Diseño de un repositorio de datos de investigación agroalimentaria

Autores:
Jorge García Pérez y Antonio Jesús Sánchez Padial

INIA Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Redia

Jornadas técnicas de RedIris Noviembre de 2015

RedIRIS

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
INIA Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Presentación del INIA

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
INIA
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

OPI – MINECO Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Responsable de la **gestión y coordinación** de la investigación en materia de **I+D+i agroalimentaria** en el ámbito estatal, así como de la ejecución de proyectos de **investigación**, colaboración con los correspondientes sectores socioeconómicos.

INIA 1929-1936

1926

(INIEAF)

Instituto Nacional de Investigaciones y Experiencias Agronómicas y Forestales

1929

(IFIE)

Instituto Agrícola de Alfonso XII en la Ciudad Universitaria

+

1931

Laboratorio de Fauna Forestal de Piscicultura y Ornitología.

1932

(INIA)

Instituto Nacional de Investigaciones Agronómicas

1931

Dirección General de Ganadería y el Instituto de Biología Animal

Agronomía
Forestal
Pecuaria

+

Piscicultura
Ornitología

+

Fisiozootecnia
Patología animal
Contrastación

INIA-1936-2015

INIA
Nueva ubicación en el Monte del Pardo

+

1952

Patronato de Biología Animal

+

EE TT SS de Ingenieros Agrónomos UPM

+

Facultad de Veterinaria de la UCM

CCAA

1991

OPI
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

INIA en la actualidad

Nacional

Control y Verificación de las ayudas

Ente Financiador

Cofinanciación Comunitaria

Producción Científica

Colaboración Público Privada

Centros de Investigación

Patentes

INIA
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Representación en los organismos Internacionales del sector

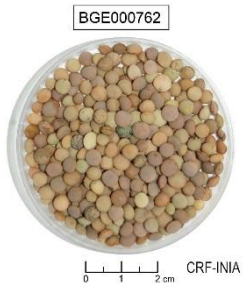
A las AAPP

Servicios

Coordinación Investigación Nacional

Otros centros

Publicaciones



Centros de Investigación del INIA

CISA Centro Investigación en Sanidad Animal

Epidemiología,
Vigilancia enfermedades emergentes
Inmunología y Vacunas
Laboratorio seguridad Biológica niveles: 3 y 3+

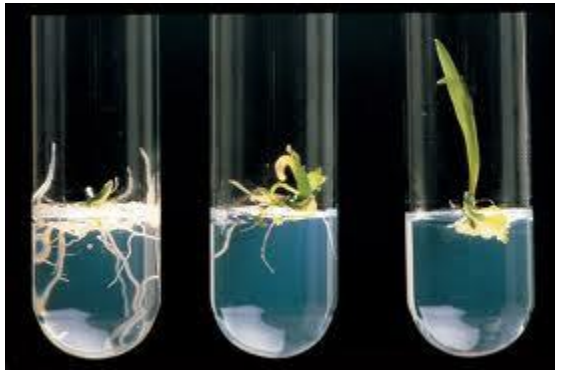
CRF Centro de Recursos Fitogenéticos

Investigación aplicada a la conservación y
Utilización de Los Recursos Fitogenéticos para la
Agricultura y la Alimentación
Inventario Nacional de Recursos Fitogenéticos

CIFOR Centro De Investigación Forestal

Cultivos y Plantaciones Forestales
Dinámica y Funcionamiento de ecosistemas forestales
Protección Forestal
Productos Forestales

Departamentos



BIOTECNOLOGÍA

MEDIO AMBIENTE

MEJORA GENÉTICA ANIMAL

PROTECCIÓN VEGETAL

REPRODUCCIÓN ANIMAL

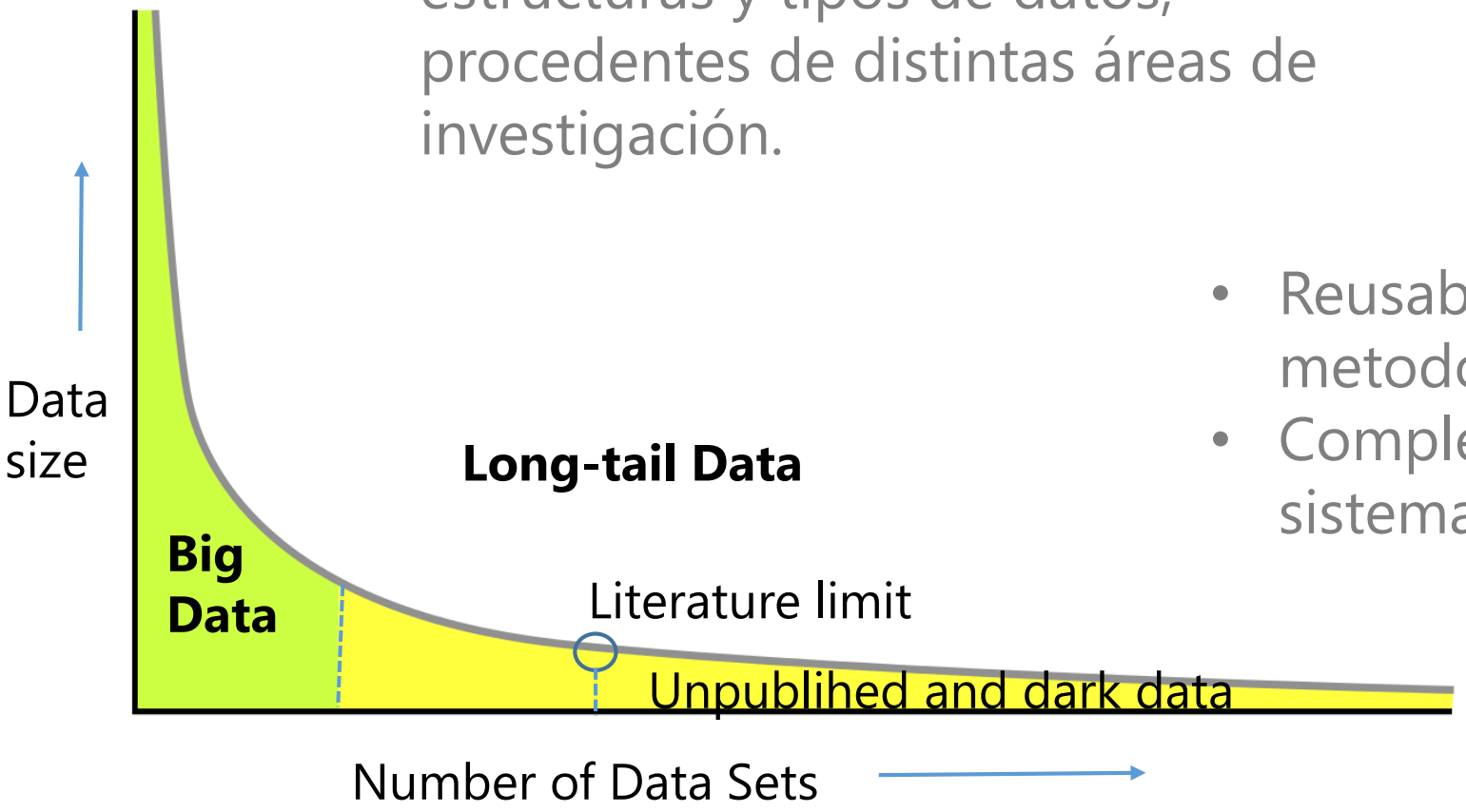
CENTRO CALIDAD ALIMENTOS SORIA

DIRECCIÓN TÉCNICA DE EVALUACIÓN DE VARIEDADES Y PRODUCTOS FITOSANITARIOS (DTEVPF)

Long Tail Data en lugar de Big Data

INIA-> Multitud de pequeñas colecciones de datos con diferentes estructuras y tipos de datos, procedentes de distintas áreas de investigación.

Long Tail Data



- Reusabilidad interdisciplinar (datos y metodología).
- Complejidad en los análisis y los sistemas de gestión.

¿Datos de Investigación?

***son hechos**, **observaciones** o **experiencias** en que se basa el **argumento**, la **teoría** o la **prueba**.*

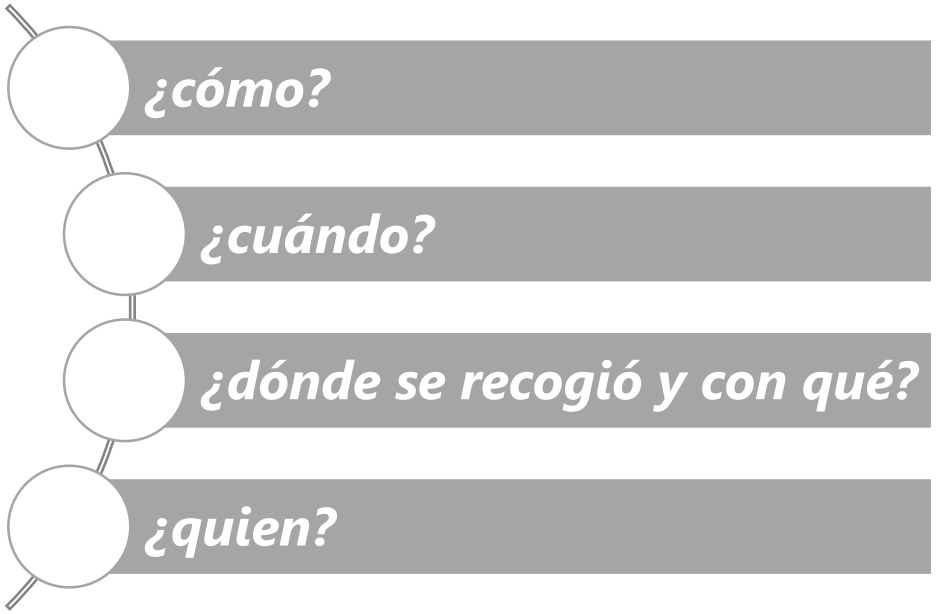
Documento de política de gestión de datos de la Universidad de Melbourne.

Metadatos

El término **metadato**

- datos sobre datos
- informaciones sobre datos
- datos sobre informaciones
- informaciones sobre informaciones

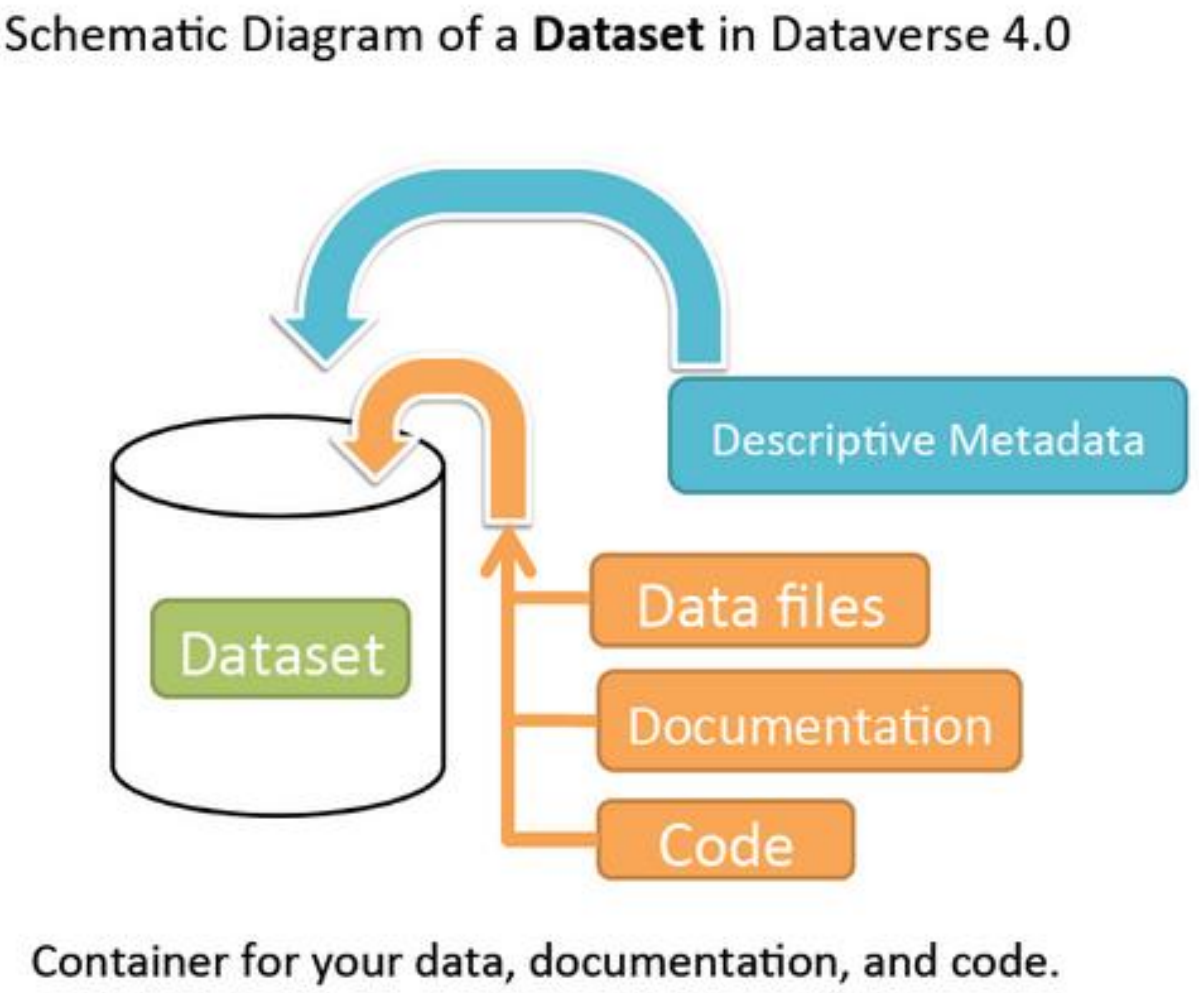
Datos estructurados y codificados que **describen características** de instancias conteniendo **informaciones** para ayudar a **identificar**, **descubrir**, **valorar** y **administrar** las instancias descritas.[8]



W. R. Durrell. **Data Administration. A Practical Guide to Data Administration.** McGraw-Hill, 1985

Dataset - Conjunto de Datos

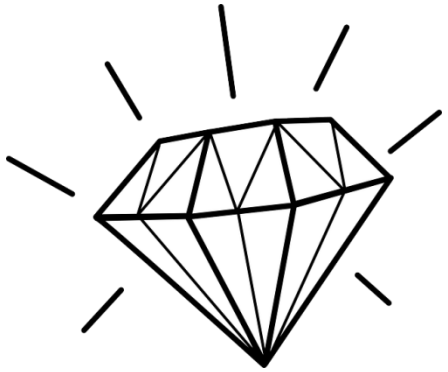
Un Dataset o Conjunto de Datos puede componerse de varios archivos y sus metadatos asociados correspondientes



¿Por qué un repositorio de datos de investigación? I

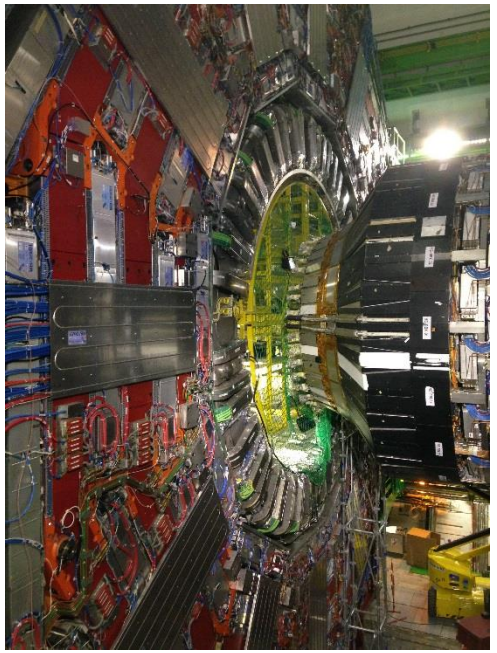
Los **datos** de investigación son un recurso

costoso, **valioso** **y** **volátil**.

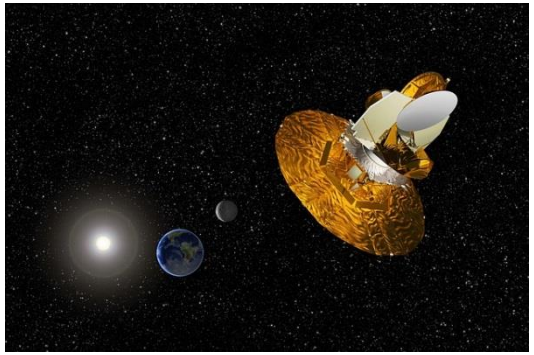


Obtener datos es costoso

Una gran parte de los recursos de investigación se dedican a la preparación y adquisición de datos de investigación.
(Pero muy pocos a la clasificación y conservación de los datos)



La obtención de datos conlleva:



- equipamiento
- personal
- espacio
- desplazamientos
- tiempo
- conocimiento (Know How)
- …

Ciclo vida de los datos

https://okfn.org/

El valor de los datos

Datos como un recurso Valioso, la robustez, consistencia, precisión, completitud…

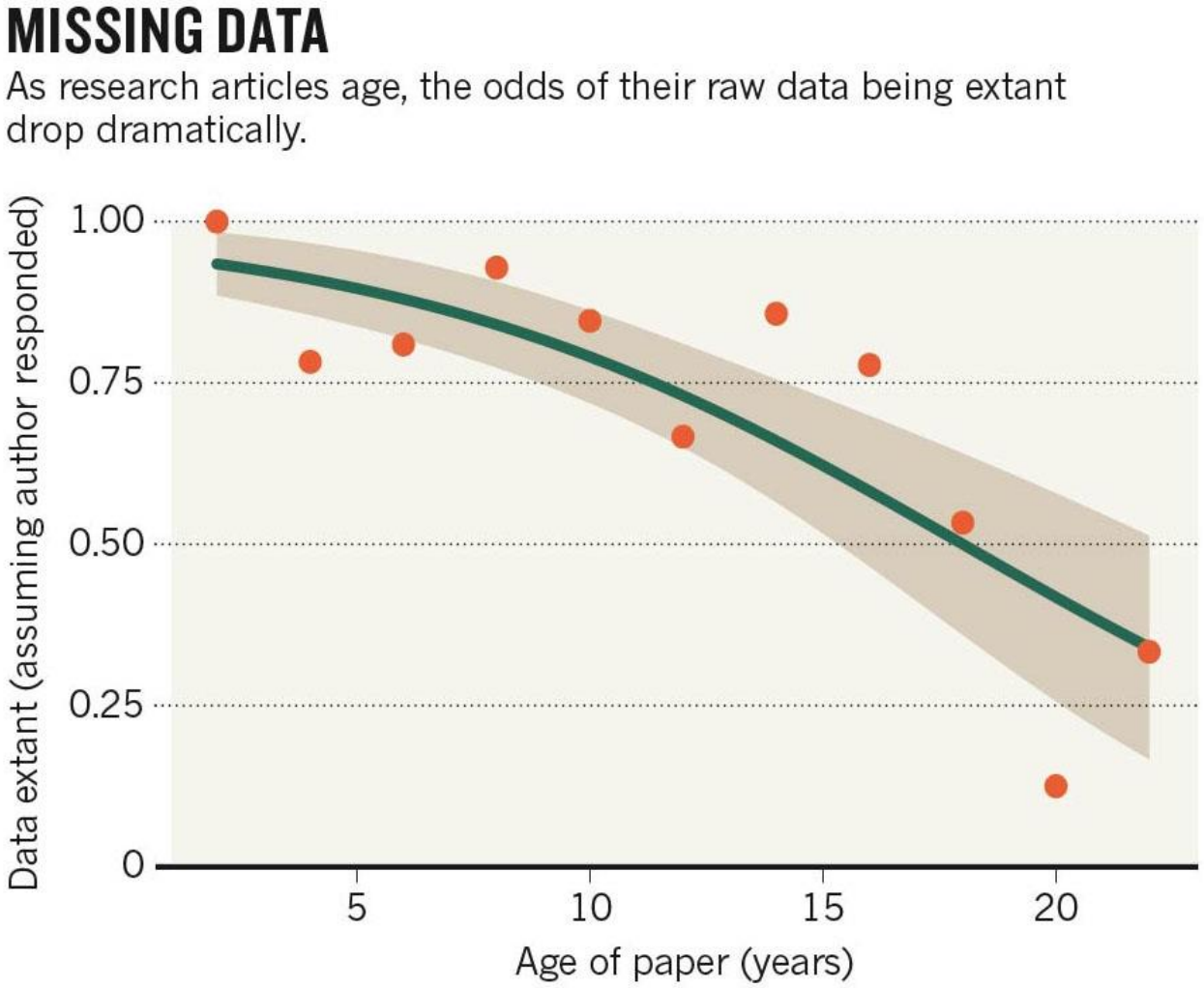
Calidad de los Datos › Calidad de la Investigación

Irreversibilidad de la toma de datos (por ejemplo datos de un cuestionario a un paciente).

Condiciones de contorno datos asociados METADATOS.

Pérdida de datos.

The availability of research data declines rapidly with article age. *Vines et al*



Estudio de 2013 sobre 516 artículos publicados con antigüedad de 2 a 22 años.

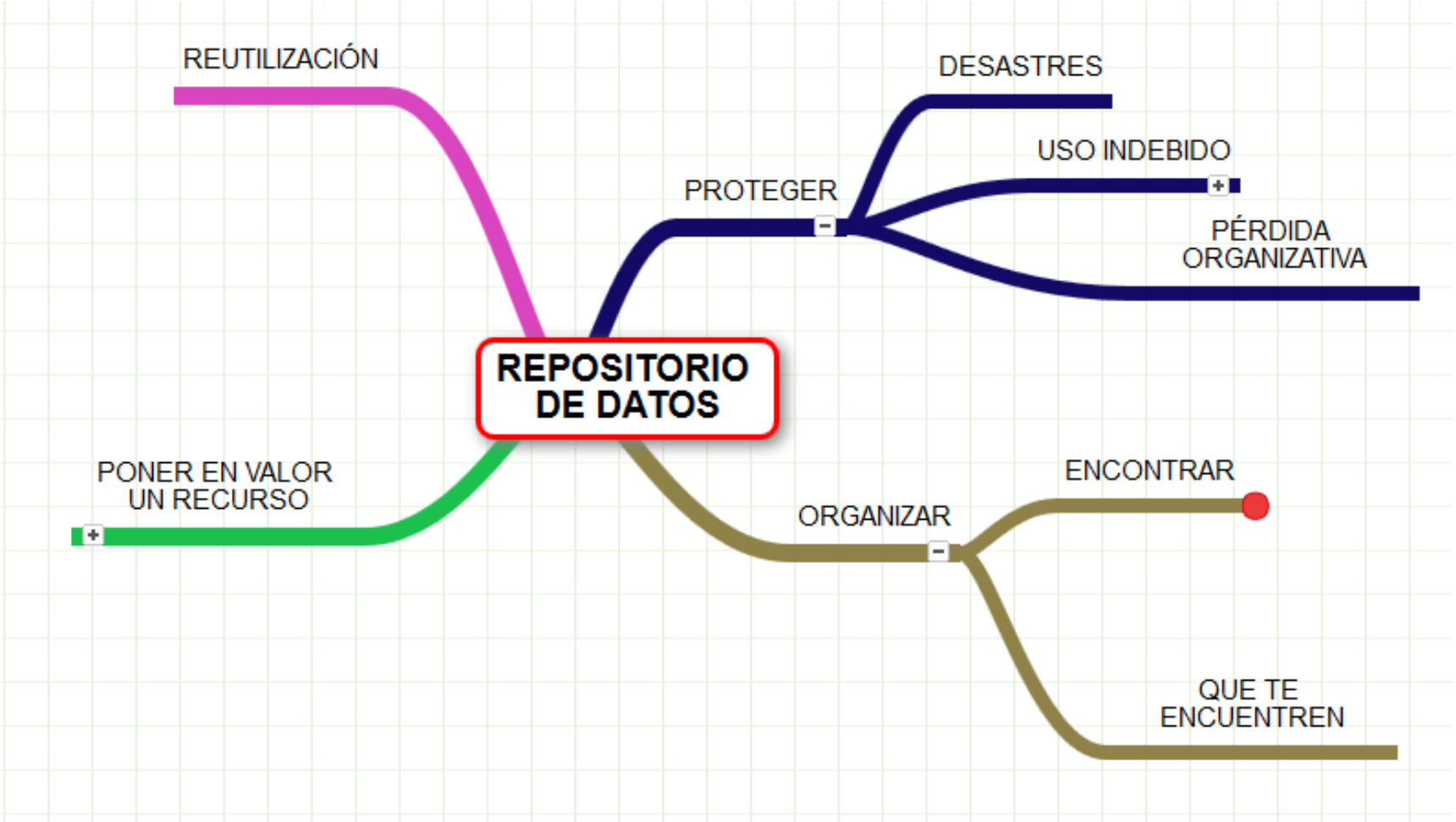
Mediante una consulta a los autores sobre los datos de los artículos.

Resultados:
bastante lineales con pérdidas anuales de datos del 7% al 17%.

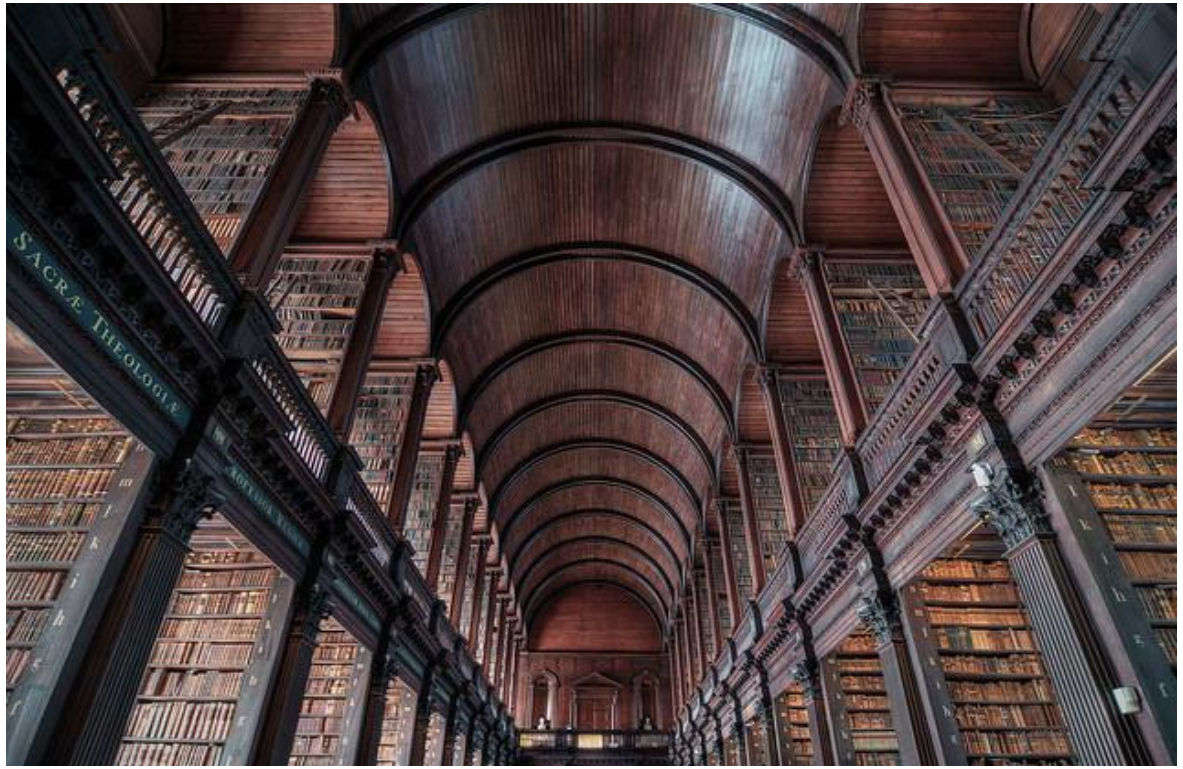
http://arxiv.org/abs/1312.5670

¿Por qué un repositorio de datos de investigación? II

Preservación, Clasificación, Accesibilidad y Control



Preservación



Clasificación

Engineering Plans Storage, 2001. *Seattle Municipal Archive*
https://flic.kr/p/6Jvi25

https://www.flickr.com/photos/proforged/with/4634075958/
brent flanders
Ernst and Son Hardware
cool old hardware store downtown Lawrence, KS

Proyecto Repositorio INIA

Servicio a los Investigadores del centro

Servicio a la comunidad

Recolección de datos

Fase 1

Fase 2

Fase 3

Proyecto Repositorio INIA

Redia

Repositorio Electrónico de Datos de Investigación Agroalimentarios

Comisiones

Comisión Científica

Requerimientos

Legales

- Licencias
- Permisos
- Propiedad
- Embargos

Organizativos

- PGD plan de gestión de datos
- Relaciones con otros centros
- Incentivos
- Divulgación entre los investigadores del INIA

Científicos

- Áreas de conocimiento
- Vocabularios controlados
- Funcionalidades de apoyo a la investigación

Comisión Técnica

- **Estado del arte**
  - **Contactos con expertos**
  - **Formaciones.**
  - **Wiki Interna.**
  - **RDM.Agripa.org**
- **Estimación de tipos y volúmenes de datos.**
  - **Censo de datos de Investigación**
  - **Adquisición de recursos de almacenamiento**
- **Evaluación de plataformas.**
  - **Pilotos en Plataformas.**
  - **Test de evaluación de requerimientos.**

Contactos

datos.gob.es
reutiliza la información pública

Instituto de Salud Carlos III

Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

cita
CENTRO DE INVESTIGACIÓN Y TECNOLOGÍA AGROALIMENTARIA DE ARAGÓN

FECYT
FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA

PAGODA

Organización de las Naciones Unidas para la Alimentación y la Agricultura

FUNDACIÓN JUAN MARCH

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Universidad de Alcalá

Seminarios

Open Data Support Madrid (Junio 2015)

red.es

Webinars Sep 2015
RIOXX: the pragmatic development of a metadata application profile

The Haye Nov 2015
2nd Workshop Creating Impact with Open Data in Agriculture and Nutrition

Iniciativas de datos agrarios

CIARD
Open Agricultural Knowledge for Development

SemaGrow

GODAN
Global Open Data for Agriculture and Nutrition

Estimación de datos

Estimación cuantitativa

Recursos de almacenamiento adquiridos en INIA en los últimos años

Estimación de recursos de almacenamiento necesario para los próximos años

Estimación cualitativa

Censo de Datos de Investigación

- **Número**
- **Tipo**
- **Volumen**
- **Formato**
- **Área de investigación**

Selección de las plataformas

The Dataverse Project

HARVARD UNIVERSITY

(IQSS) Institute for Quantitative Social Science
y
(HUIT) Harvard University Information Technology

Características técnicas:

Java

GlassFish

PostgreSQL

www.dateverse.org

Características funcionales

- Potente, tablas de hasta 2GB archivos hasta 10 GB
- Muy orientada a datos.
- Presentación cuidada.
- Poco extensible (no tiene un sistemas de extensiones)
- Muy utilizada mediante hosting en el repositorio de Harvard.
- Herramientas de visualización compatibles, TwoRavens, Geo.

Interfaz

The Dataverse Project

Files | Metadata | License + Terms | Versions

Download

ug-cdbk-ques-drone.pdf
Adobe PDF - 2.6 MB - Mar 18, 2015
MD5: b491ccbf5d83a5be2ba46e5264358b51;
User Guide, Codebook
Documentation

Download

Drones-UK.tab
Tabular Data - 230.6 KB - Mar 18, 2015
Original File MD5: 8f7dec1f28a2e9ab9486a528838ce275;
87 Variables, 2008 Observations - UNF:6:9fTztQZCgAewmZTs1NeZ8Q==
UK dataset
Data

Drones-Canada.tab
Tabular Data - 357.1 KB - Mar 18, 2015
Original File MD5: 5b1d93418ecaf5c39c8003fce1831eb0;
85 Variables, 3045 Observations - UNF:6:9AJGZe9Cxf71CpJO8cjGSg==
Canada dataset
Data

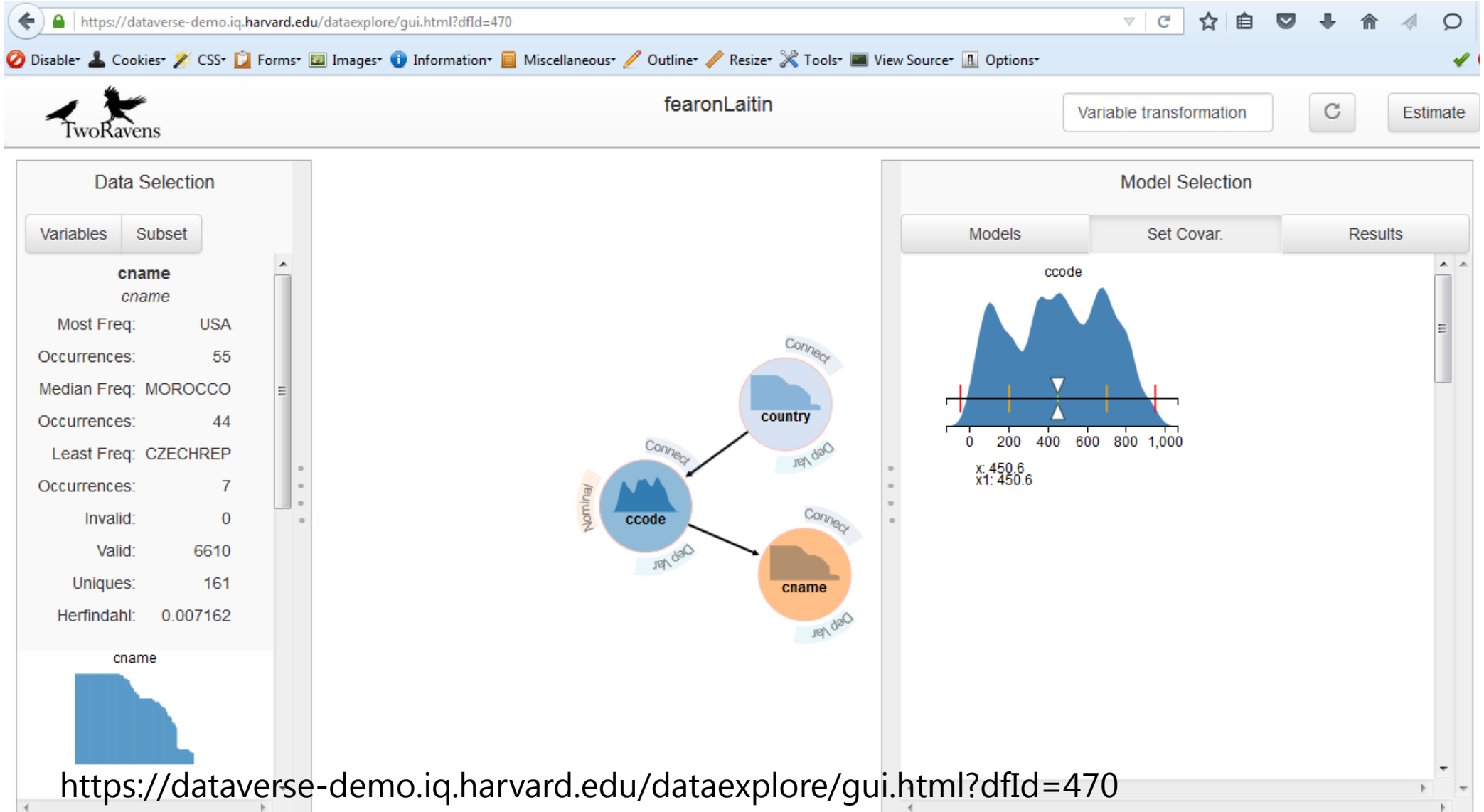
Drone -US(revised).tab
Tabular Data - 366.5 KB - Mar 18, 2015
Original File MD5: 7c042a7631d29343ca332f8c81637fd3;

Explore | Download

All File Formats + Information

Original File Format (SPSS SAV)
Tab-Delimited
RData Format
Variable Metadata
Data Subset
Data File Citation

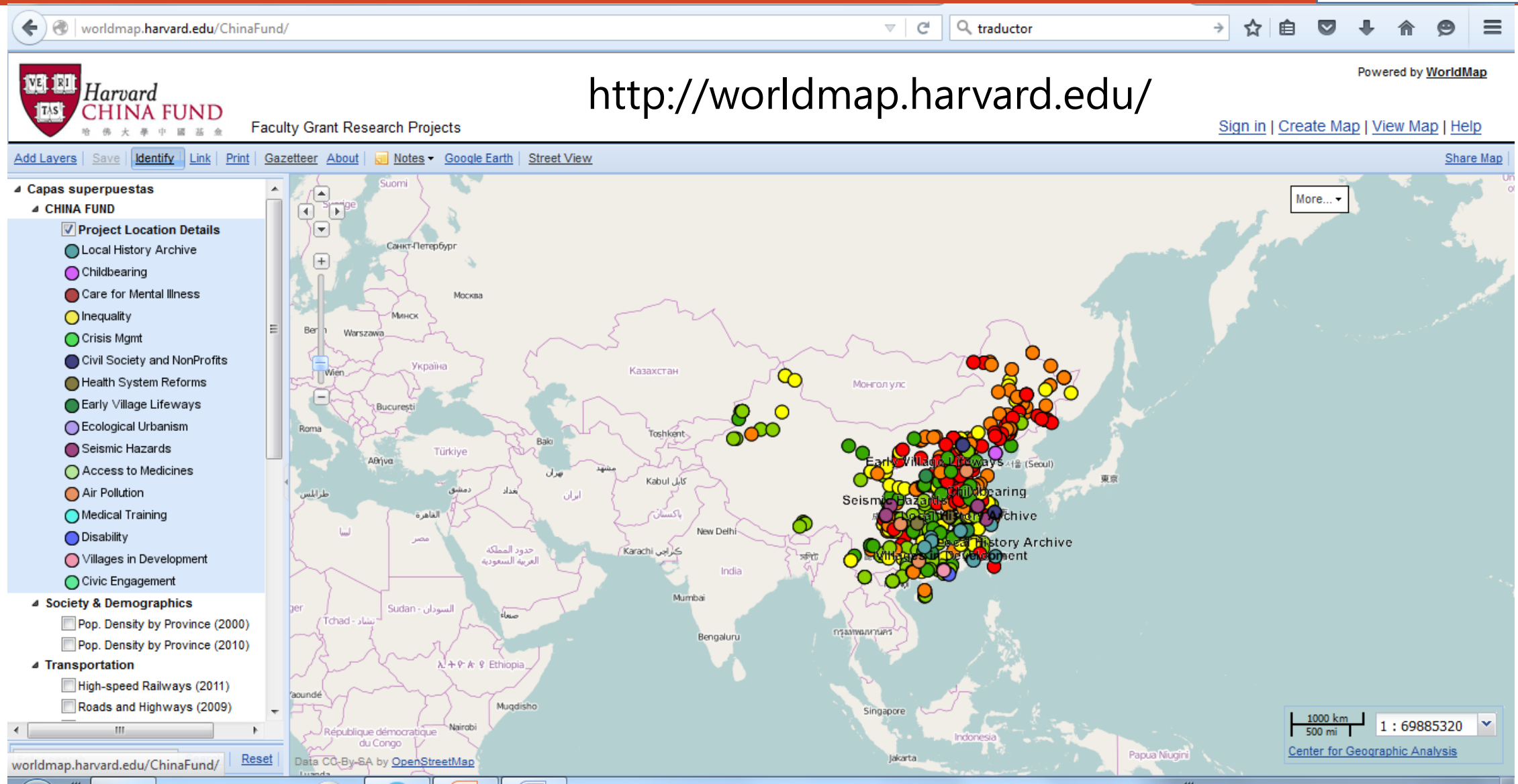
Herramienta de analisis y visualización estadístico integrado con Dataverse



https://dataverse-demo.iq.harvard.edu/dataexplore/gui.html?dfId=470

Visualización de datos geo-referenciados
Harvard's WorldMap platform

The Dataverse Project



Ckan

ckan The open source data portal software

Open Knowledge Foundation (OKF)

- Orientado a datos
- Muy extensible
- Una comunidad desarrollo muy activa
- Extensiones para visualización de datos.
- Muy utilizada en portales de datos gubernamentales.

Características técnicas:

Python and Javascript

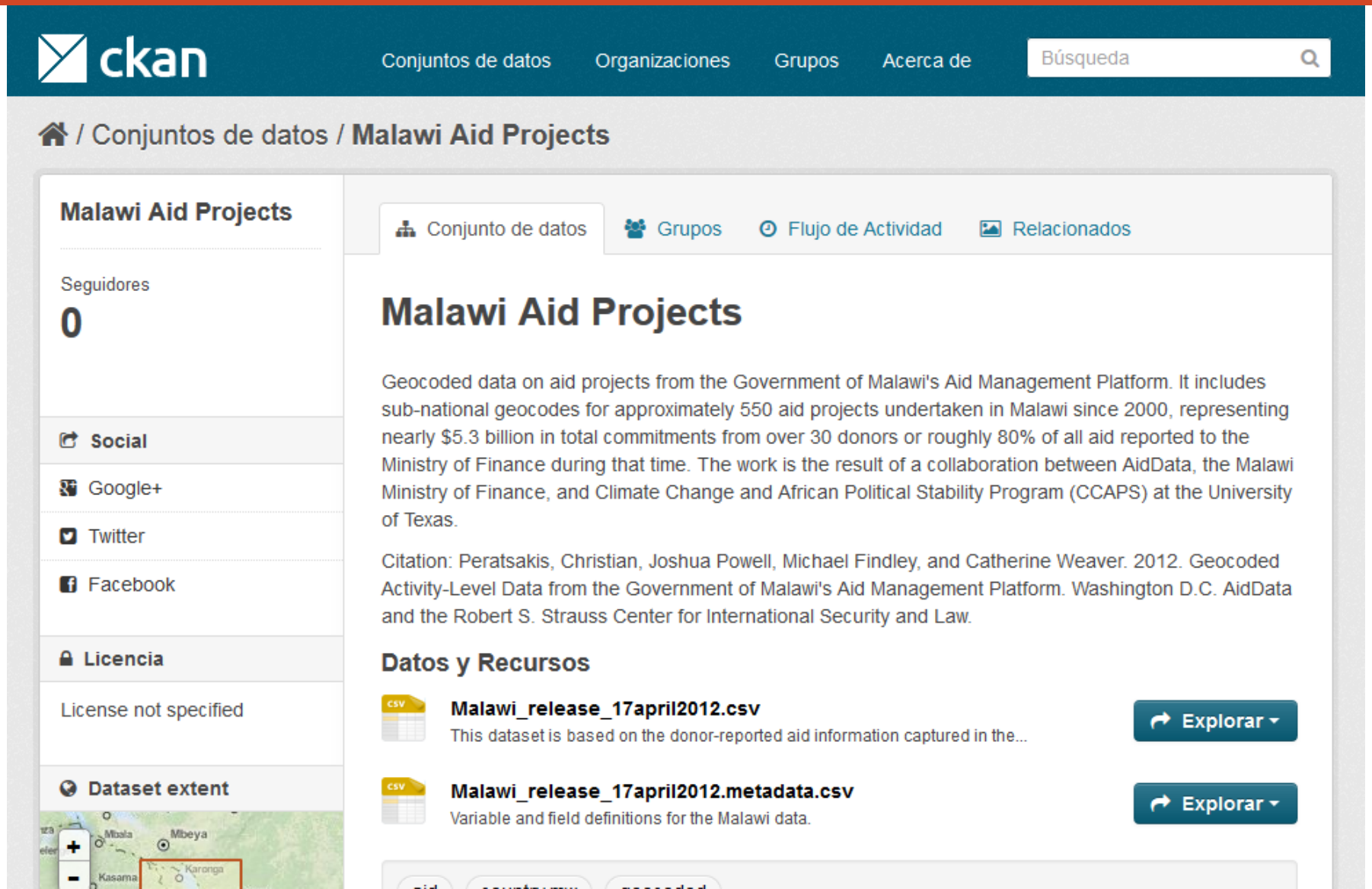
SQLAlchemy

PostgreSQL

SOLR

data.gov.uk

Interfaz I


ckan
Conjuntos de datos
Organizaciones
Grupos
Acerca de
Búsqueda
/ Conjuntos de datos / Malawi Aid Projects
Malawi Aid Projects
Seguidores
0
Social
Google+
Twitter
Facebook
Licencia
License not specified
Dataset extent
Mbala
Mbeya
Kasama
Karonga
Conjunto de datos
Grupos
Flujo de Actividad
Relacionados
Malawi Aid Projects
Geocoded data on aid projects from the Government of Malawi's Aid Management Platform. It includes sub-national geocodes for approximately 550 aid projects undertaken in Malawi since 2000, representing nearly $5.3 billion in total commitments from over 30 donors or roughly 80% of all aid reported to the Ministry of Finance during that time. The work is the result of a collaboration between AidData, the Malawi Ministry of Finance, and Climate Change and African Political Stability Program (CCAPS) at the University of Texas.
Citation: Peratsakis, Christian, Joshua Powell, Michael Findley, and Catherine Weaver. 2012. Geocoded Activity-Level Data from the Government of Malawi's Aid Management Platform. Washington D.C. AidData and the Robert S. Strauss Center for International Security and Law.
Datos y Recursos
Malawi_release_17april2012.csv
This dataset is based on the donor-reported aid information captured in the...
Explorar
Malawi_release_17april2012.metadata.csv
Variable and field definitions for the Malawi data.
Explorar

Interfaz II

ckan

Twitter

Facebook

Licencia

License not specified

Dataset extent

Map data © OpenStreetMap contributors

Tiles by MapQuest

of Texas.

Citation: Peratsakis, Christian, Joshua Powell, Michael Findley, and Catherine Weaver. 2012. Geocoded Activity-Level Data from the Government of Malawi's Aid Management Platform. Washington D.C. AidData and the Robert S. Strauss Center for International Security and Law.

Datos y Recursos

Malawi_release_17april2012.csv
This dataset is based on the donor-reported aid information captured in the...
Explorar

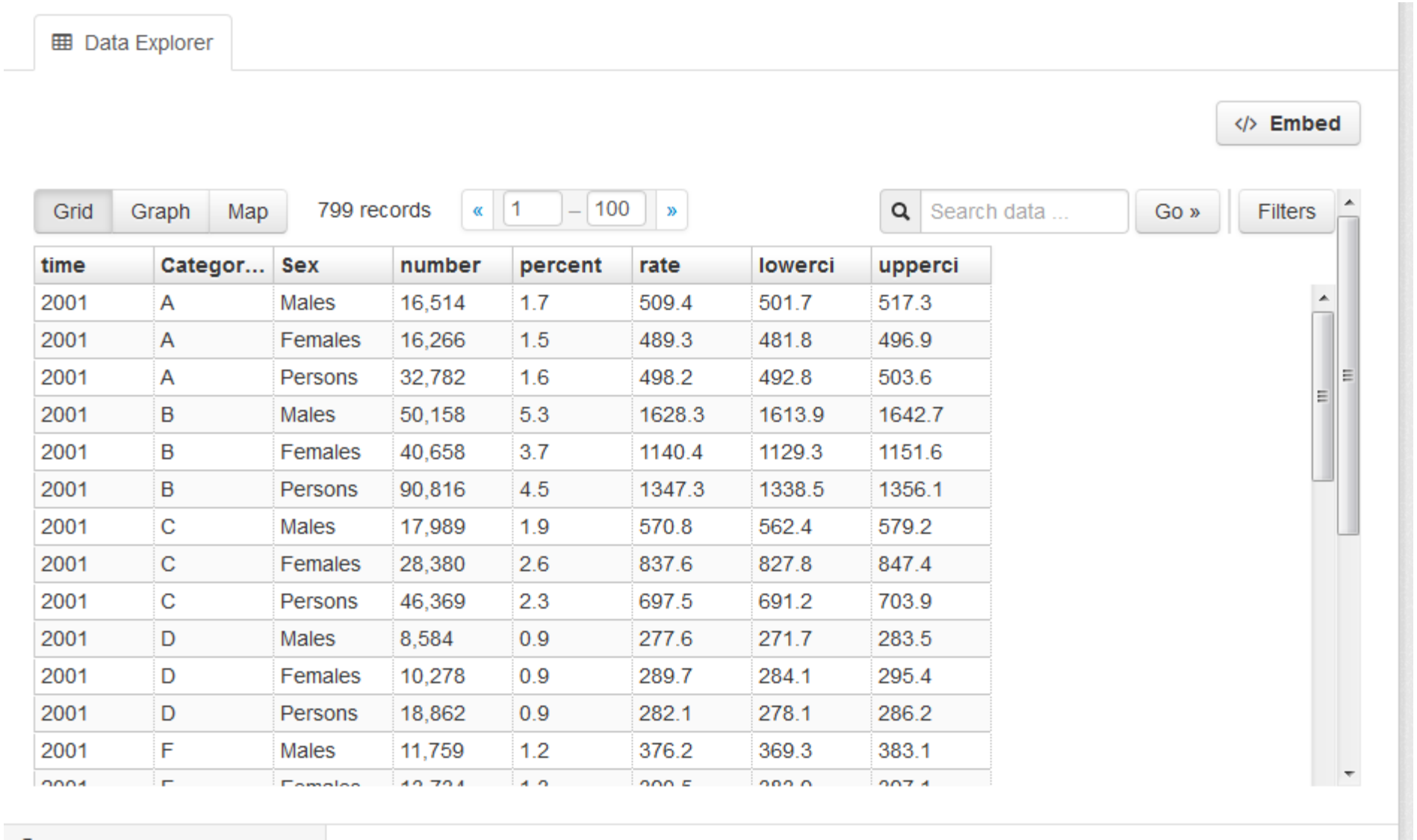
Malawi_release_17april2012.metadata.csv
Variable and field definitions for the Malawi data.
Explorar

aid | country.mw | geocoded

Información Adicional

Campo	Valor
Fuente	http://blog.aiddata.org/2012/04/where-are-donors-working-in-malawi-new.html
Última actualización	27 Julio, 2015, 17:23
Creado	22 Julio, 2015, 14:30
spatial	{ "type": "Polygon", "coordinates": [[[32.689701, -17.135811],[32.689701, -9.373335], [35.92416, -9.373335], [35.92416, -17.135811], [32.689701, -17.135811]]] }

Tabular Data



time	Categor...	Sex	number	percent	rate	lowerci	upperci
2001	A	Males	16,514	1.7	509.4	501.7	517.3
2001	A	Females	16,266	1.5	489.3	481.8	496.9
2001	A	Persons	32,782	1.6	498.2	492.8	503.6
2001	B	Males	50,158	5.3	1628.3	1613.9	1642.7
2001	B	Females	40,658	3.7	1140.4	1129.3	1151.6
2001	B	Persons	90,816	4.5	1347.3	1338.5	1356.1
2001	C	Males	17,989	1.9	570.8	562.4	579.2
2001	C	Females	28,380	2.6	837.6	827.8	847.4
2001	C	Persons	46,369	2.3	697.5	691.2	703.9
2001	D	Males	8,584	0.9	277.6	271.7	283.5
2001	D	Females	10,278	0.9	289.7	284.1	295.4
2001	D	Persons	18,862	0.9	282.1	278.1	286.2
2001	F	Males	11,759	1.2	376.2	369.3	383.1

Gráficos

US National Foreclosure Statistics - By State - January 2012

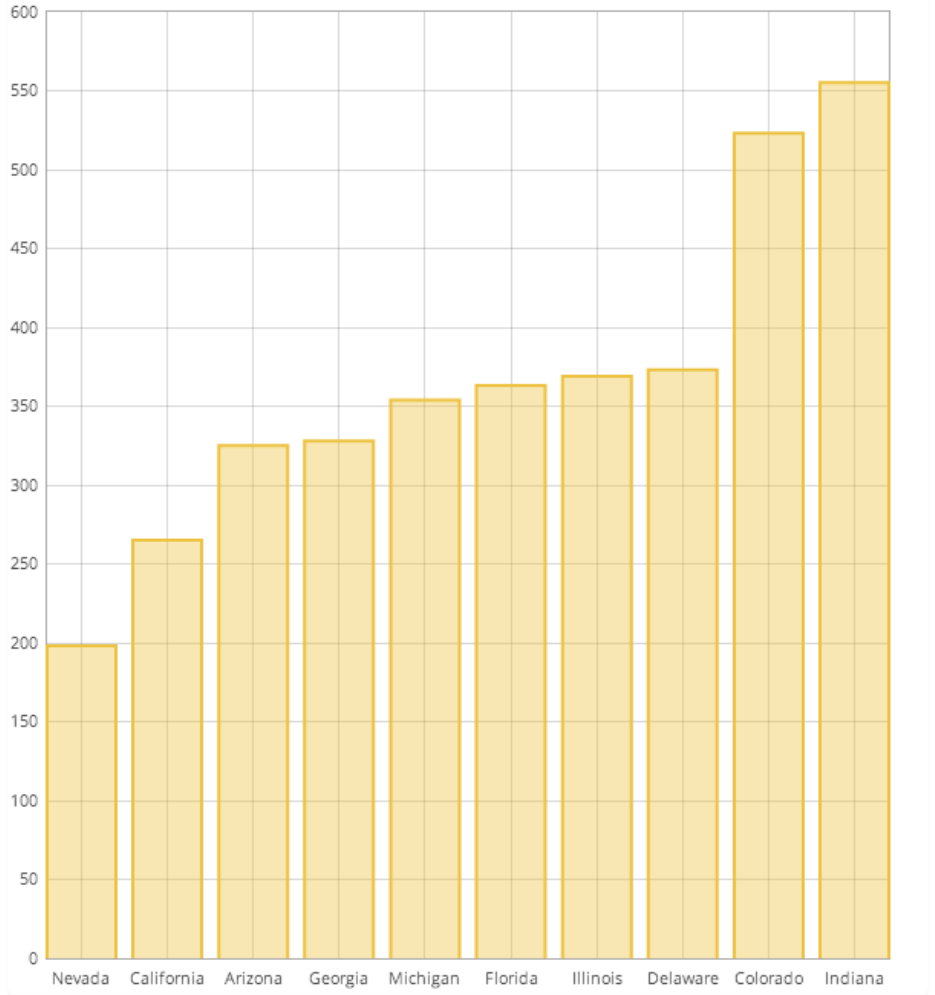
us_foreclosures_jan_2012_by_state.csv

Embed

Grid | Graph | Map

51 records « 1 - 10 »

Search data ... Go » Filters Fields



Graph Type
Columns

Group Column (Axis 1)
state

Series A (Axis 2) [Remove]
foreclosure.ratio

Add Series

School of Electronics and Computer Science University of Southampton V 1.0

Nov 2.000

Enfocado a la creación de repositorios institucionales no enfocados específicamente a datos.

Muy configurable y abierto.

http://www.eprints.org

Características Técnicas LAMP:

Apache MySQL Perl

Interfaz

Effects of four tanniferous plant extracts on the in vitro exsheathment of third-stage larvae of parasitic nematodes

D., BAHUAUD; C., MARTINEZ-ORTIZ DE MONTELLANO; S., CHAUVEAU; F, PREVOT; F., TORRES-ACOSTA; I., FOURASTE and H., HOSTE (2005) Effects of four tanniferous plant extracts on the in vitro exsheathment of third-stage larvae of parasitic nematodes. *Parasitology*.

PDF
Limited to [Depositor and staff only]
243Kb

Summary

The anthelmintic properties of tanniferous plants and of their secondary metabolites represent one possible alternative to chemotherapy that is currently being explored as a means of achieving sustainable control of gastrointestinal nematodes in ruminants. Previous in vivo and in vitro results suggest that tanniferous plants can have direct anti-parasitic effect against different stages of nematodes. However, the mode of action of the bioactive plant compounds remains obscure. The objectives of the current study were (1) to examine the hypothesis that extracts of tanniferous plants might interfere with the exsheathment of third-stage infective larvae (L3); (2) to assess the role of tannins in the process by examining the consequence of adding an inhibitor of tannins (polyethylene glycol: PEG) to extracts. The effects of 4 tanniferous plant extracts on exsheathment have been examined on L3 of Haemonchus contortus and Trichostrongylus colubriformis. Artificial exsheathment was induced in vitro by adding hypochlorite solution to larval suspension. The evolution of exsheathment with time was measured by repeated observations at 10-min interval for 60 min. The selected plants were: genista (Sarothamnus scoparius), heather (Erica erigena), pine tree (Pinus sylvestris), and chestnut tree (Castanea sativa), with tannin contents ranging from 1.5 to 24.7% of DM. Extracts of a non-tanniferous plant (rye grass, tannin content: 0.3% of DM) were included in the assay as negative controls. The extracts were tested at the concentration of 600 mg/ml and the effects were compared to the rate of exsheathment of control larvae in PBS. No statistical differences in the pattern of exsheathment was observed after addition of rye grass or genista extracts for both nematode species and with heather extracts for T. colubriformis. In contrast, pine tree extracts on larvae of both species and heather extracts with H. contortus induced a significant delay in exsheathment. Last, contact with chest nut extracts led to a total inhibition of the process for both nematodes. These results suggest that extracts of tanniferous plants might affect a key process in the very early stages of larval invasion of the host. In most cases, the addition of PEG led to a total or partial restoration towards control values. This suggests that tannins are largely involved in the inhibitory process. However, other secondary metabolites may also interfere with the process that would help to explain some of the differences in response observed between the two nematode species.

EPrint Type:	Journal paper
Keywords:	parasitic nematodes, exsheathment, third-stage larvae, tanniferous plant, PEG
Subjects:	Animal husbandry > Production systems > Sheep and goats Animal husbandry > Health and welfare
Research affiliation:	France > INRA - Institut National de la Recherche Agronomique

Test de requerimientos
Data Management System INIA I

1-Permisos (acceso a documentos, tipos de Usuarios LDAP)

2-Embargos

3-Backups (facilidad de gestión y recuperación)

4-Logs (Historial de modificaciones del documento, Accesos, popularidad, frecuencia)

5-Taxonomías Vocabularios (los que incorpora, facilidad para incorporar nuevos)

6- Interoperabilidad combatividad con OAI/PMH otros repositorios y recolectores

7- Accesibilidad (legislación, URLs únicas perdurables)

Test de requerimientos
Data Management System INIA II

8-Licencias (Gestión de licencias personalizada, obras relacionadas)

9-Documentos (Gestión flexible de documentos, control de versiones)

10-Citas (Generador configurable de citas. Cita de elementos individuales)

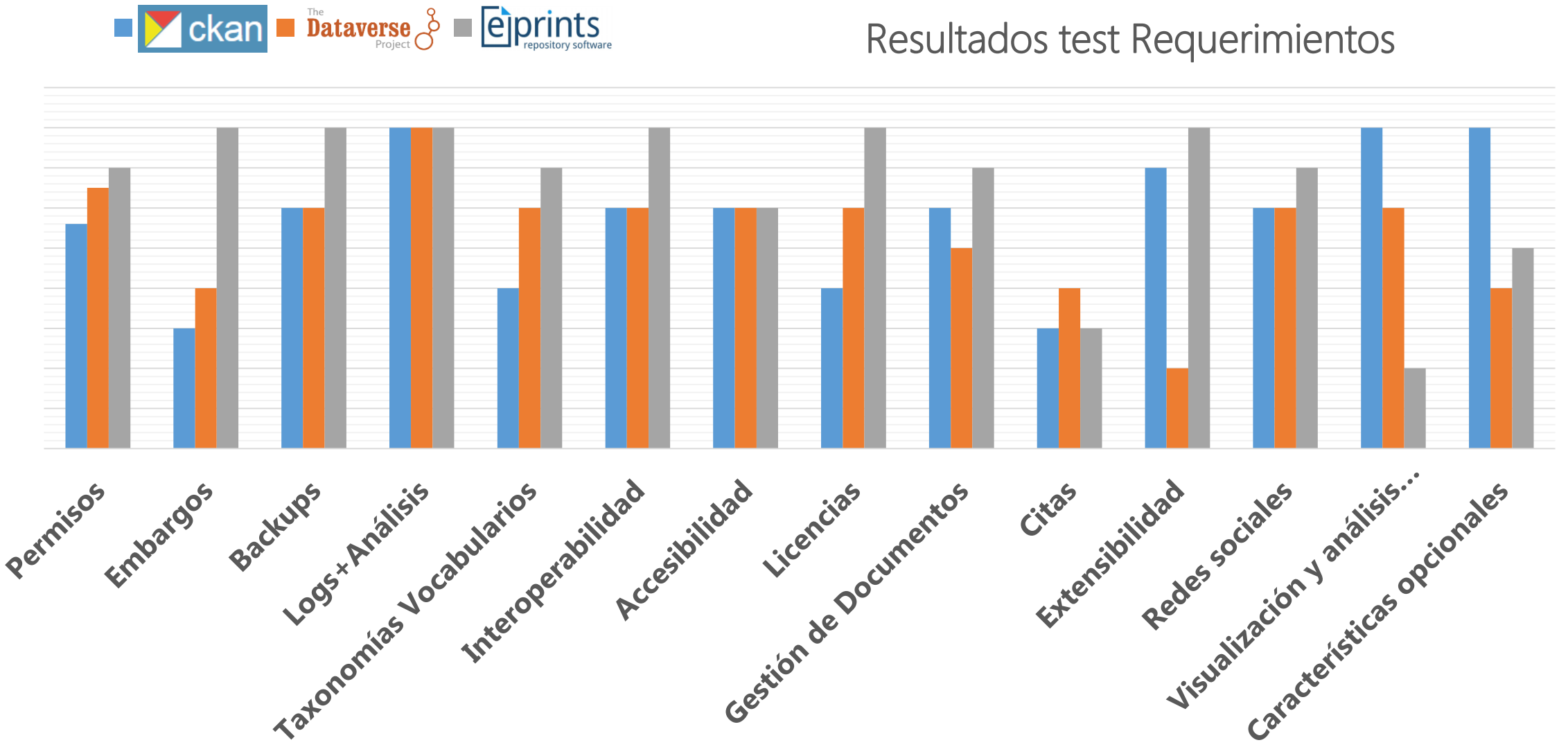
11-Extensibilidad

12-Redes sociales

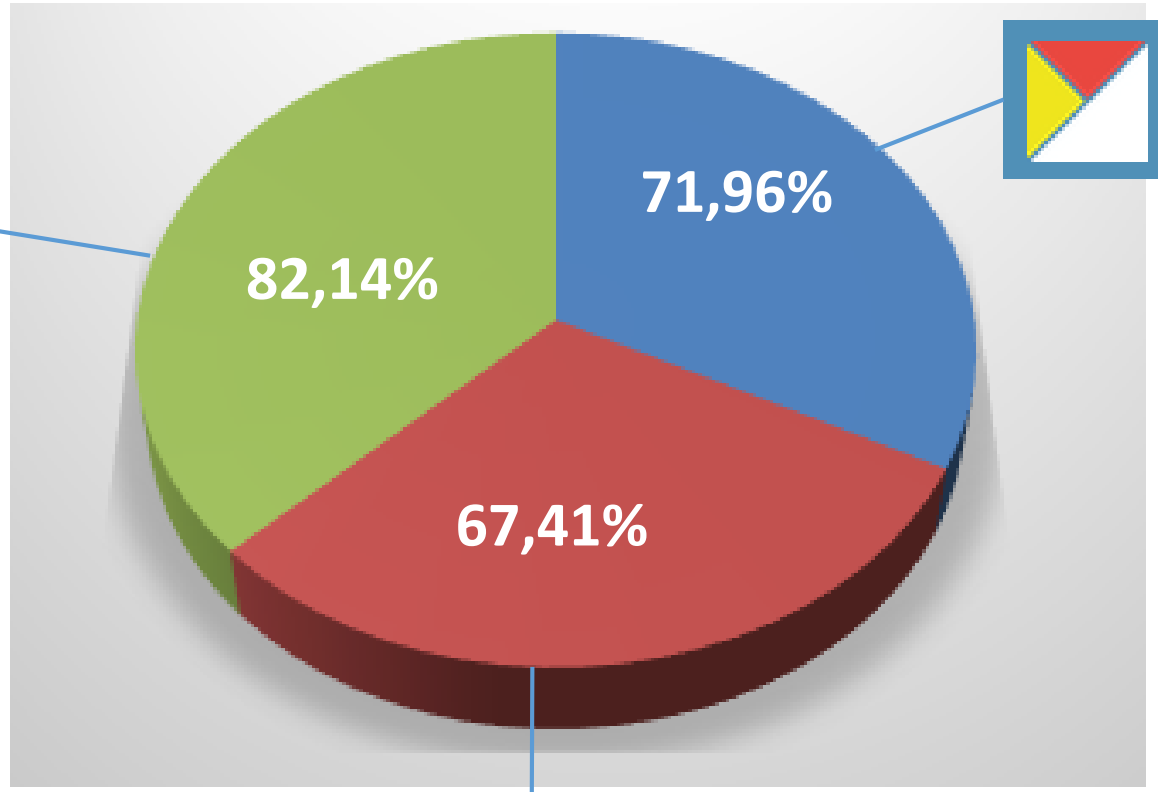
13- Visualización y análisis de datos

14-Características opcionales

Resultados del test I



Resultados del test II



Gracias por su atención

Antonio Jesús Sánchez Padial
Jefe del Servicio de Biometría.
antonio.sanchez@inia.es INIA

Jorge García Pérez
Jefe Sección del Área de Informática.
jorge.garcia@inia.es INIA

INIA
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Redia
Repositorio Electrónico de Datos de Investigación Agroalimentarios