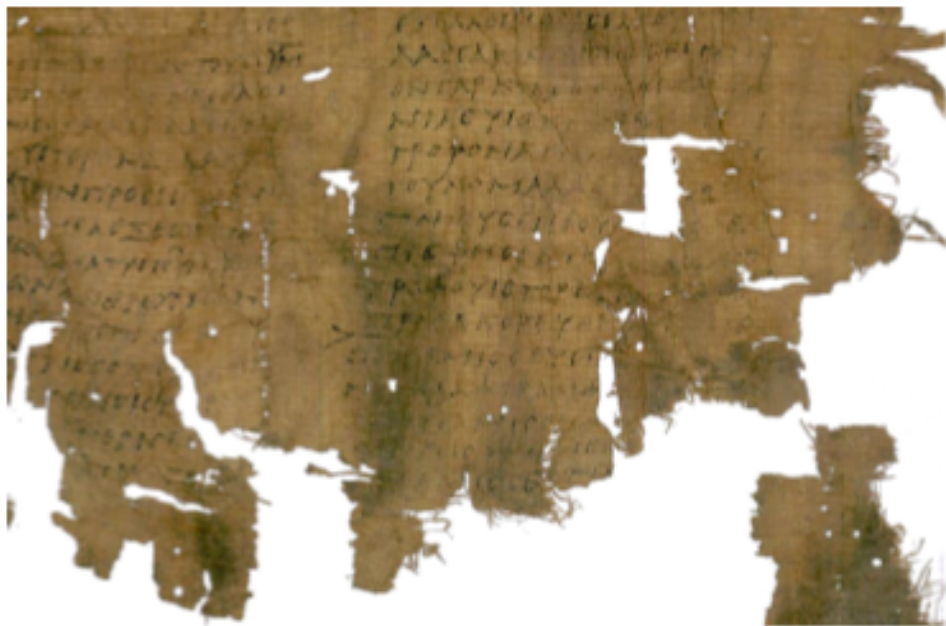


<http://www.rcsb.org/pdb/explore.do?structureId=1BA4>



SEMINARIO SOBRE DATASETS





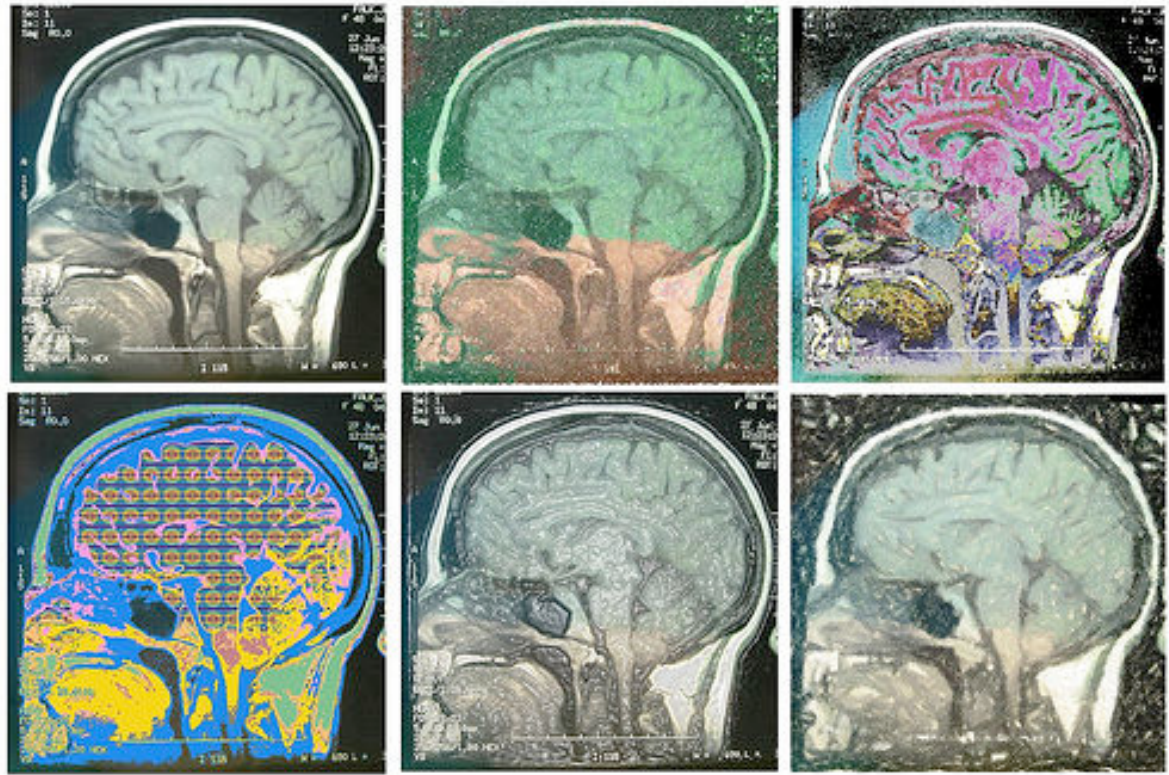
From Building a VRE for the Humanities poster
presented at All Hands Meeting 2007

<http://www.flickr.com/photos/wrowlands/2270729405/>

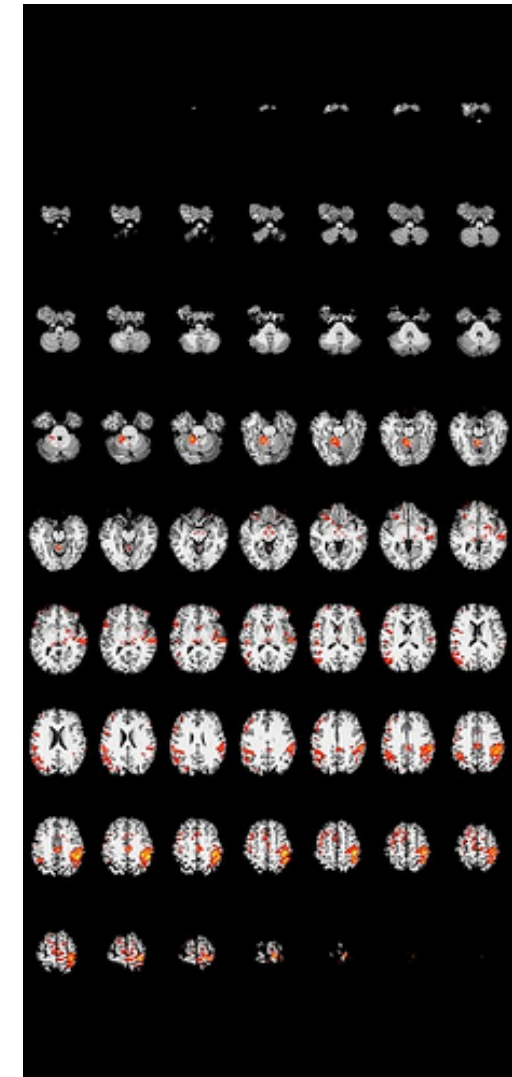


SEMINARIO SOBRE DATASETS





<http://www.flickr.com/photos/piper/22584430/>



http://www.flickr.com/photos/kiss_this/144493954/



SEMINARIO SOBRE DATASETS





<http://www.flickr.com/photos/edyson/59524184/>



<http://www.flickr.com/photos/alhouse/273160052/>



Madroño

SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD

<http://www.flickr.com/photos/thivierr/540241947/>



Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

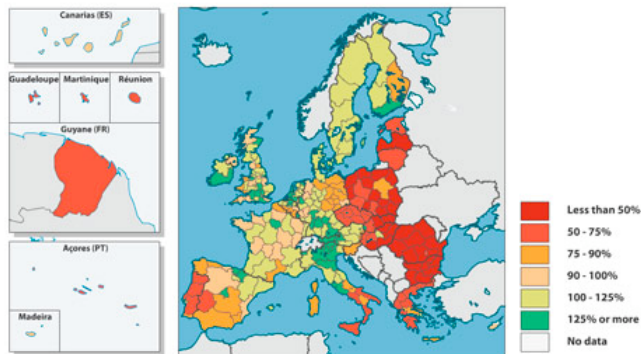
caseno intday intmon intyr inttime a1 a2 a3day a3mon a3yr

caseno	intday	intmon	intyr	inttime	a1	a2	a3day	a3mon	a3yr	
1	2068201	20	10	97	1	3	13	1	8	83
2	2087115	20	10	97	1	2	12	14	8	85
3	2001204	9	10	97	1	1	11	31	5	86
4	2001211	9	10	97	1	5	15	14	6	82
5	2001217	9	10	97	1	2	12	16	9	85
6	2001218	9	10	97	1	2	12	16	8	85
7	2001220	9	10	97	1	3	13	14	4	84
8	2001222	9	10	97	1	1	11	21	10	85
9	2001223	9	10	97	1	1	11	9	7	86
10	2001224	9	10	97	1	1	11	29	3	86
11	2001225	9	10	97	1	5	15	21	10	81
12	2002102	15	10	97	1	3	13	28	4	84
13	2002104	15	10	97	1	3	13	13	3	84
14	2002107	15	10	97	1	3	13	15	8	84
15	2002108	15	10	97	1	3	13	1	4	84
16	2002109	15	10	97	1	3	14	9	9	83
17	2002111	15	10	97	1	4	14	4	12	82
18	2002112	15	10	97	1	3	13	30	5	84
19	2002113	15	10	97	1	3	13	12	3	84
20	2002115	15	10	97	1	4	15	13	10	82
21	2002123	15	10	97	1	5	15	24	9	82
22	2002125	15	10	97	1	4	14	18	10	82

**EXPECTATIONS FOR THE NEXT TWELVE MONTHS:
CHANGE IN INDEX SPRING 2008 (EB69) – AUTUMN 2007 (EB68)**

Your life in general		The financial situation of your household		The employment situation in (OUR COUNTRY)		The economic situation in (OUR COUNTRY)	
EU27	-1	EU27	-9	EU27	-19	EU27	-28
CY	+22	CY	+10	MT	+13	CY	+12
CZ	+12	CZ	+9	CY	+5	CZ	+8
SE	+10	LT	+6	CZ	+2	MT	+8
DK	+7	NL	+4	PL	+2	:	:
NL	+6	:	:	:	:	EE	-24
MT	+5	FR	-11	FR	-28	SE	-24
SI	+5	UK	-13	PT	-28	PT	-26
SK	+5	BE	-15	DK	-29	NL	-27
:	:	PT	-21	LV	-30	DE	-29
LV	-5	EL	-22	EE	-32	IE	-32
IT	-8	ES	-23	IE	-35	DK	-34
ES	-16	:	:	EL	-38	LU	-34
PT	-20	SE	-43	ES	-43	UK	-35
				SE	-43	EL	-38
						ES	-39
						FR	-47

o GDP per inhabitant in PPS as a percentage of the EU-27 average (2004)



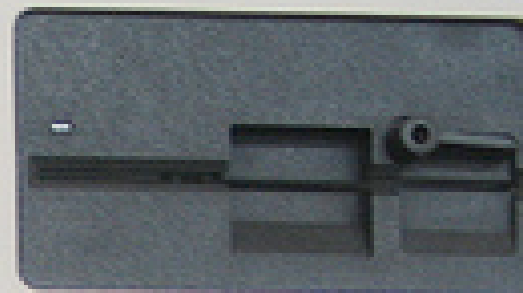
EUROBAROMETER 69 PUBLIC OPINION IN THE EUROPEAN UNION FIRST RESULTS



SEMINARIO SOBRE DATASETS



http://www.flickr.com/photos/hanan_cohen/455238557/



TANDY 1980-82 MODEL 100 MICROCOMPUTER SYSTEM



Data Management and Curation

Research Data Management

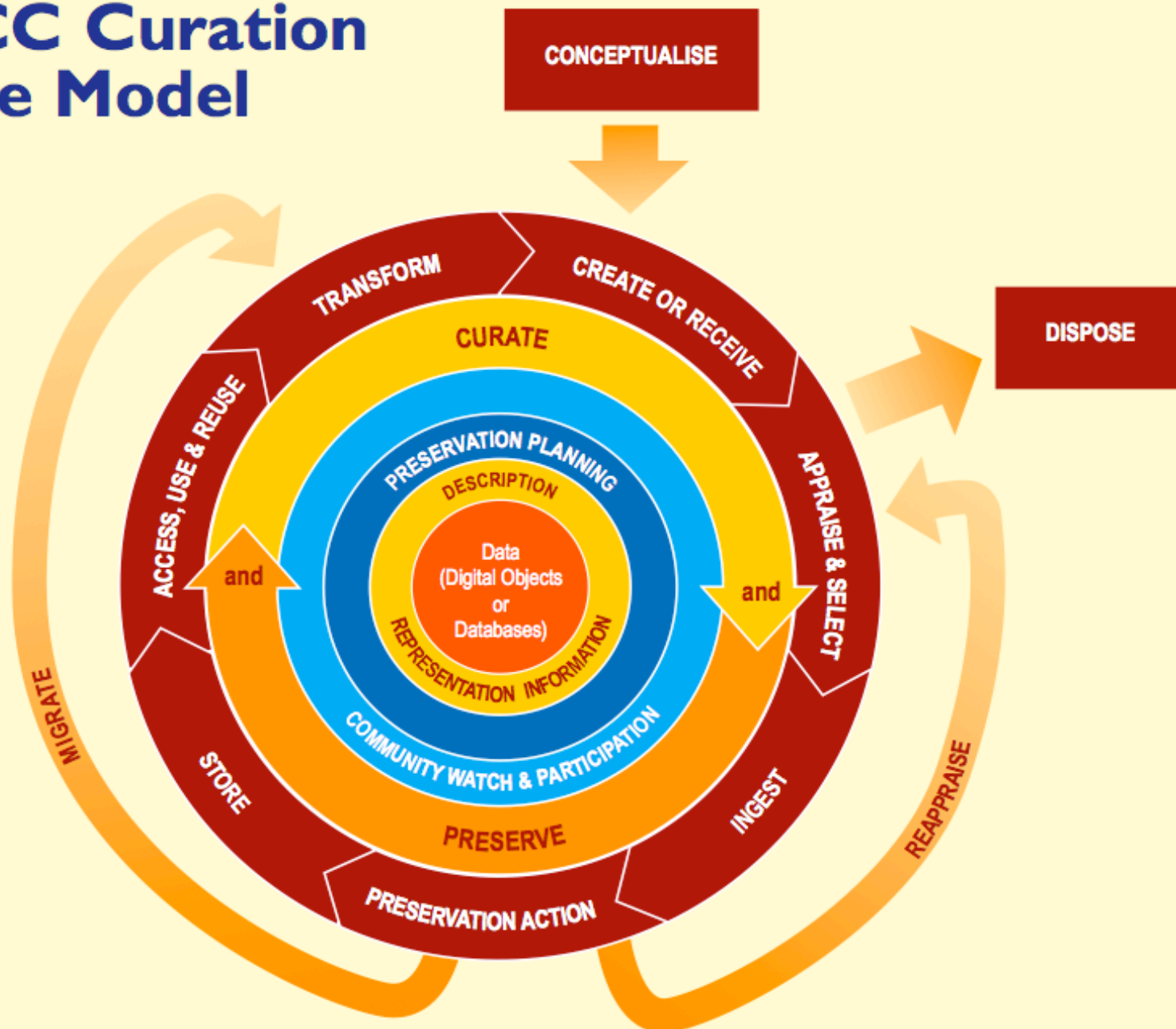
- De disciplinas como *knowledge/information management*
- “...es comprender las necesidades actuales y futuras”
(Departamento de Defensa de EEUU)
- Un medio para llegar a un fin
- No solo la infraestructura técnica, también políticas y procedimientos
- *Digital curation*



SEMINARIO SOBRE DATASETS



The DCC Curation Lifecycle Model



SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD

Beneficios

- Calidad y autenticidad
- No re-inventar la rueda, ¡crear datos puede ser caro!
- Mejora de acceso a datos (en muchos casos financiados con fondos públicos) mejora la calidad de la investigación
- Preservación
- Cumplir con los requisitos de las agencias de financiación



SEMINARIO SOBRE DATASETS





<http://www.flickr.com/photos/steve-wall/524803118/>

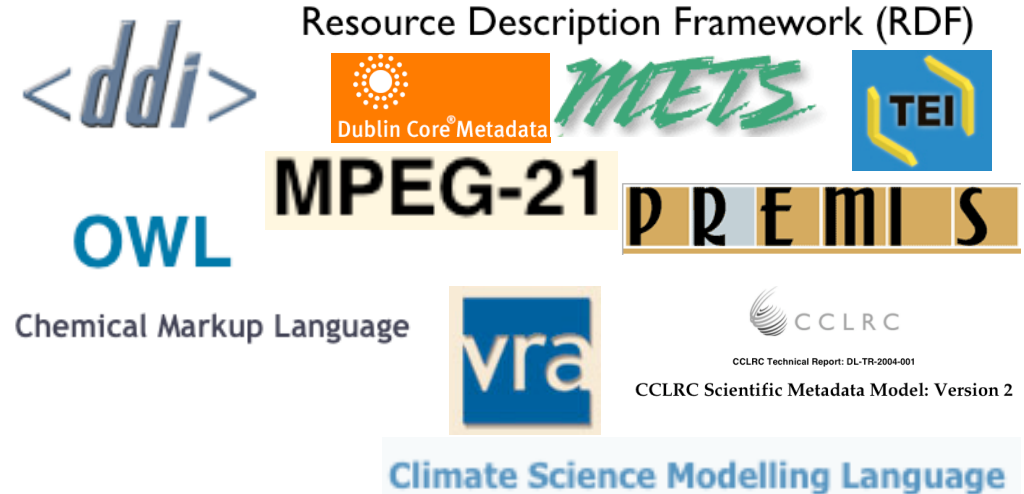
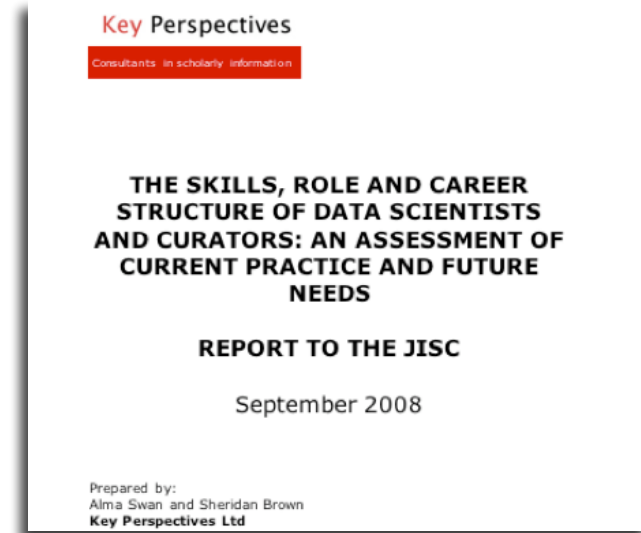
Desafíos

Desafíos

RIN (2007) “Stewardship of digital research data”

Roles y responsabilidades - ¿Quién es responsable de las distintas labores?

Estándares y garantía de calidad-
¿Qué estándares y procedimientos deben utilizarse? ¿Cómo podemos asegurar la calidad de los datos?
(Data peer-review)



SEMINARIO SOBRE DATASETS



Desafíos

Acceso, uso y atribución - ¿A quién le pertenecen los datos? ¿Cómo se pueden proteger los datos del mal uso? ¿Cómo se deben de citar los datos?



<http://www.flickr.com/photos/paulthewineguy/2495933369/>



<http://www.flickr.com/photos/wwworks/2960675738/>

Beneficios y rentabilidad - ¿Qué modelos son eficientes y rentables para gestionar y proveer acceso a datos investigativos?

Preservación y sustentación - ¿De qué manera se deben seleccionar los datos?



<http://www.flickr.com/photos/lollie-pop/64838763/in/set-910971/>



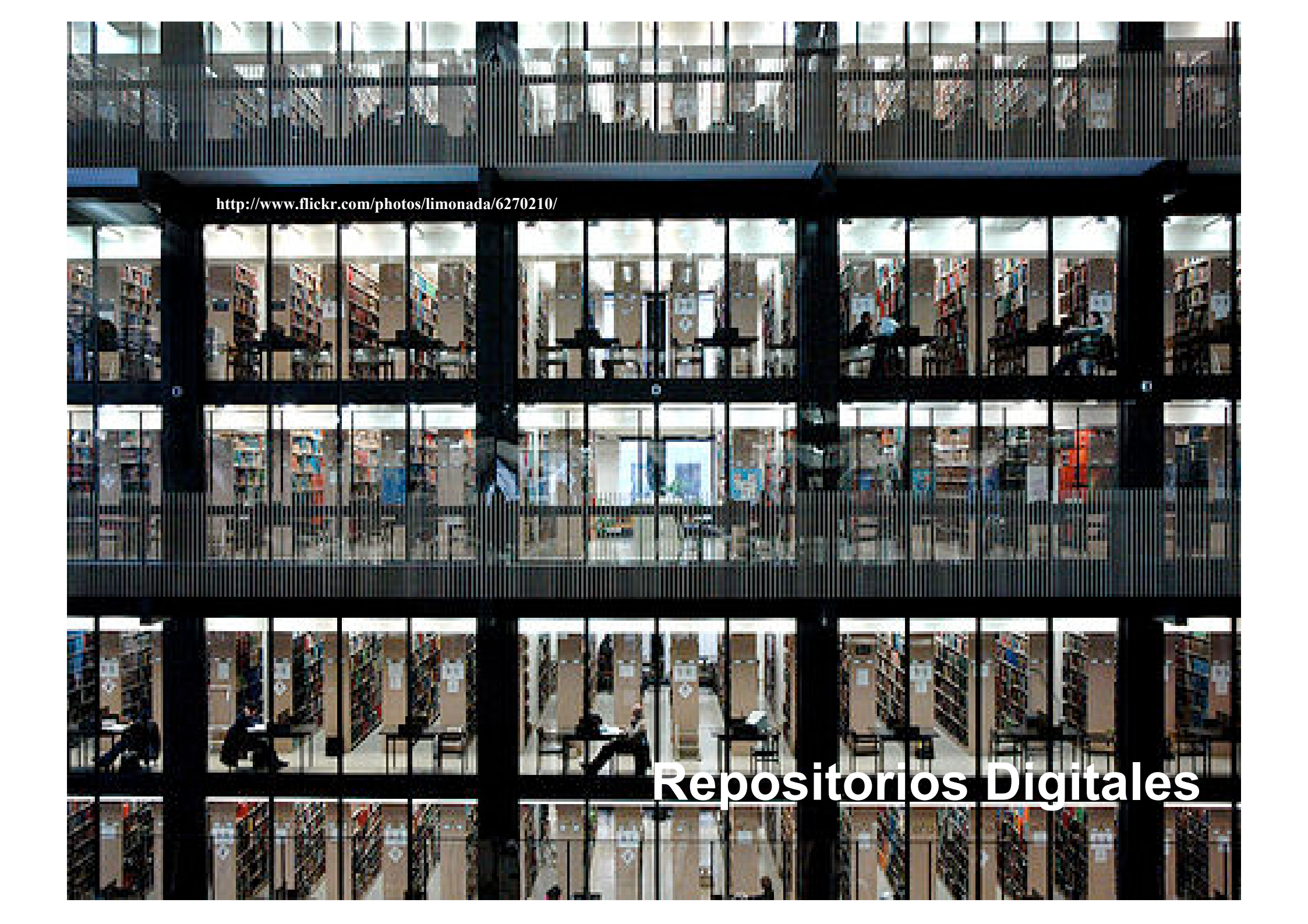
<http://www.flickr.com/photos/nickwheeleroz/2295584401/>



SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD



<http://www.flickr.com/photos/limonada/6270210/>

Repositorios Digitales

Repositorios Digitales

- El movimiento de Open Access
- Bibliotecas desarrollando repositorios para almacenar, proporcionar acceso y preservar los resultados de la investigación producidos en la institución (artículos y tesis en su mayoría)
- ¿Y los datos qué?



FedoraCommons



DSPACE

eprints



SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD

Repositorios de Datos

- Repositorios actuales validos para:
 - Almacenamiento/Acceso/Preservación/Descripción
- Los datos de investigación necesitan otras cosas:
 - Tamaños enormes en casos de GRID datasets
 - Control de Acceso
 - Visualizaciones
 - Análisis
 - Minería de datos
 - Anotación
 - Temas legales de creación, acceso y publicación
 - ?



SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD



La Universidad de Oxford

Scoping Digital Repository Services for Research Data Management



Project Aims

The Scoping Digital Repository Service for Research Data Management project overall aim is to scope the requirements, including underlying infrastructure and interoperability, for digital repositories services to store, disseminate and preserve research data generated at Oxford.



SEMINARIO SOBRE DATASETS



Problemas para gestionar sus datos

NO PUDE ENCONTRAR EL SENTIDO A LOS DATOS PRODUCIDOS DURANTE MI DOCTORADO HACE 5 AÑOS



QUIERO PUBLICAR MIS DATOS COMO UN RECURSO ADICIONAL A MI ARTICULO



CUANDO LOS INVESTIGADORES CAMBIAN DE TRABAJO PERDEMOS SUS DATOS

TUVIMOS QUE MIGRAR LOS FORMATOS DE TODOS NUESTROS DATOS Y TARDAMOS MESES

PARA COMPARTIR NUESTROS DATOS TENEMOS QUE TRANSPORTAR EL SERVIDOR

DATOS DE ESTUDIOS CLINICOS REALIZADOS HACE TREINTA AÑOS PERMITEN IDENTIFICAR A LAS HIJAS DE LAS MUJERES A LAS CUALES LES ADMINISTRO UNA DROGA QUE PUEDE SER LA CAUSA DE CANCER EN SUS HIJAS



CREAMOS UNOS DATOS EN UN PROYECTO INTERNACIONAL Y NO SABEMOS A QUIEN LE PERTENECEN



SEMINARIO SOBRE DATASETS



UNIVERSITY OF
OXFORD

Servicios Requeridos

- Soporte en cuestiones practicas de la gestión de datos a lo largo del ciclo de vida
- Una solución segura y fácil de usar que permita almacenar y compartir datos de gran tamaño utilizando sistemas de control de acceso
- Una infraestructura sostenible para publicar y preservar los datos



Gracias



SEMINARIO SOBRE DATASETS

