

EVENTOS

XV WORKSHOP REBIUN DE PROYECTOS
DIGITALES. DATOS Y BIBLIOTECAS

2016



Bibliotecas: velando por la calidad y el valor de los datos



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

Bibliotecas: velando por la **calidad** y el **valor** de los datos

Dra. Montserrat Garcia Alsina

Profesora de los Estudios de Ciencias de la Información y la Comunicación

Universitat Oberta de Catalunya

mgarciaals@uoc.edu



Castellón, 30 setiembre, 2016

Contenidos

¿**Por qué** datos deben velar las bibliotecas? ¿**Para qué**?

- + Datos, datos internos / datos externos
- + datos abiertos, grandes volúmenes
- + Características de los datos

Calidad de los datos

- + Gestión de datos: cadena de valor
- + Indicadores de calidad
- + Gobernanza de la información

Valor

- + Análisis
- + Generación de conocimiento para la acción
- + Gestión del conocimiento, inteligencia competitiva y gestión de datos masivos

Calidad y valor de los datos: **competencias**

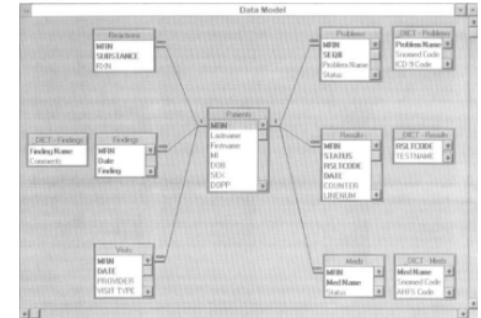
- + Habilidades y conocimientos
- + Perfiles

1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

Datos
Datos internos
Datos externos

Del papel a la informática

Estructurados



Semiestructurados

```
<?xml version="1.0"?>  
<SCHOOL>  
  <CLASS type="seminar">  
    <CLASS_TITLE>XML In The Real World</CLASS_TITLE>  
    <CLASS_NUMBER>6.031</CLASS_NUMBER>  
    <SUBJECT>XML</SUBJECT>  
    <START_DATE>6/1/2002</START_DATE>  
    <STUDENTS>  
      <STUDENT status="attending">  
        <FIRST_NAME>Edward</FIRST_NAME>  
        <LAST_NAME>Samson</LAST_NAME>  
      </STUDENT>  
      <STUDENT status="withdrawn">  
        <FIRST_NAME>Ernestine</FIRST_NAME>  
        <LAST_NAME>Johnson</LAST_NAME>  
      </STUDENT>  
    </STUDENTS>  
  </CLASS>  
</SCHOOL>
```

Desestructurados



1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

procedencias

Internet	Crowdsourcing	IoT	Datos internos (organizativos)	Datos abiertos
términos de búsquedas entradas a foros grabaciones de chats mensajes de blogs Web ...	Open call Innovación social Inteligencia colectiva Votación colectiva	GPS PDA, alarmas Teléfonos móviles sensores (temperatura, climatología, velocidad) ...	Aplicaciones informáticas Videos Imágenes Documentos Clickstreams E-mails ...	Públicos (portales de datos abiertos) Portales de transparencia Investigaciones científicas ...

Datos
Datos internos
Datos externos



The screenshot shows the 'UGR TRANSPARENTE' website. The header includes the UGR logo, 'Administración electrónica', and a search bar. The main content area features a navigation menu on the left with items like 'Inicio', 'Información Institucional', 'Administración', 'Decencia', 'Gestión e Investigación', 'Normativa Legal', and 'Solicitud de Información'. The central area has a grid of buttons for 'Información Institucional', 'Personal', 'Información Económica', 'Oferta y Demanda Académica', 'Clausuro', 'Estudiantes', 'Gobierno', 'Rendimiento', and 'Normativa Legal'. On the right, there are vertical panels for 'Portal OpenData UGR', 'Portal de Transparencia de la UGR', 'Boletín UGR', and 'La UGR en cifras'.

1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

Internet	Crowdsourcing	IoT	Datos internos (organizativos)	Datos abiertos
términos de búsquedas entradas a foros grabaciones de chats mensajes de blogs Web ...	Open call Innovación social Inteligencia colectiva Votación colectiva	GPS PDA, alarmas Teléfonos móviles sensores (temperatura, climatología, velocidad) ...	Aplicaciones informáticas Documentos Clickstreams E-mails ...	Públicos (portales de datos abiertos) Portales de transparencia Investigaciones científicas ...

Datos
 Datos internos
 Datos externos



A screenshot of the 'Portal Transparente' website for the University of Granada (UGR). The header includes the UGR logo and 'Administración electrónica'. A search bar is present. The main content area features a grid of navigation buttons: 'Información Institucional', 'Personal', 'Información Económica', 'Oferta y Demanda Académica', 'Claustro', 'Estudiantes', 'Gobierno', 'Rendimiento', and 'Normativa Legal'. On the right side, there are vertical panels for 'Portal OpenData UGR', 'Portal de Transparencia de la DGE', 'Portal dedicado a Transparencia y acceso a datos abiertos de la Delegación General de Estudiantes', 'Boletín UGR', and 'La UGR en cifras'.

1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

Características datos



Volumen

Variedad de formatos

Velocidad

Veracidad

Valor

Visualización

Verificación

Variabilidad

Viabilidad

1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

Datos

Datos internos

Datos externos

Estructurados

Datawarehouse

Semiestructurados

Business Intelligence

Desestructurados

Big Data (NoSQL)

1. ¿Por qué datos deben velar las bibliotecas? ¿Para qué?

Datos

Datos internos

Datos externos

Estructurados

Datawarehouse

Semiestructurados

Business Intelligence

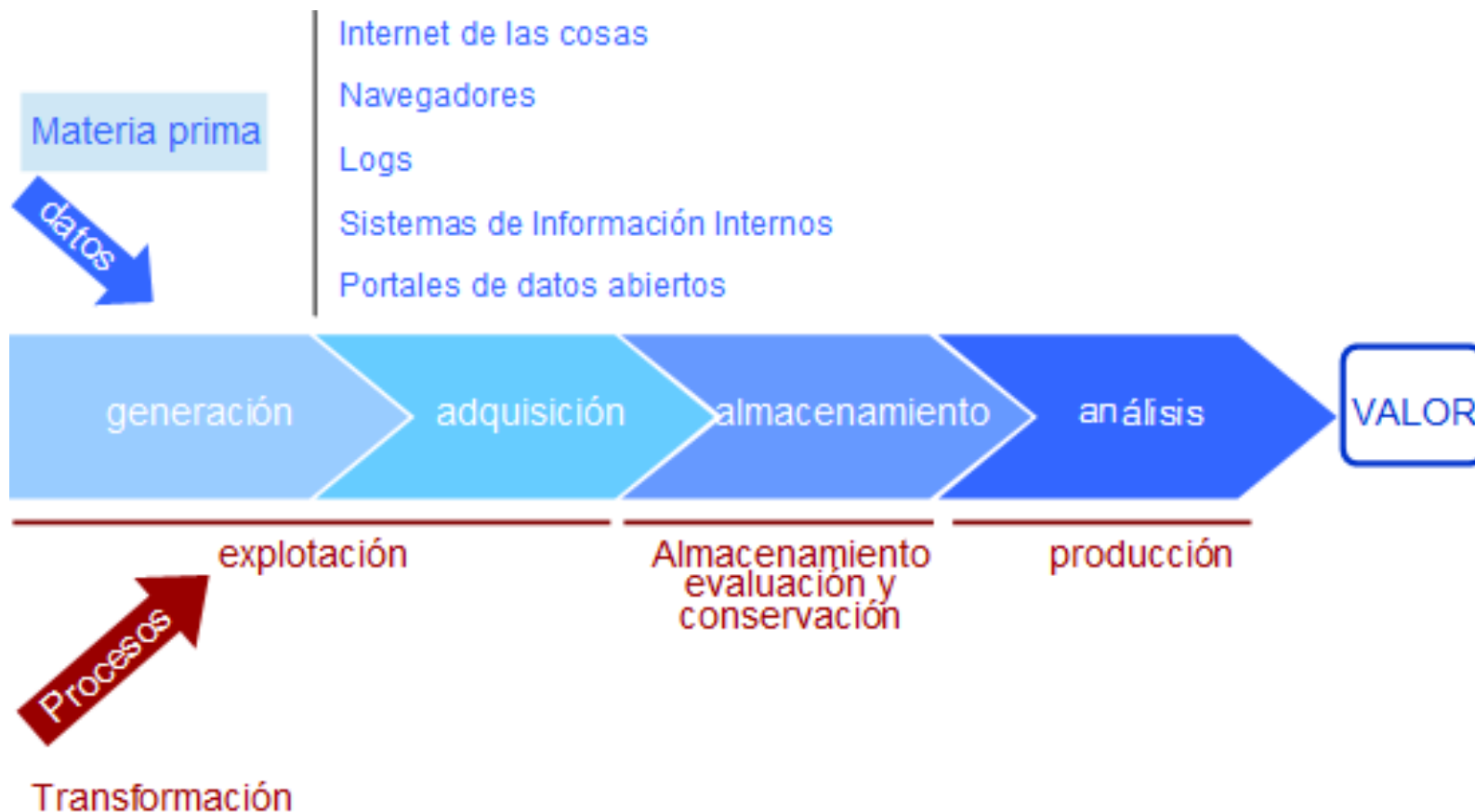
Desestructurados

Big Data (NoSQL)

VALOR

2. Calidad de los datos: gestión

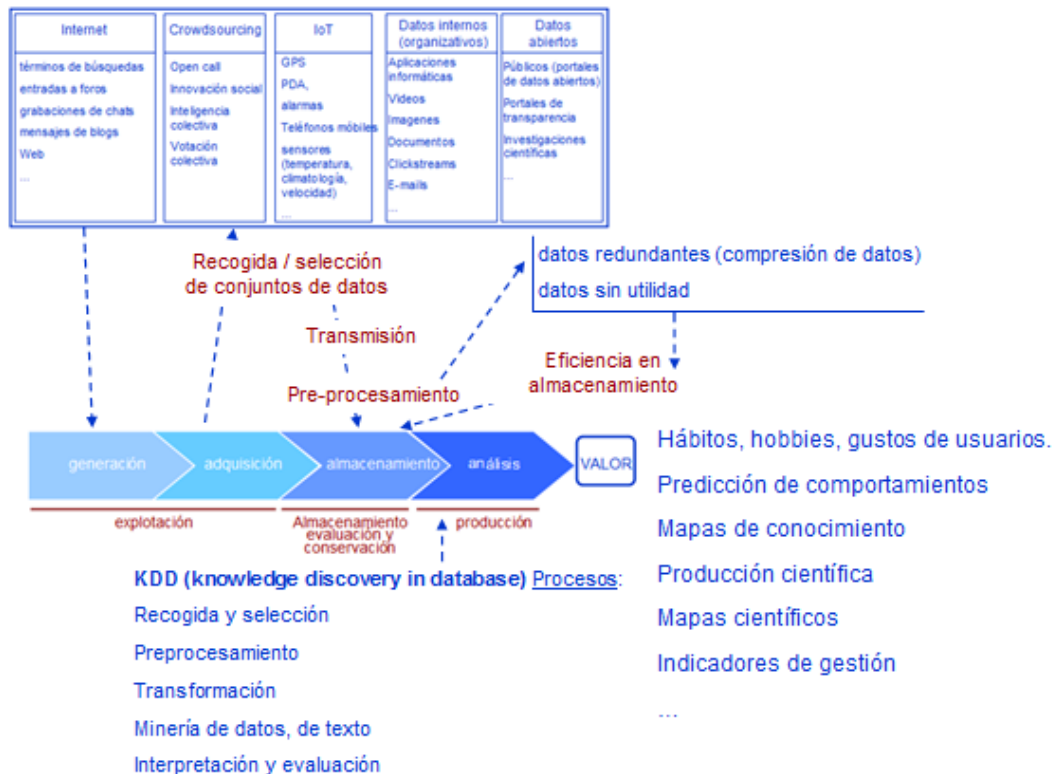
Cadena de valor



2. Calidad de los datos: cadena de valor



2. Calidad de los datos



Indicadores de calidad (datos)

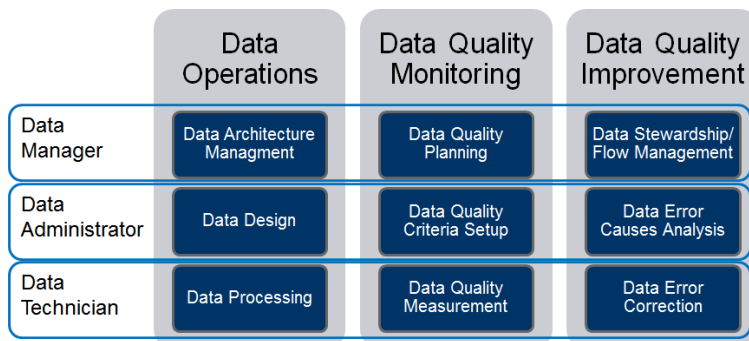
- Precisión
- Compleitud
- Relevancia
- Validez de los datos
- Proveniencia
- Autenticidad
- Veracidad
- Exactitud
- Reputación
- Credibilidad

Indicadores de calidad en gestión (procesos)

(McDonald y Léveillé, 2014; Gironés, 2013; Thurston, 2012; Zhao y Hartig, 2012; ISO 8000-150)

2. Calidad de los datos: gestión

Marco para gestionar la calidad



ISO 8000

(Fuente: Data and Process advantage blog <https://dpadvantage.wordpress.com/2014/12/23/iso-8000150a-framework-for-data-quality-management/>)

ISO/TS 8000-1:2011 Data quality -- Part 1: Overview.

ISO/TS 8000-100:2009 Data quality -- Part 100: Master data: Overview.

ISO 8000-2:2012 Data quality -- Part 2: Vocabulary.

ISO/TS 8000-120:2009 Data quality -- Part 120: Master data: Exchange of characteristic data: Provenance.

ISO/TS 8000-130:2009 Data quality -- Part 130: Master data: Exchange of characteristic data: Accuracy.

ISO 8000-110:2009 Data quality -- Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification.

ISO/TS 8000-140:2009 Data quality -- Part 140: Master data: Exchange of characteristic data: Completeness.

ISO/TS 8000-150:2011 Data quality -- Part 150: Master data: Quality management framework.

ISO/TS 8000-311:2012 Data quality -- Part 311: Guidance for the application of product data quality for shape (PDQ-S).

ISO 8000-8:2015 Data quality -- Part 8: Information and data quality: Concepts and measuring.

2. Calidad de los datos → gobernanza de la información

Necesidades de datos, información

Necesidad de gestionar y crear conocimiento

Crecimiento de aplicaciones

Años '90
(XX) →



Años '10
(XXI) →

- Auditoria (administraciones públicas)
- *New Public Administration*
- Rendimiento de cuentas,
- Cumplimiento de legislación
- Evidencias de las actividades

- Gobierno abierto
- Transparencia
- Datos en abierto
- Ley de reutilización de la información
- Datos masivos (*big data*)
- Potencia de los sistemas informáticos
- Incertidumbre del entorno
- ...

2. Calidad de los datos → gobernanza de la información

¿Qué es gobernanza de la información?

¿Cómo se gobiernan los datos, la información, el conocimiento?

¿Por qué?

¿Quién? Cuando? ¿Dónde? (aspectos organizativos)

(Guetat y Dakhli, 2015)

“**Governance** refers to the structures, processes, cultures and systems that engender the successful operations of the organizations”. Therefore, **governance is broader than management** since it encloses both a managerial and a cultural dimension” (Keasey et al.)

Information governance is “the activities and practices developed to control the use of information, including, but not limited to, practices mandated by law” (McLennan, 2014)

“The main **goal of information governance** consists in contributing to organizations competitive advantage by creating a holistic approach to manage important organizational information” (Guetat i Dakhli, 2015)

“**Information governance** is the formulation of policy, to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions” (Soares, 2011)

Information governance “involves establishing an environment of opportunities, rules, and decision making rights for the valuation, creation, collection, analysis, distribution, storage, use and control of information”, (Koooper et al., 2011)

2. Calidad de los datos

Gestión de la información

Gobernanza de la información



3. Valor de los datos: Análisis de datos



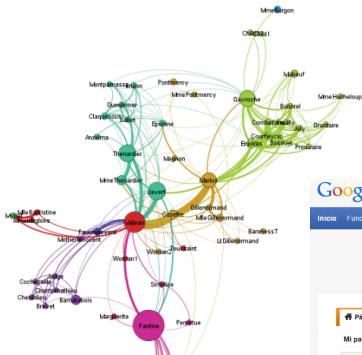
De acuerdo a las **necesidades**

- ◆ Táctico
- ◆ Operativo
- ◆ Estratégico

Diversidad de **fuentes**: procedencia de diferentes áreas y procesos

Conexión de datos

- Variación parámetros e indicadores
- Conexiones
- Patrones



Puntos de vista

- ◆ Perspectiva o descriptiva (ahora y como)
- ◆ Prospectiva o predictiva (futuro)
- ◆ Prescriptiva: opciones



Generación de conocimiento para la acción

3. Valor de los datos: generación de conocimiento para la acción



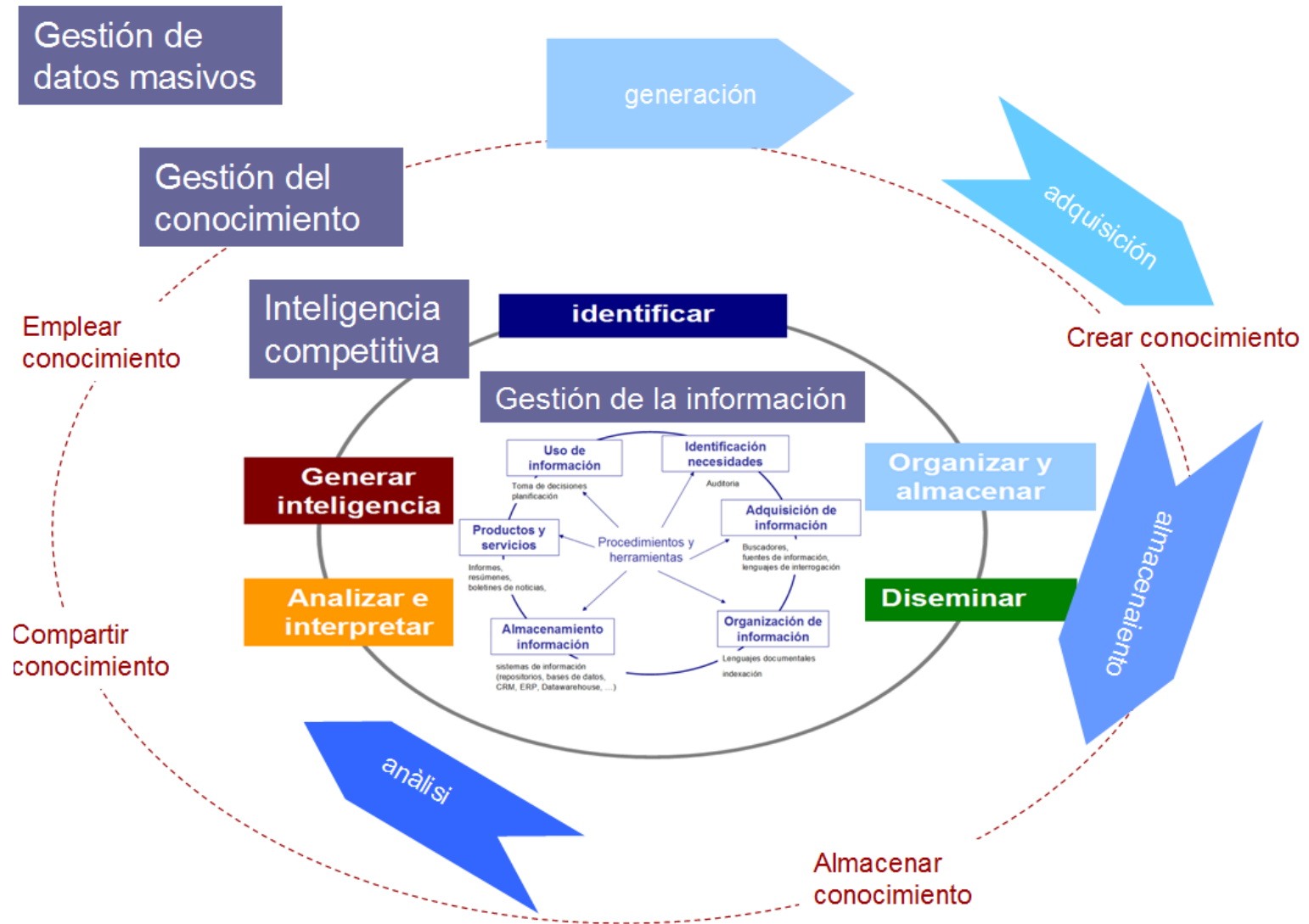
Gestión organizativa
Cuadro de mando
Conocimiento del entorno (IC)
Producción
Contabilidad y finanzas

Conocimiento del cliente:

OPAC: Número de usuarios,
frecuencia de visitas
Evolución de perfiles
Documentos más solicitados

Big data: Acciones del usuario (trazabilidad)
Contenidos buscados
Explotación de la colección:
descubrimiento de conocimiento,
mapas...
Tendencias en investigación

3. Valor de los datos: gestión del conocimiento, inteligencia competitiva y gestión de datos masivos



4. Calidad y valor de los datos: competencias y habilidades

Top 10 IT skills demanded:

1 - process/methodological skills

- Agile Software Development
- TDD
- OO
- Scrum
- SOA
- Continuous Integration
- Unit Testing
- BDD
- Multithreading
- OOP

2 - applications/technologies/vendors

- Oracle
- Java
- SQL
- Linux
- JavaScript
- SQL Server
- Python
- PHP
- UNIX
- CSS

Top 10 data-related skills demanded:

1 - process/methodological skills

- Business Intelligence
- NoSQL
- Data Warehouse
- big data
- ETL
- Analytics
- Data Modelling
- RDBMS
- Data Mining
- Data Management

2 - applications/technologies/vendors

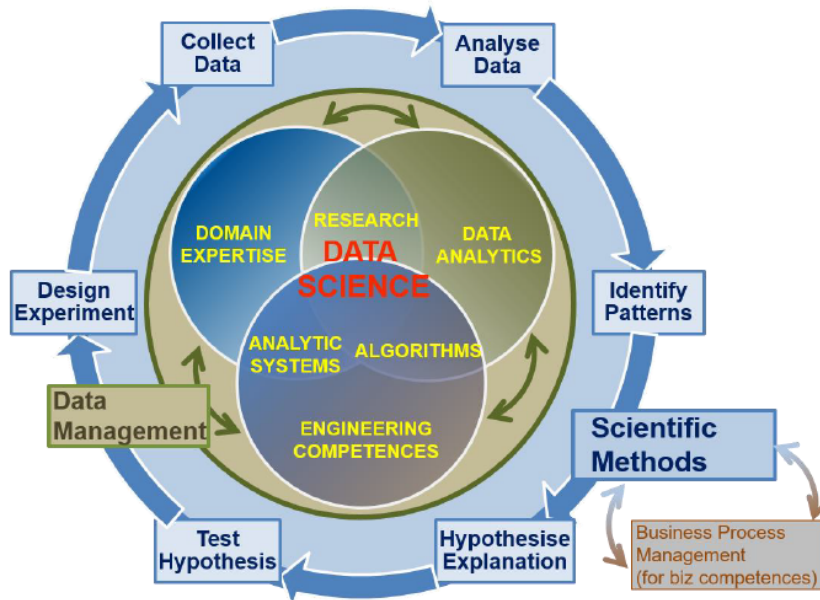
- Oracle BI EE
- MongoDB
- MySQL
- Hadoop
- Informatica
- Amazon EC2
- Apache Cassandra
- Oracle 11g
- PL/SQL
- Oracle EBS

4. Calidad y valor de los datos: competencias y habilidades



Education for Data Intensive Science to Open
New science frontiers

5 grupos de competencias para el Científico de datos:



- Analista de datos (métodos estadísticos, lenguaje de máquina i análisis de negocio)
- Ingeniería: software e infraestructura.
- Competencias y conocimiento del campo científico.
- Gestión de datos, *curation*, preservación
- Métodos científicos o de investigación

4. Calidad v valor de los datos: competencias y habilidades



Education for Data Intensive Science to Open New science frontiers

Grupo “Gestión de datos, *curation*, preservación”

- Tratamiento y análisis de datos complejos, de gran volumen, alta dimensionalidad procedentes de varias fuentes
- Fuentes de datos y técnicas para mejorar los datos y focalizarse en el cliente y en mejores procesos de negocio.
- Modelos de datos.
- Experiencia de trabajo con extensos data sets.
- Datos no relacionales i datos desestructurados.
- Almacenamiento basado en la nube y gestión de datos.
- Planificación de gestión de datos.
- Registro y gestión de metadatos.
- Citación de datos, metadatos, PID (Persistent Identifier), Registros de tipos de datos, políticas de gestión de datos.

4. Calidad y valor de los datos: perfiles / ocupaciones



Education for Data Intensive Science to Open New science frontiers

Top level	Hierarchies existing and new	Occupations group (if any)	Occupations
Managers			
	Production and specialised services managers	Data Science/Big Data Infrastructure Managers	Data Science/Big Data Infrastructure Manager
		Research Infrastructure Managers	RI Manager
			RI Data storage facilities manager
Professionals			
	Science and engineering professionals	Data Science Professionals	Data Scientist
		Data professionals not elsewhere classified	Data Science Researcher
			(Big) Data Analyst
			Data Science (Application) Programmer
			Business Analyst
		Database and network professionals	Large scale (cloud) data storage designers and administrators
			Large scale (cloud) database designer*)
			Database designers and administrators
			Large scale (cloud) database administrator*)
			Database and network professionals not elsewhere classified
			Scientific database administrator*)
	Information and communications technology professionals	Data technology professionals	Digital Librarian
		Data professionals handling not elsewhere classified	Data Archivist
			Data Steward
			Data curator
Technicians and associate professionals			
	Science and engineering associate professionals	Data Technology Professionals	Big Data facilities Operators
		Data infrastructure engineers and technicians	Large scale (cloud) data storage operators
			Scientific database operator*)
		Database and network professionals not elsewhere classified	
Clerical support workers			
	General and keyboard clerks		Digital Librarian
	Data handling support workers (alternative)	Data and information entry and access	Digital Librarian
		Digital Archivists and Librarians	Data Archivist
			Data Steward
			Data curator

NIVELES

GRUPOS DE OCUPACIONES

OCUPACIONES

• Gestores

• Profesionales: ➤ Tratamiento de datos

➤ bibliotecario digital

➤ archiveros de datos

➤ Data Steward

➤ Data curator

• Profesionales asociados y técnicos

• Administrativos de soporte

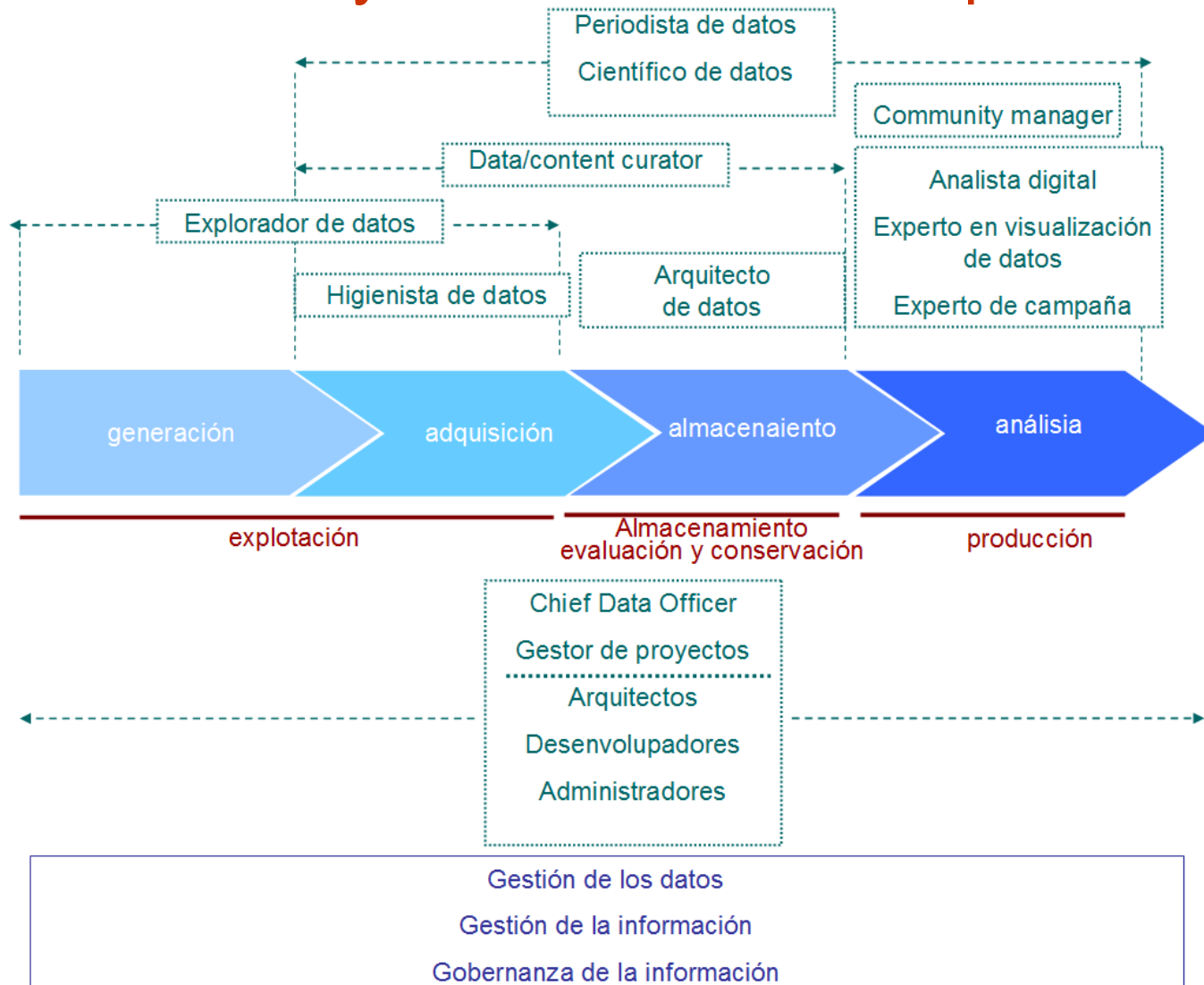
➤ Bibliotecario digital

➤ Archivero de datos

➤ *Data Steward*

➤ *Data curator*

4. Calidad y valor de los datos: perfiles





XV WORKSHOP REBIUN

DATOS Y BIBLIOTECAS · Universitat Jaume I (Castellón), 29-30 septiembre 2016



Preguntas

Debate

mgarciaals@uoc.edu