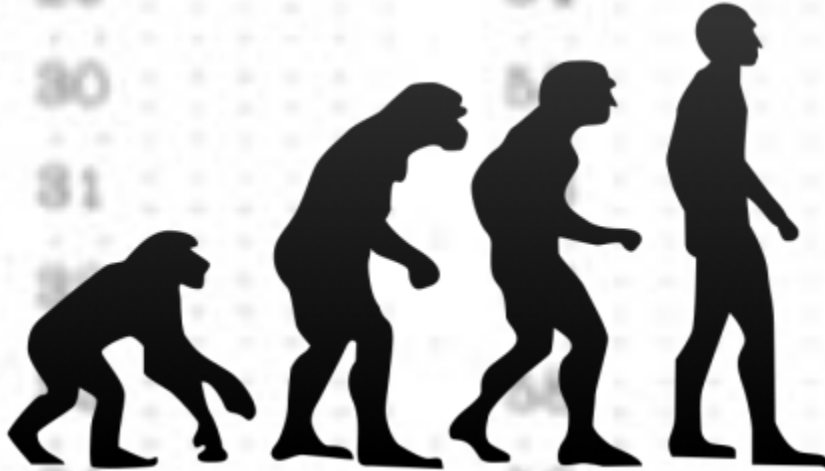


Una

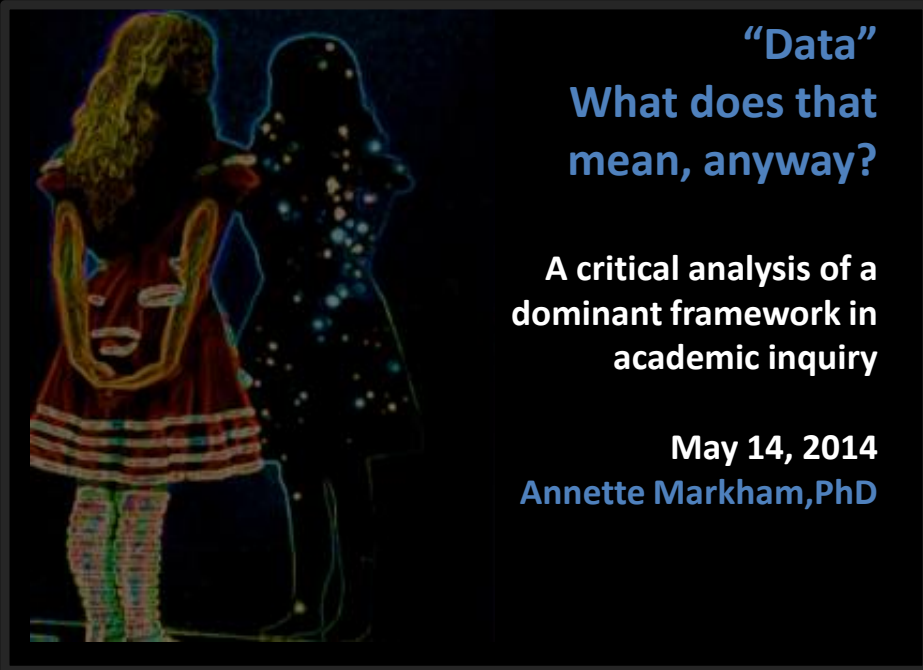
Historia de los servicios de datos



...desde el Data Librarian al comienzo del Research Data Management, pasando por el Data Curator, y llegando al Data Scientist.

Contenido





“Data”
What does that
mean, anyway?

A critical analysis of a
dominant framework in
academic inquiry

May 14, 2014
Annette Markham, PhD

“The term data comes *from the Latin verb dare, to give*. In light of this, ‘data’ is something that is already *given* in the argument.

“‘Data’ is often an easy term to toss around, as it stands for *unspecified stuff*. Ultimately, ‘data’ is “a lot of highly specific but unspecified stuff”, that we use to make sense of the world around us. The term ‘data’ is a arguably quite a powerfully rhetorical word in humanities and social sciences, in that it *shapes what we see and what we think*.”



TIPOLOGÍAS DE ROLES DE DATOS

Key Perspectives

Consultants in scholarly information

THE SKILLS, ROLE AND CAREER STRUCTURE OF DATA SCIENTISTS AND CURATORS: AN ASSESSMENT OF CURRENT PRACTICE AND FUTURE NEEDS

REPORT TO THE JISC

September 2008

Prepared by:
Alma Swan and Sheridan Brown

- *Data Creators*
 - autores de los datos
- *Data Scientist*
 - trabajando junto al investigador
 - con conocimiento del dominio
- *Data Managers*
 - expertos en computación, almacenamiento y preservación
- *Data Librarians*
 - Especializados en *curation*, preservación y archivos

<http://eprints.soton.ac.uk/id/eprint/266675>

What is a Data Scientist? (...Data Scientists in the Wild...)

Dr Liz Lyon,
Associate Director, Digital Curation Centre,
Director, UKOLN, University of Bath, UK
Dr Kenji Takeda,
Microsoft Research Connections

Microsoft eScience Workshop, Chicago, October 2012



This work is licensed under a Creative Commons Licence
Attribution-ShareAlike 2.0



UKOLN is supported by:



www.ukoln.ac.uk

A centre of expertise in digital information management

JISC Microsoft Research



- *Data Engineer*
 - desarrollo de software y programación
- Data Analyst
 - análisis y estadística
- Data Librarian
 - gestión de datos
- Data Steward
 - preservación, repositorios y archivos
- Data Journalist
 - generación de noticias e historias

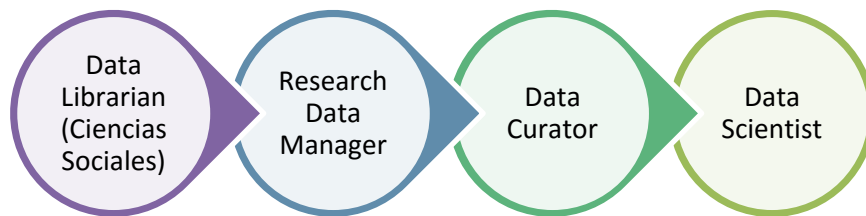
<http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/liz-lyon-microsoft-escience-chicago-october-2012-final.pdf>

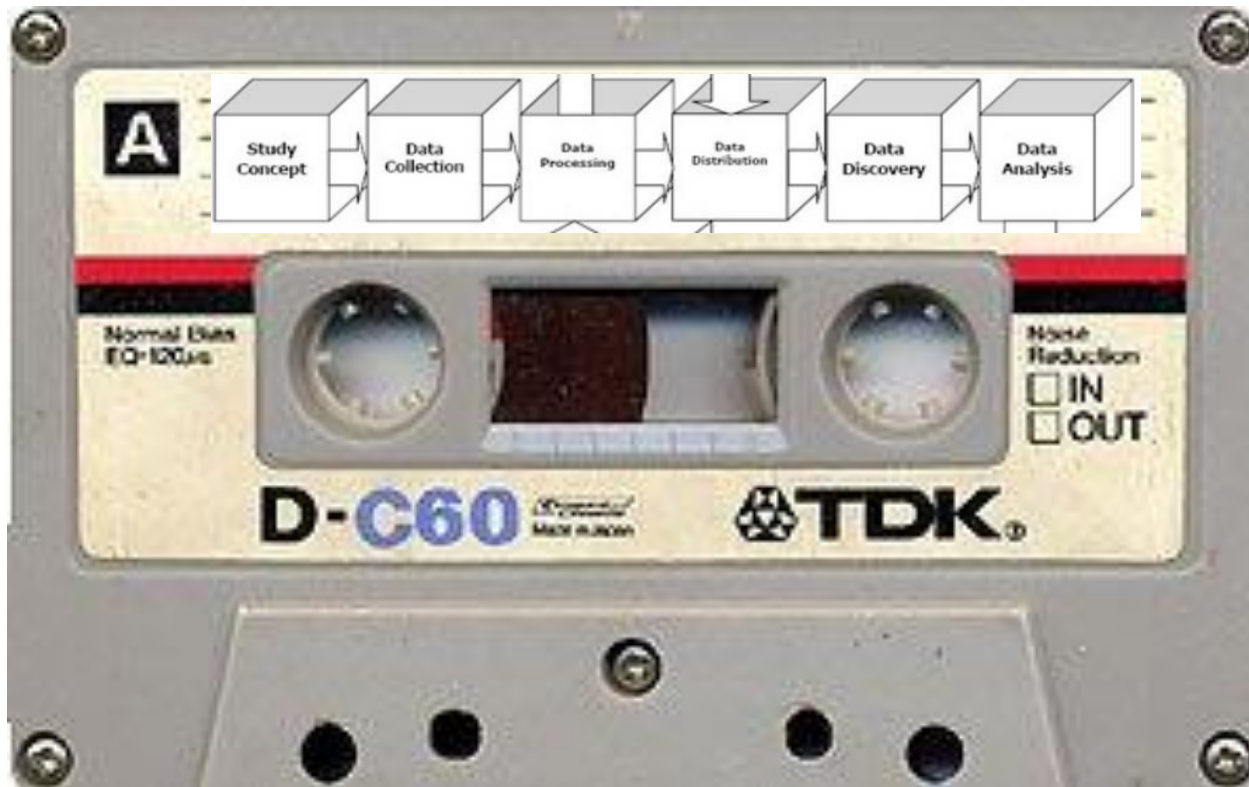
Una perspectiva evolutiva desde el punto de vista de un *Data Activist*



"The activist is not the man who says the river is dirty. The activist is the man who cleans up the river."

Ross Perot





DATA LIBRARIAN (CIENCIAS SOCIALES)



*“social science data archives have developed in recent years because of the increasing need of many social scientist to have **access** to the growing body of quantitative data produced by research organizations all over the world”*

Ralph L. Bisco *“Social Science Data Archives: A review of Developments”*
The American Political Science Review Vol 6. No.1 **1966**

Data Libraries & Data Centres

Data Libraries



Berckley (1958)



Edinburgh (1983)



Cornell (1987)



Oxford (1988)

Data Centres



Roper Center (1957)



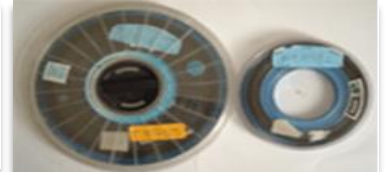
ICPSR (1960)



UKDA (1967)



IOP (1963) y CIS (1977)



Fotos tomadas de www.data-archive.ac.uk/ukda40/

Datos en Ciencias Sociales

- **Cualitativos:** medidas expresadas en lenguaje natural

Pregunta ¿En que consiste el nacionalismo?

Respuesta: El ser nacionalista supone un doble sentimiento...

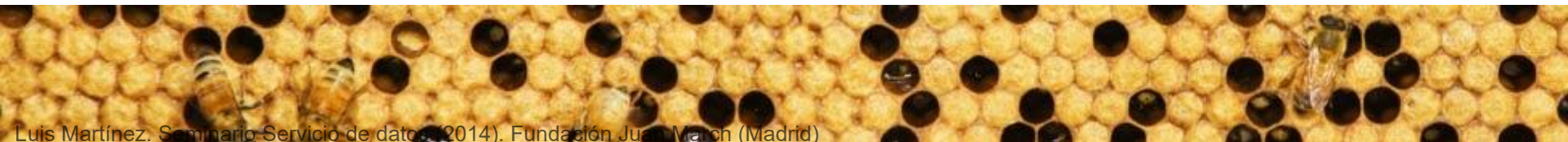
- **Cuantitativos:** medidas expresadas en términos de números

Pregunta ¿Cómo califica la gestión del gobierno?

Respuesta Muy buena -1 Buena -2 Mala -3

...

- **Microdatos:** Son datos a nivel de individuo que presentan sus respuestas a un cuestionario. Suelen ser encuestas del gobierno o proyectos de investigación, datos longitudinales y de opinión.
- **Datos agregados:** Son datos que se agregan por área geográfica. Datos socioeconómicos de organizaciones intergubernamentales mayoritariamente.
- **Datos geográficos (sistemas de información geográfica):** Combinación de datos con referencias geográficas, mapas y software de sistemas de información geográfica.



CEACS

Centro de Estudios Avanzados
en Ciencias Sociales



Biblioteca CEACS, Instituto Juan March

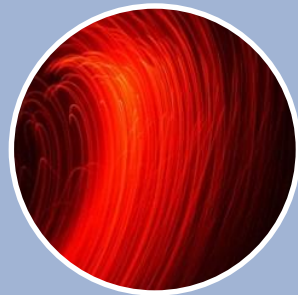
- CEACS comienza en 1987
- Ciencia política, sociología y economía política
- Enfoque anglosajón, modelo cuantitativo
- Biblioteca CEACS
 - Trabajo de cerca con su comunidad científica
 - Servicios virtuales a recursos electrónicos
 - Archivo Linz de la Transición Española
 - Un servicio de datos

Biblioteca de Datos de CEACS



Acceso

- Gestión y catalogación de la colección
- Adquisición de nuevos datos y licencias
- Servicio de referencia



Uso

- Software especializado
- Soporte uso y manipulación

← INFRAESTRUCTURA , SOPORTE, FORMACIÓN →

Data Librarian - London School of Economics



You will be responsible for ***managing the Data Library's collection*** of electronic datasets as well as ***providing site support for usage of national data centres***. You will provide a ***data advisory service*** consisting of training programmes and individual dataset surgeries and will manage the datasets budget.

Job description September 2013



“By open access, we mean the **free availability** of articles on the public internet, permitting any users to **read, download, copy, ...**”

- The Budapest Open Access Initiative –
February 14, 2002

Growth of the OpenDOAR Database - Worldwide

OpenDOAR - 09-Jul-2014



Requisitos de las agencias de financiación



*“... **open access to scientific data** should be adopted as the international norm for the exchange of scientific data derived **from publicly funded research.**”*

OECD Principles and Guidelines for Access to Research Data from Public Funding (2004-2007)

*“**requires, in all proposals** a supplementary document of no more than two pages describing **a Data Management Plan** for the proposed research. “*

The National Science Foundation, January 2011

*“...**primary data**, as well as data-related products such as computer codes, is deposited in the relevant databases **as soon as possible**, preferably immediately after publication and in any case **not later than six months after the date of publication.**”*

European Research Council , Open Access Guidelines for Researchers, June 2012

Requisitos e incentivos de las revistas y editores

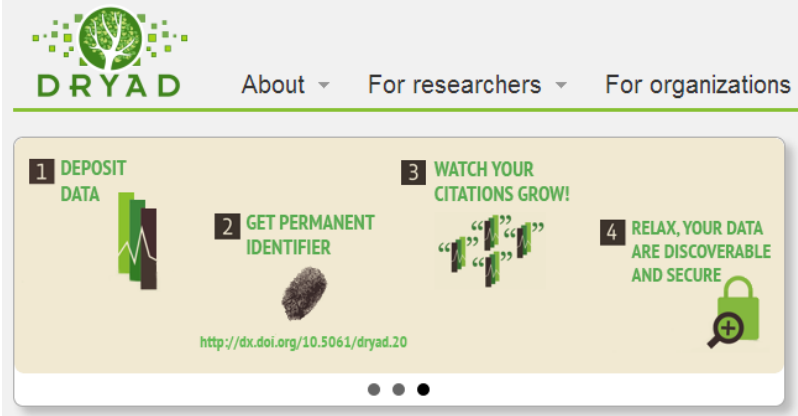


Data Policy



PLOS' New Data Policy: Public Access to Data

By Liz Silva
Posted: February 24, 2014



Requisitos de las Universidades

“**Research data** of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, **will be offered** and assessed **for deposit** and retention in an appropriate **national or international data service** or domain repository, or a University repository.”

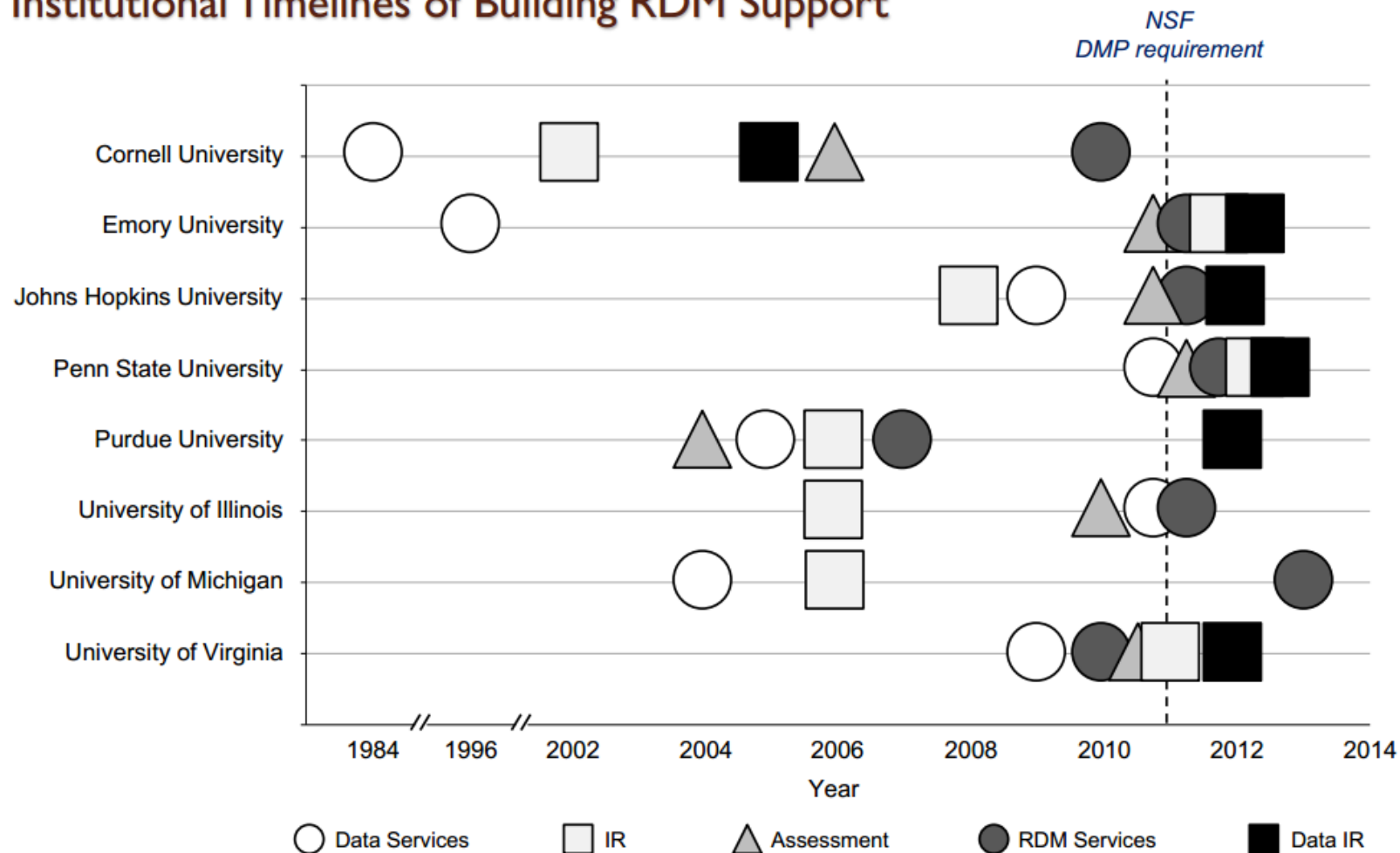
“The **University will provide mechanisms** and services for **storage, backup, registration, deposit and retention** of research data assets in support of current and future access, during and after completion of research projects.”

“Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.”

Edinburgh University Data Policy 2011



Institutional Timelines of Building RDM Support



Biblioteca de Datos CEACS

Servicios de gestión de datos

Harvard Dataverse Network >

Instituto Juan March - Center for Advanced Study in the Social Sciences (CEACS) Dataverse



Welcome to the Instituto Juan March - CEACS Dataverse!

The Instituto Juan March CEACS - Dataverse is administered by the CEACS Data Library to manage, disseminate and curate the data generated by CEACS researchers. For any queries please email the CEACS Data Librarian.

The screenshot displays the CEACS Dataverse search results page. On the left, there is a navigation menu with "Instituto Juan March - CEACS" selected, and sub-items for "Explaining Terrorist Target Selection" and "Replication data". The main content area features a search bar with "Search Studies" and a "Search" button. Below the search bar, there are options for "Sort By: Global ID" and "Advanced Search Tips". The results show two entries. The first entry is "The Domestic Terrorist Victims dataset" by Ignacio Sánchez-Cuenca and Luis de la Calle Robles, with a description of the dataset covering Western European countries from 1965-2005. It includes production and distribution dates (2009), related publications, and replication data for "Revolutionary Terrorism in the Developed World, 1970-2000". The second entry is "Replication data for: Comparative lottery sales data in Europe 1850-1913". Each entry includes an HDI link, download/analysis statistics, and the last release date (May 20, 2010).

- A partir del 2010
- Soporte en la creación de datos primarios
 - Codificación de datos
 - Documentación y creación de libros de códigos
 - “Data scraping”
- Espacio de almacenamiento y para compartir
- [Repositorio para publicación y preservación](#)

Final report of the LIBER working group on E-Science / Research Data Management

Ten recommendations for libraries to get started with research data management

1. Offer research data management support, including data management plans for grant applications, intellectual property rights advice and information materials. Assist faculty with data management plans and the integration of data management into the curriculum.
2. Engage in the development of metadata and data standards and provide metadata services for research data.
3. Create Data Librarian posts and develop professional staff skills for data librarianship.
4. Actively participate in institutional research data policy development, including resource plans. Encourage and adopt open data policies where appropriate in the research data life cycle.

5. Liaise and partner with researchers, research groups, data archives and data centers to foster an interoperable infrastructure for data access, discovery and data sharing.
6. Support the lifecycle for research data by providing services for storage, discovery and permanent access.
7. Promote research data citation by applying persistent identifiers to research data.
8. Provide an institutional Data Catalogue or Data Repository, depending on available infrastructure.
9. Get involved in subject specific data management practice.
10. Offer or mediate secure storage for dynamic and static research data in co-operation with institutional IT units and/or seek exploitation of appropriate cloud services.

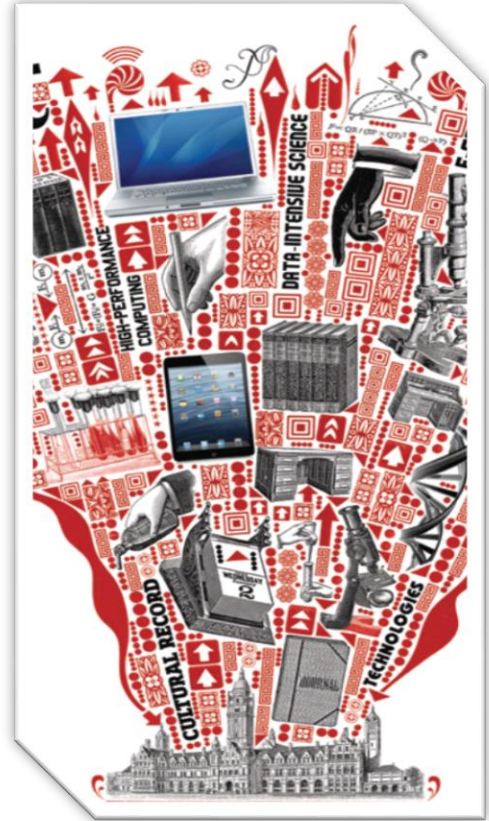
Research Data Services Manager - University of Michigan

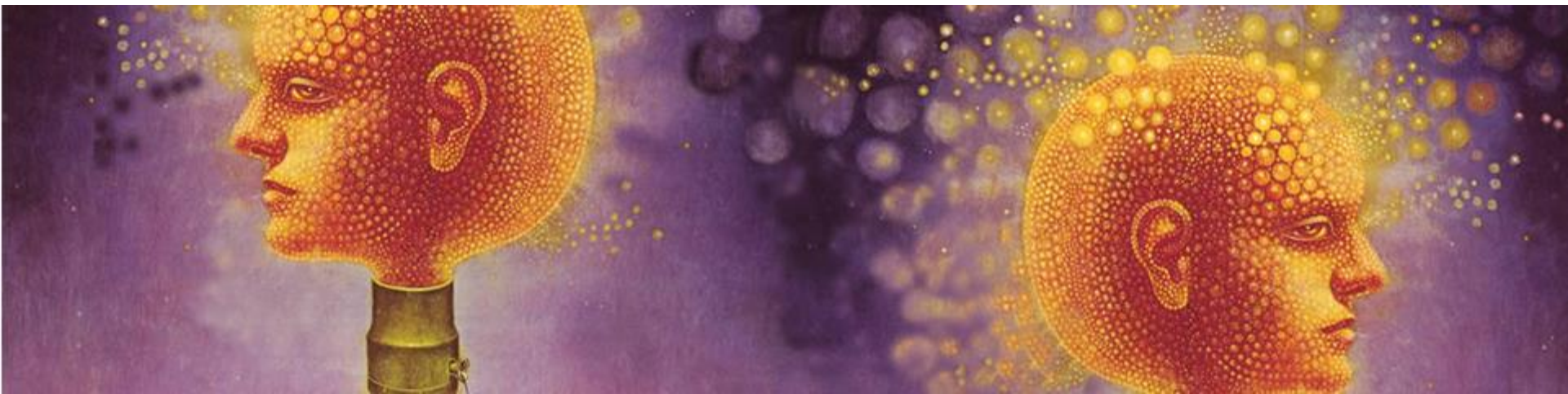
The Research Data Services Manager will help build and enhance Research Data Services (RDS), which are responsible for strategic planning, coordination and deployment of research data services directed at **facilitating the research lifecycle**. This includes creating and implementing **data management assistance** for the campus, **outreach** to faculty in collaboration with subject specialists, **training**, and assessment of RDS programs and services.

The Research Data Services Manager develops and promotes 4 key areas:

- 1) Education, Awareness and Community Building,
- 2) **Infrastructure**,
- 3) **Policy** and Strategy, and
- 4) **Consultation** and Services (provision of **Data Management Plan** support services)

Job description December 2013





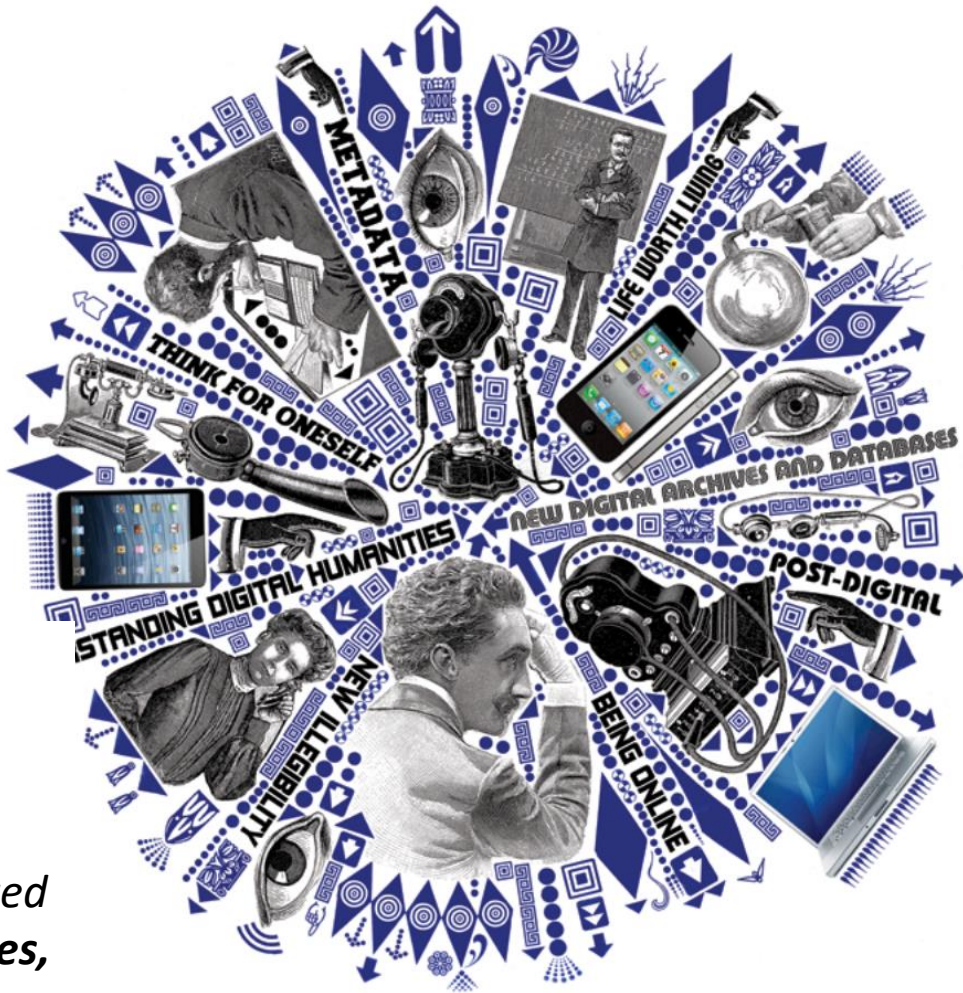
DATA CURATOR

Post-Digital Humanities:

Computation and Cultural Critique
in the Arts and Humanities

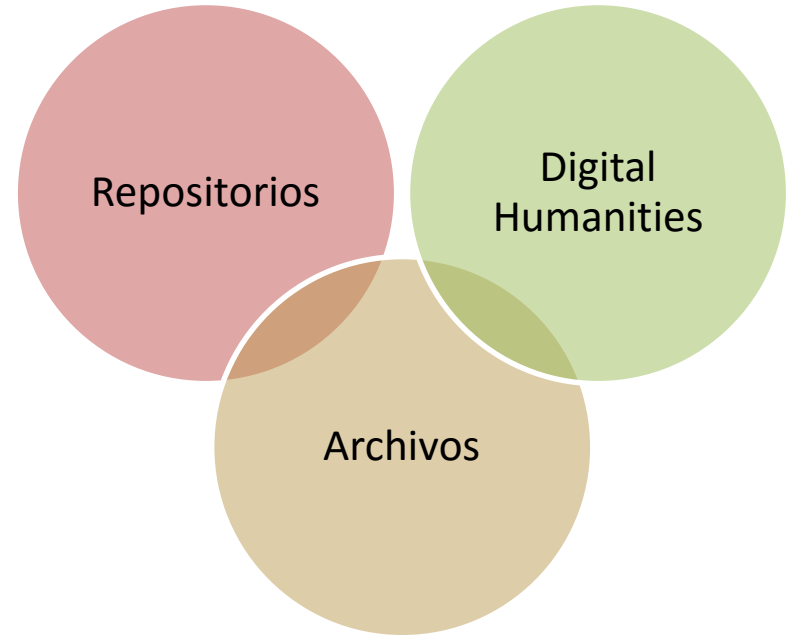
By David M. Berry

*“The **digital humanities**, at its most straightforward, is the application of **computational principles, processes, and machinery** to humanities texts—and here I use “texts” to refer to all forms of materialized cultural forms such as **images, books, articles, sound, film, video, and so on**”*



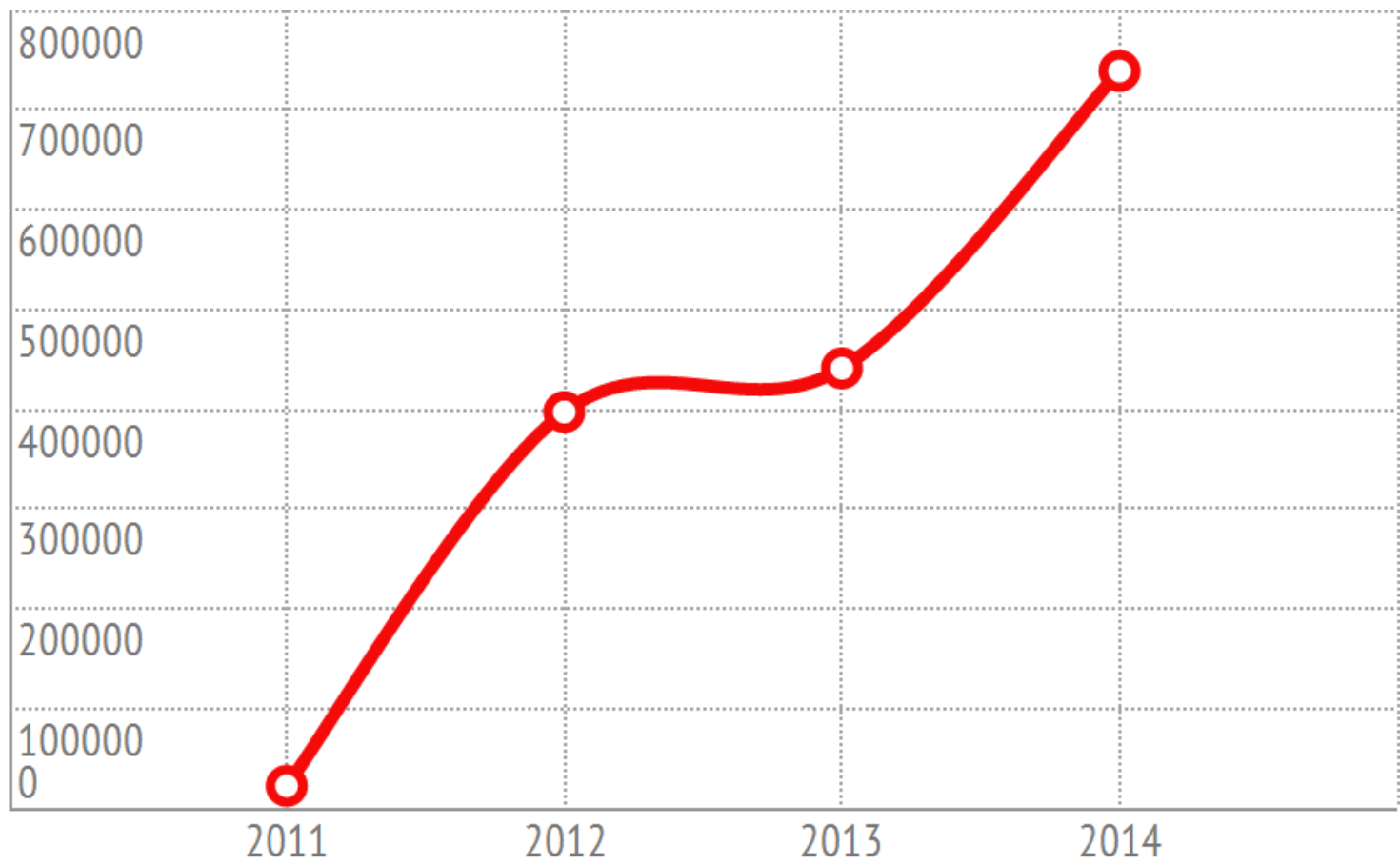


Bibliotecas Digitales





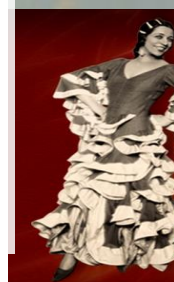
Descripción | Archivo de C



objetos en el repositorio



ecopilar,
olia
cos de la



Data Curator - New York University

You will work in **close collaboration with individual faculty and researchers** across the country in preparing files for deposit, including

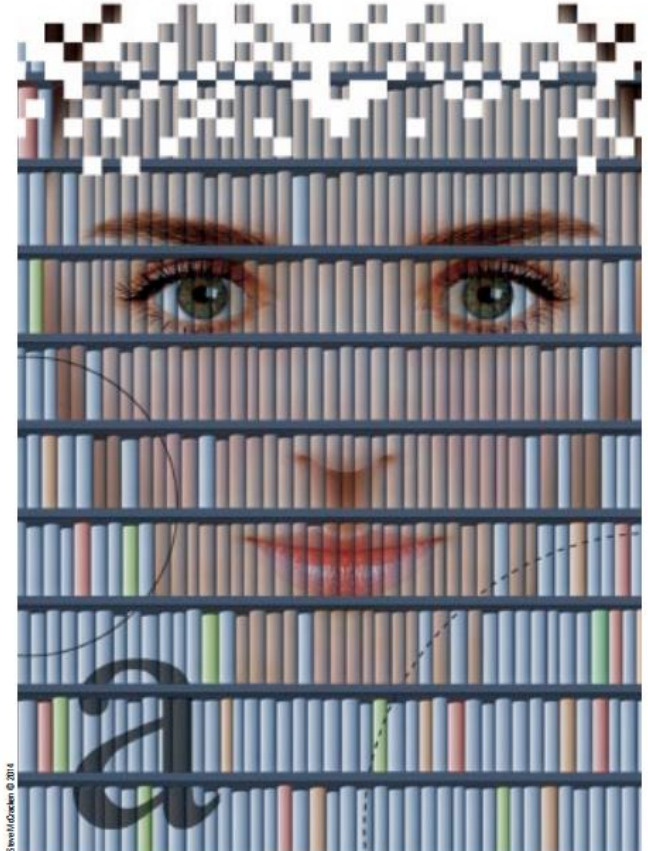
- **Analysis** of the structure of study data;
- **Data normalization**, cleaning, authority management;
- Organization of **digital and physical inventories**.

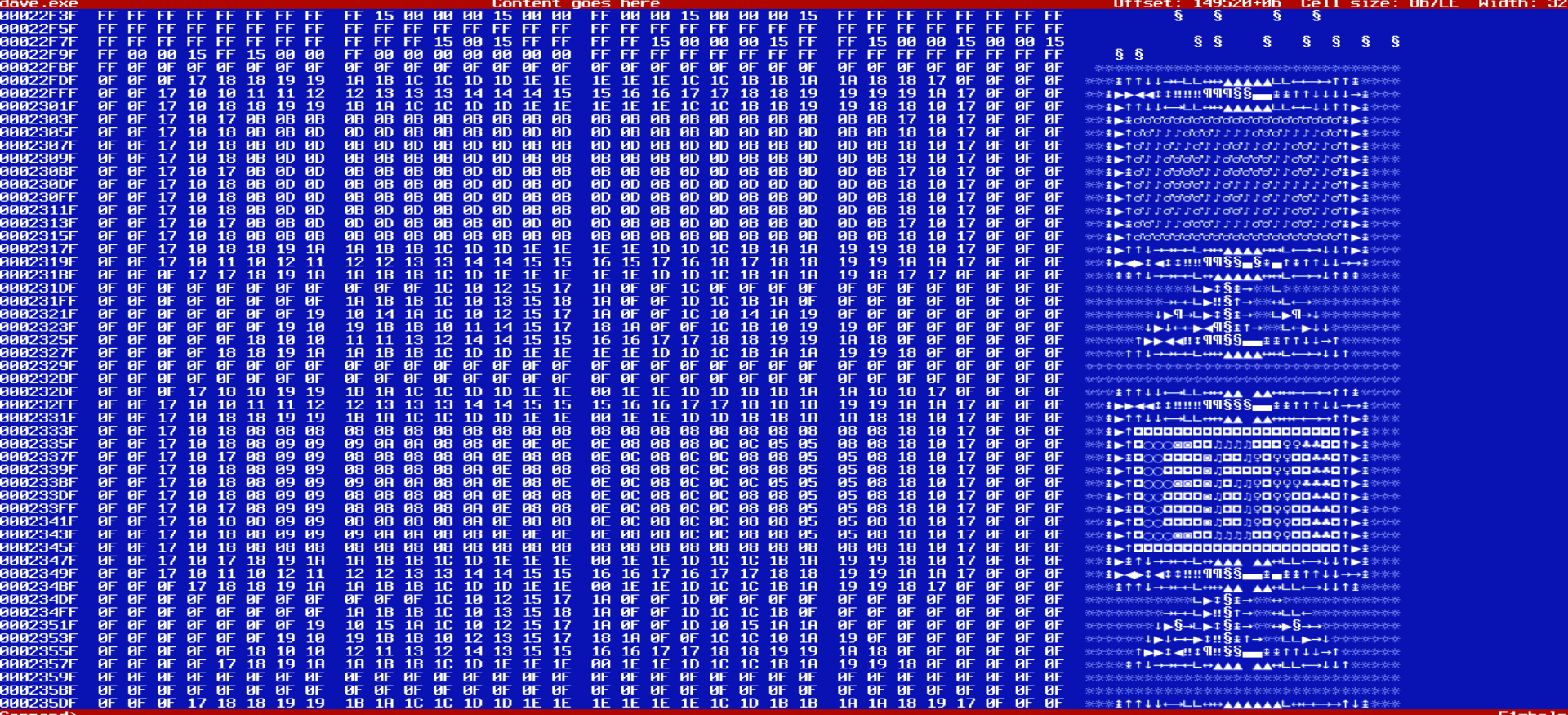
You will also work in **close collaboration** with the Datalibrary **software developers** to

- Devise **metadata schema** that are flexible but that also offer incentives to consistency;
- Create best practices for **long-term sustainability** of data formats, including video formats, such as format migration and transcoding;
- Create **deposit, maintenance, and annotation** workflows;
- Create **taxonomy(ies)** and initial tag sets.

Job description June 2013

Luis Martínez. Seminario Servicio de datos (2014). Fundación Juan March (Madrid)

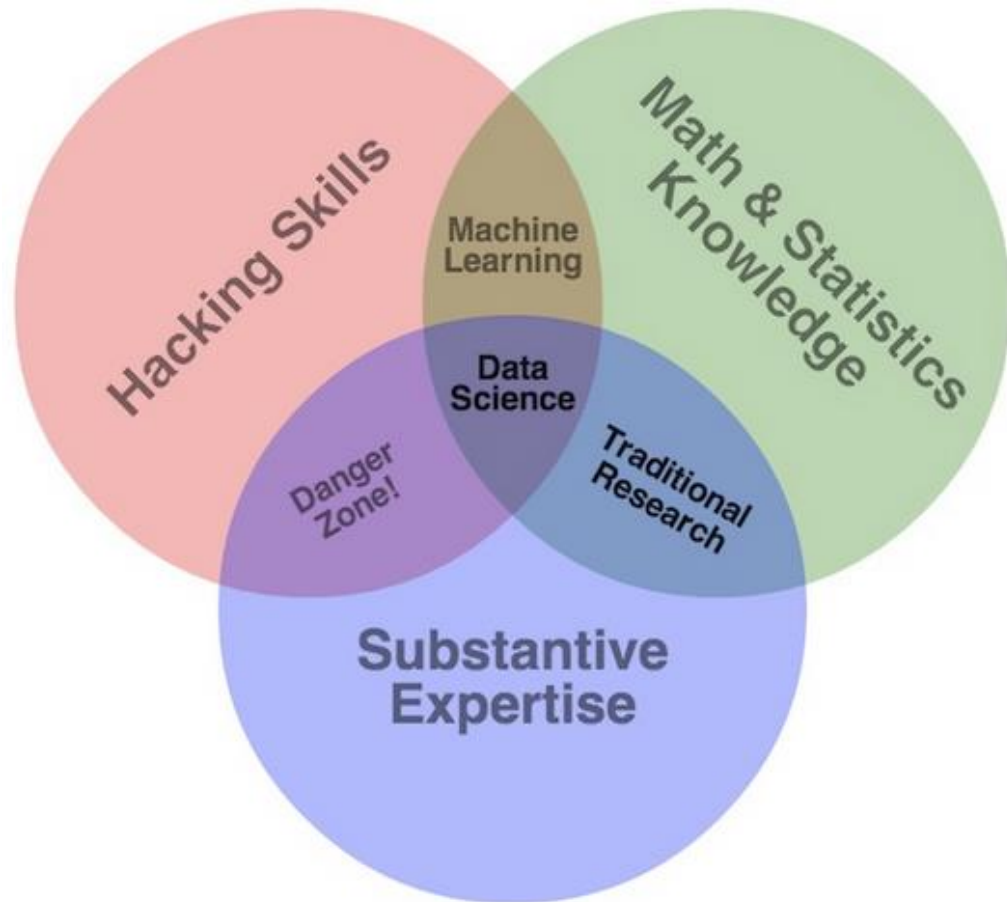




DATA SCIENTIST

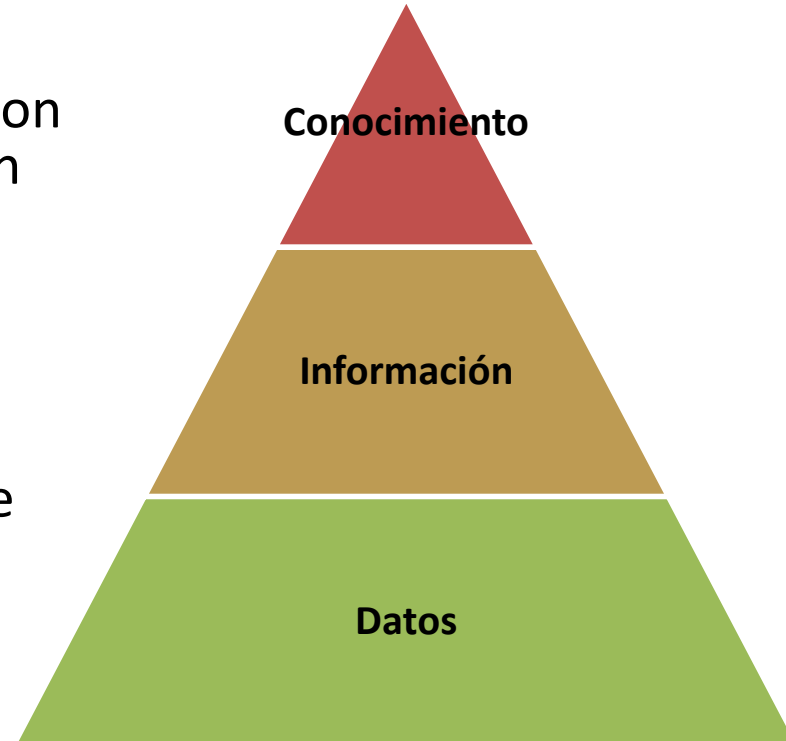
...data science helps an organization turn data into information, information into knowledge and insights and valuable, actionable insights into better decision making and game changing strategies

Data Science Central, Julio 2013



Organización del Conocimiento Digital de la Fundación Juan March

- Desde Septiembre 2013
- Un nuevo eje de trabajo de la Biblioteca con los demás departamentos de la Fundación
- Objetivos
 - analizar y enriquecer los datos
 - evaluar, descubrir, crecer
 - comunicar
- Febrero 2014 convenio con la Facultad de Matemáticas de la UCM
 - Master de tratamiento Estadístico Computacional de la Información



14.30 – 15.30 Casos prácticos

- Conversión de datos textuales a datos numéricos a partir de una base de datos de recortes de prensa. (Almudena Knecht)
- Estudio de caso de algoritmos de *Machine learning* para asignación automática de materias. (Daria Dubinoshka)

16.00 – 16.30 Casos prácticos

- Visualizaciones y gráficos de datos con el programa Gephi (Diego Vera)
- Explicación, interpretación de los resultados, análisis y puntos a mejorar (Diego Vera, Paz Fernández, Luis Martínez)

