Acceso enriquecido a la información: metadatos e interoperabilidad en el contexto de la ciencia abierta

Citar como: Peset, F ; Subirats, I. (2022). Acceso enriquecido a la información: metadatos e interoperabilidad en el contexto de la ciencia abierta. En 2º Hatun Tinkuy Perú, 17 y 18/8/2022 https://hatuntinkuy.concytec.gob.pe/inicio

CONCYTEC Hatun Tinkuy

DATAUSE

Materiales **procedentes de las presentaciones mostradas durante los últimos AGRIS y AGROVOC meetings 2022**, por el grupo de la FAO sobre Normas para la gestión de la información agraria AIMS

http://aims.fao.org/es

Comunicación científica	Acceso abierto	Datos enlazados	Datos de Investigación	Open Data	Measuring data	Blockchain Science/data	OD AGRI-TRANS	SDG DATAUSE	Reusar Analizar	Linkar Explotar datos

Reuse of open data in agriculture

Preservación y difusión digital del patrimonio científico y cultural HUM2004-02722/HIST

Premisas

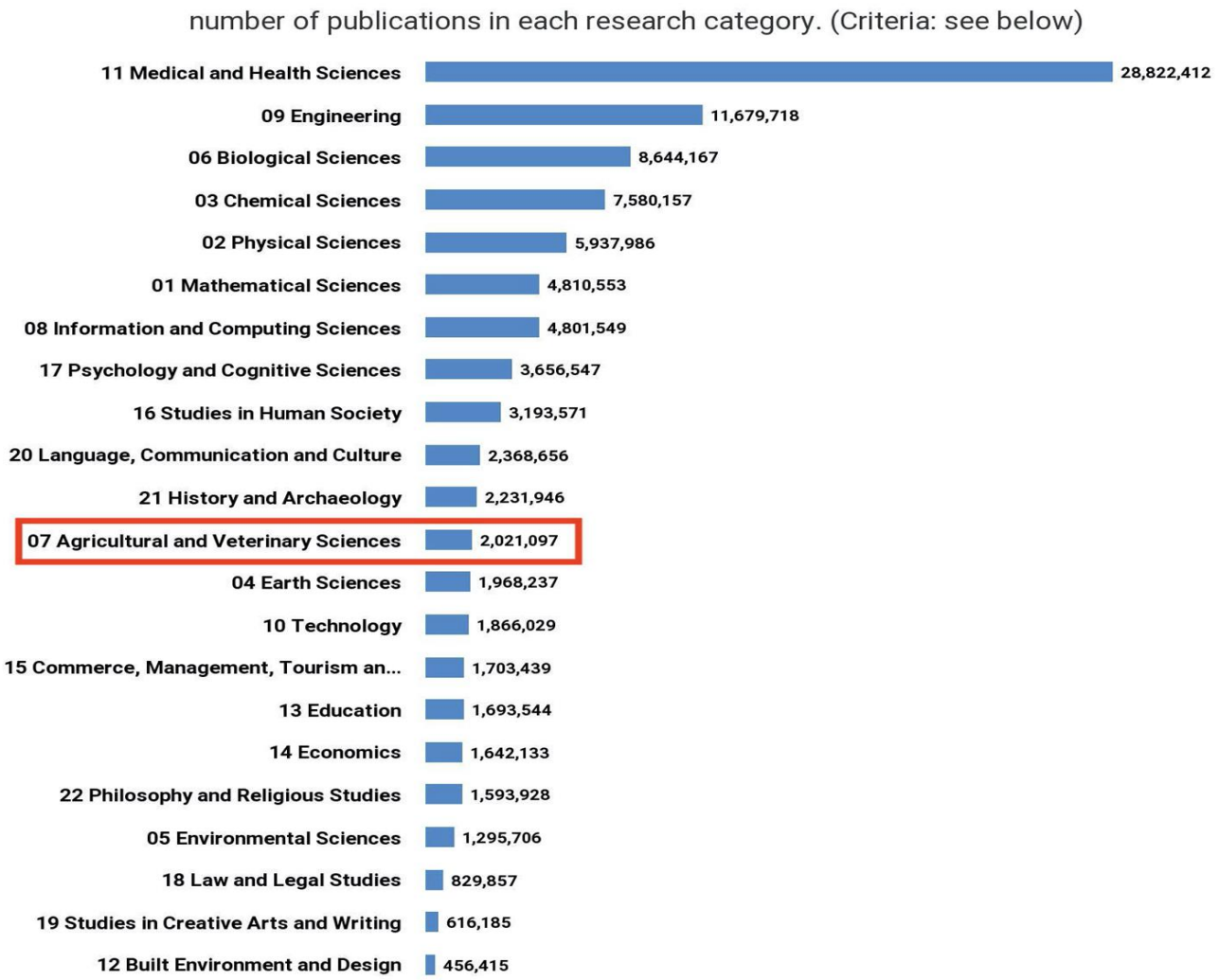
- **Derecho de acceso a la información científica**
- Especialmente en el área de los **datos de investigación y datos de la administración**
- **No está al alcance** de todos aprovechar su potencial
- Solo sobre información y datos se pueden construir **modelos capaces de describir un fenómeno y predecir el éxito** de una estrategia determinada definiendo índices con análisis multivariantes=**Inteligencia artificial**

Agricultura

Ciencia aplicada

La mayor parte del resultado de la investigación NO se encuentra en la literatura revisada por pares

Gran parte de la investigación la realizan empresas privadas y organizaciones públicas.



Source: https://login.research4life.org/tacsgr1app_dimensions_ai
Exported: July 01, 2020
Criteria: none.

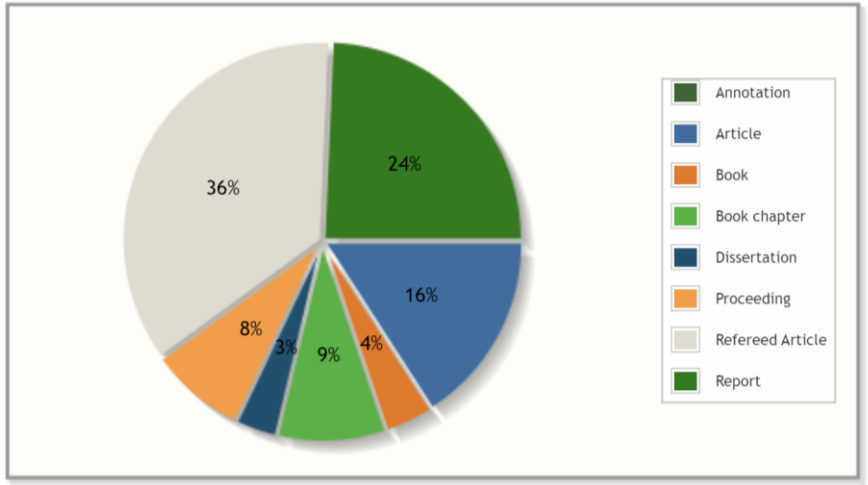
© 2019 Digital Science and Research Solutions Inc. All rights reserved. Non-commercial redistribution / external re-use of this work is permitted subject to appropriate acknowledgement. This work is sourced from Dimensions® at www.dimensions.ai.

Diversidad de tipologías de documentos

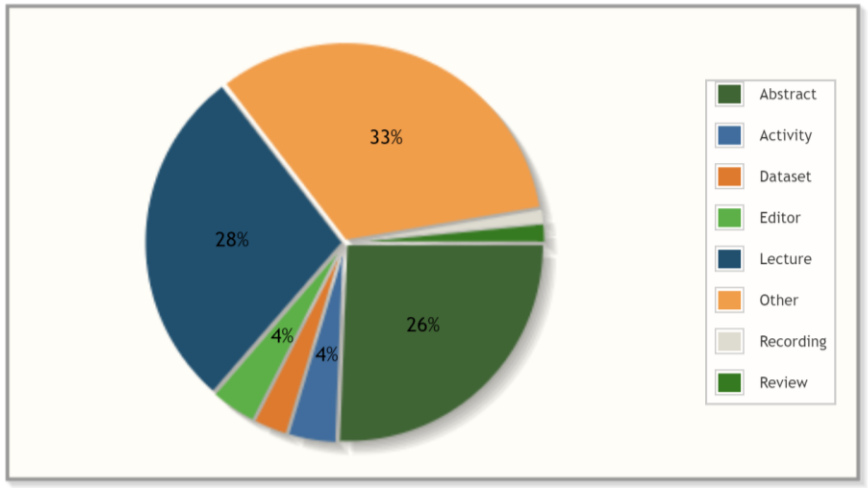
Las áreas principales de actividad de la Universidad e Investigación de Wageningen (WUR) son la alimentación, la producción de alimentos, el entorno de vida, la salud, el estilo de vida y los medios de subsistencia.

La diversidad en su producción de investigación incluye no solo artículos.

242847 publications



85742 other output



Source: Wageningen University & Research Output by document type. https://library.wur.nl/WebQuery/wurpubs/show Retrieved 2022.

Dar visibilidad a la investigación agrícola y alimentaria

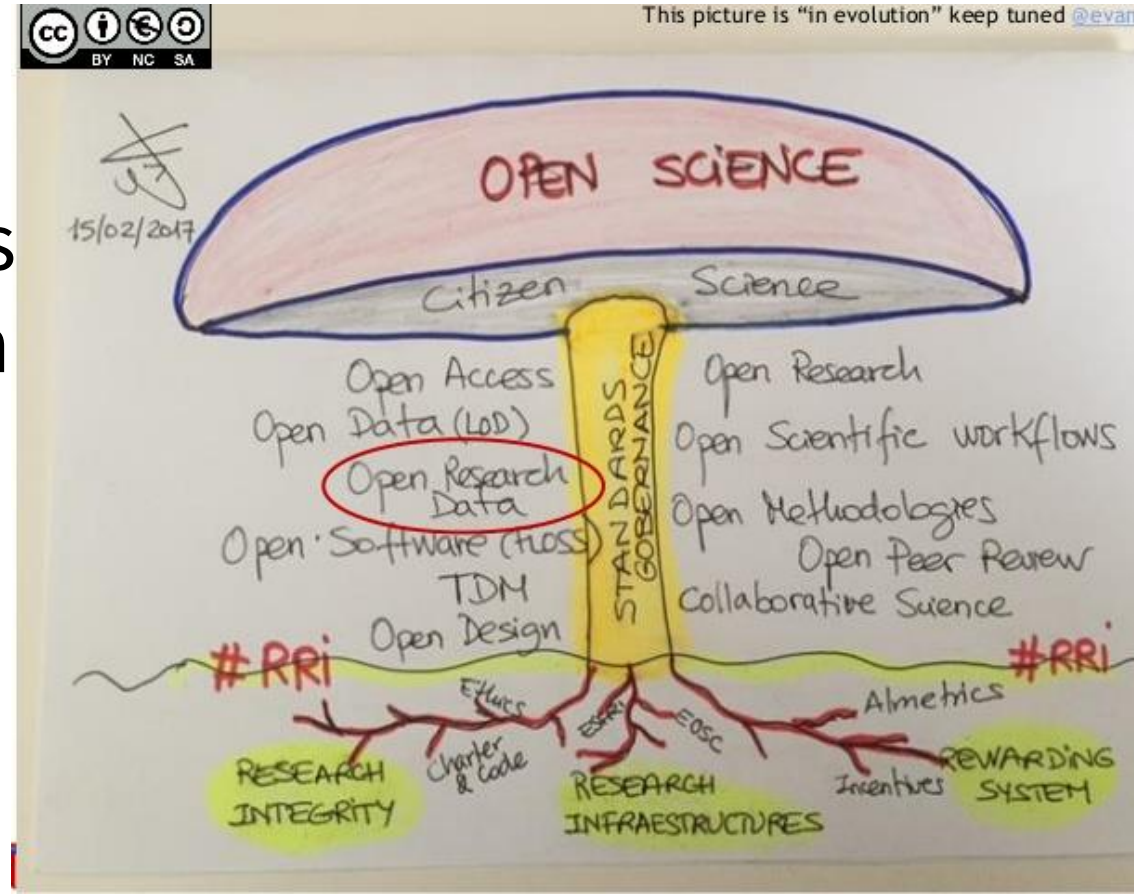
La FAO

La Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) promueve el intercambio de investigación científica y conocimientos técnicos relacionados con todos los aspectos de la alimentación y la agricultura.

Open Science

Es un “constructo inclusivo que combina diversos movimientos y prácticas con el fin de que los conocimientos científicos multilingües estén abiertamente disponibles y sean accesibles para todos, así como reutilizables”: **Open Access+Open Data+Open to Society**

https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa



Méndez, Eva (2021). OpenScience por defecto. La nueva normalidad para la investigación.Arbor, 197(799): a587. https://doi.org/10.3989/arbor.2021.799002

Open access

AGRIS

http://agris.fao.org

Sistema de información internacional para las ciencias y tecnologías agrícolas desde 1975

AGRIS recopila información bibliográfica de todo el mundo sobre publicaciones relacionadas con la alimentación y la agricultura

Indexa referencias bibliográficas pero no almacena el texto completo. Recursos previos a 2010 pueden no tener enlace al texto

Estadísticas de uso

2020

9.7 M
Visitas a páginas

5.9 M
Usuarios

2021

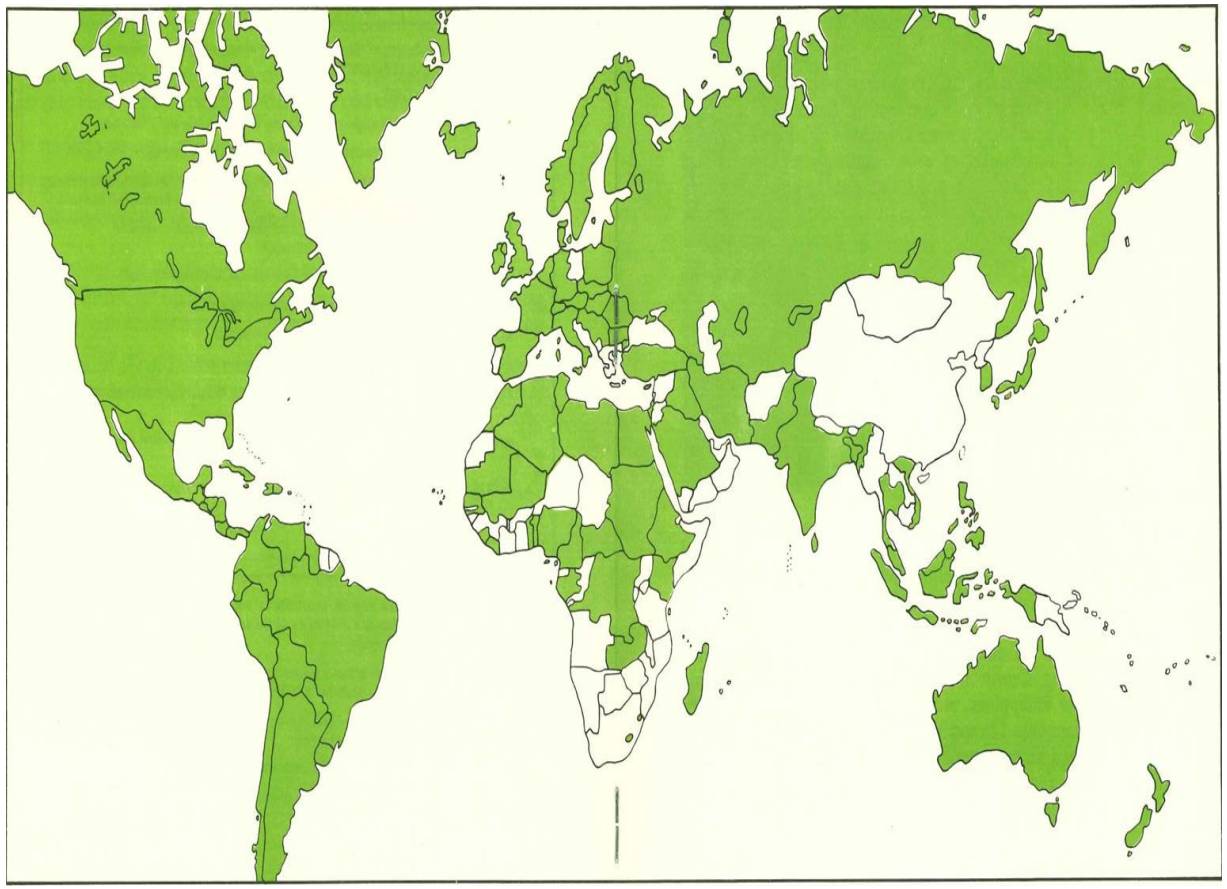
12 M
Visitas a páginas

8 M
Usuarios

50% nuevos usuarios

Fuente: Google Analytics

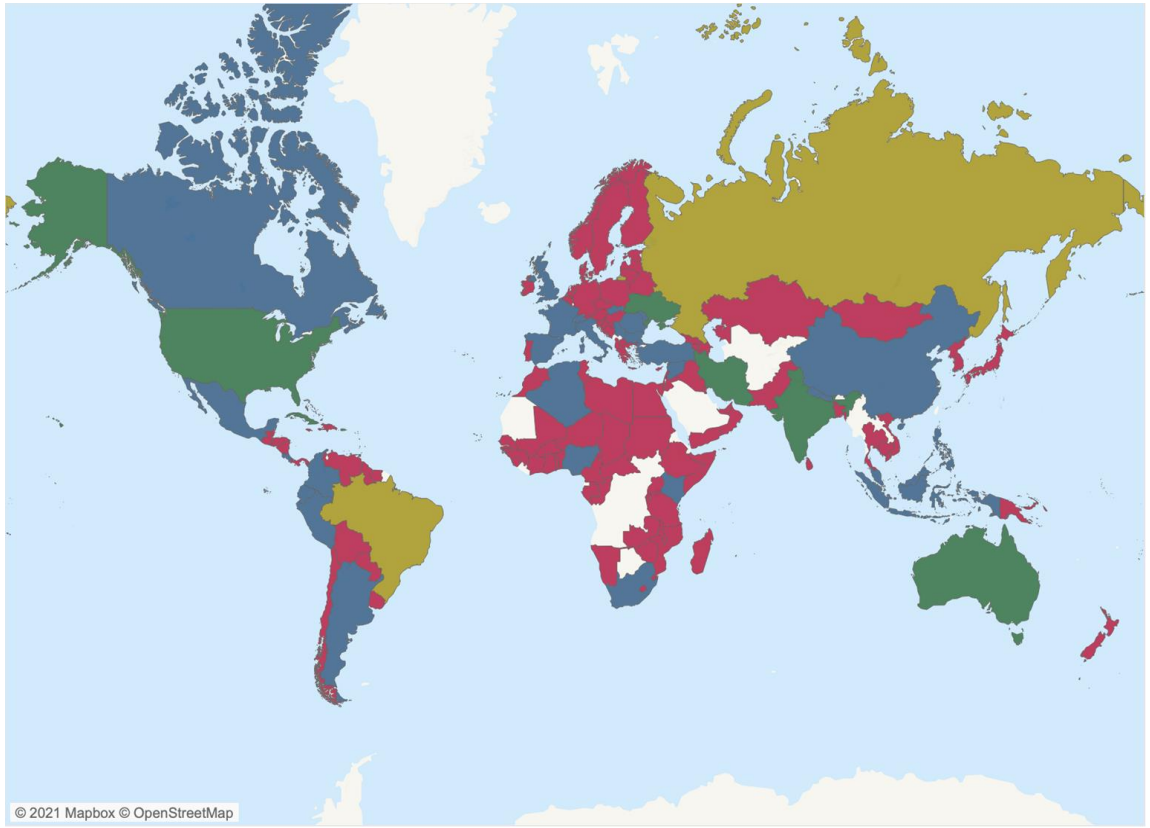
1978 1 organización por país



2022

465

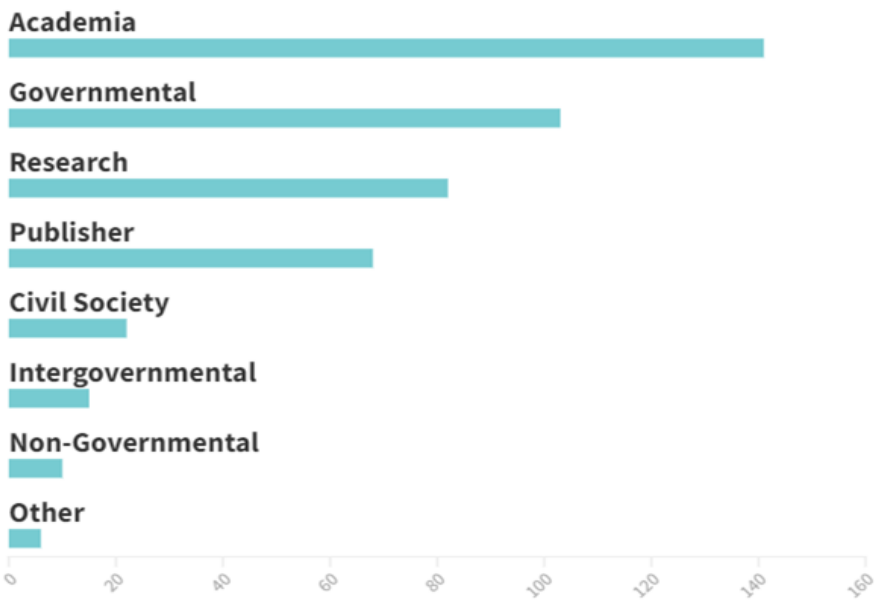
Data Providers



Proveedores de datos

De 150 países: el **multilingüismo** es una característica fundamental de AGRIS

Países con más de 10 proveedores de datos



Proveedores de datos por tipo

¿Por qué ser parte de la Red AGRIS?

- Ser **parte de una red internacional dinámica** con un enfoque en contenido agrícola que es accesible, utilizado intensivamente y distribuido masivamente en el panorama de información global;
- **Intercambio de información y datos** entre diferentes países y partes interesadas;
- **Compartir materiales únicos**, como informes científicos y técnicos no publicados, publicaciones gubernamentales, materiales de conferencias, conjuntos de datos…
- Aumentar la **visibilidad** de la organización a nivel global;
- Aumentar el **uso** de los resultados de la investigación institucional;

Open research data

Datos

- Necesitan una gestión muy cuidadosa si queremos reutilizarlos
- Son clave en la reproducibilidad de la ciencia y en la rendición de cuentas públicas
- Muchos datos de investigación tienen un periodo de embargo y se suele utilizar el CC By
- Algunos datos de la administración nunca estarán en abierto (ej. Militares)
- Los datos de investigación y administrativos son muy heterogéneos

FAIR Data

El acrónimo agrupa los cuatro principios que le dan significado, referidos a cualidades precisas y medibles que debería tener toda publicación formal de datos:

- **Findable**: se pueden encontrar, para lo que se acompañan de metadatos que identifican, describen y permiten localizar los datos.
- **Accessible**: son accesibles, porque pueden recuperarse mediante protocolos estandarizados de comunicación y los metadatos persisten aún cuando los datos dejen de estar disponibles.
- **Interoperable**: se presentan de forma que resultan aplicables e incluyen referencias a otros datos.
- **Reusable**: se pueden reutilizar, porque queda clara la procedencia de los datos y las condiciones de su reutilización.

Interoperabilidad2:
Los (meta) datos usan vocabularios que siguen los principios FAIR

¿Qué significa?

El vocabulario controlado que se utiliza para describir los conjuntos de datos debe documentarse y resolverse mediante identificadores persistentes y únicos a nivel mundial.

Cualquier persona que utilice el conjunto de datos debe poder encontrar y acceder fácilmente a esta documentación.

AGROVOC

1980, la FAO ha apoyado los esfuerzos para un mayor intercambio de conocimiento, visibilidad y acceso mediante la estandarización y el mejoramiento de métodos para encontrar información sobre alimentación y agricultura, mediante la publicación de AGROVOC. Es el Tesauro más utilizado para etiquetar artículos publicados y conjuntos de datos, para facilitar a los investigadores la búsqueda de información y enlaces semánticos.

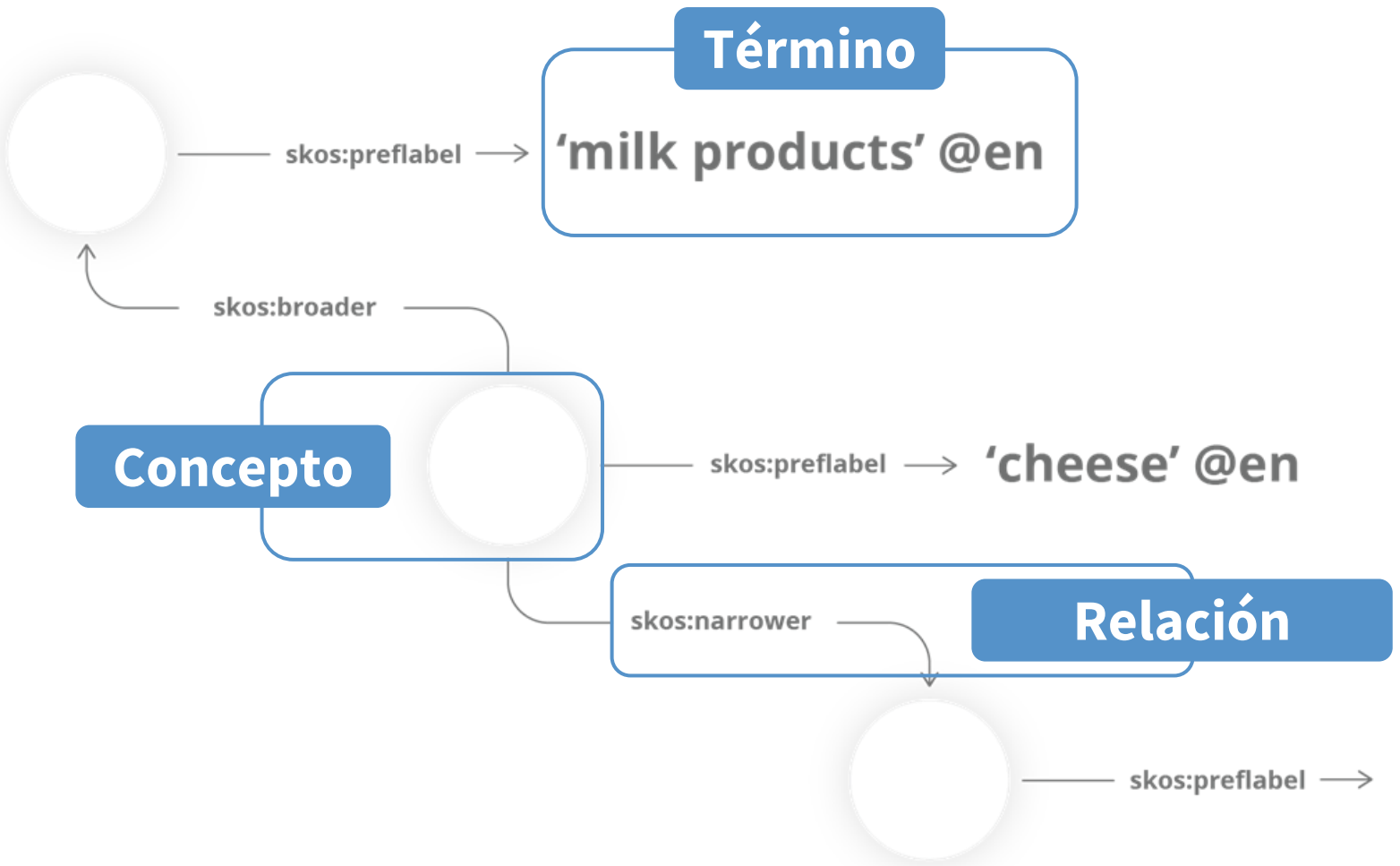
AGROVOC

Originalmente concebido con el fin de describir documentos y otros recursos de información para indexación y búsqueda, AGROVOC ha evolucionado desde los catálogos impresos hasta las **tecnologías semánticas de la web**.

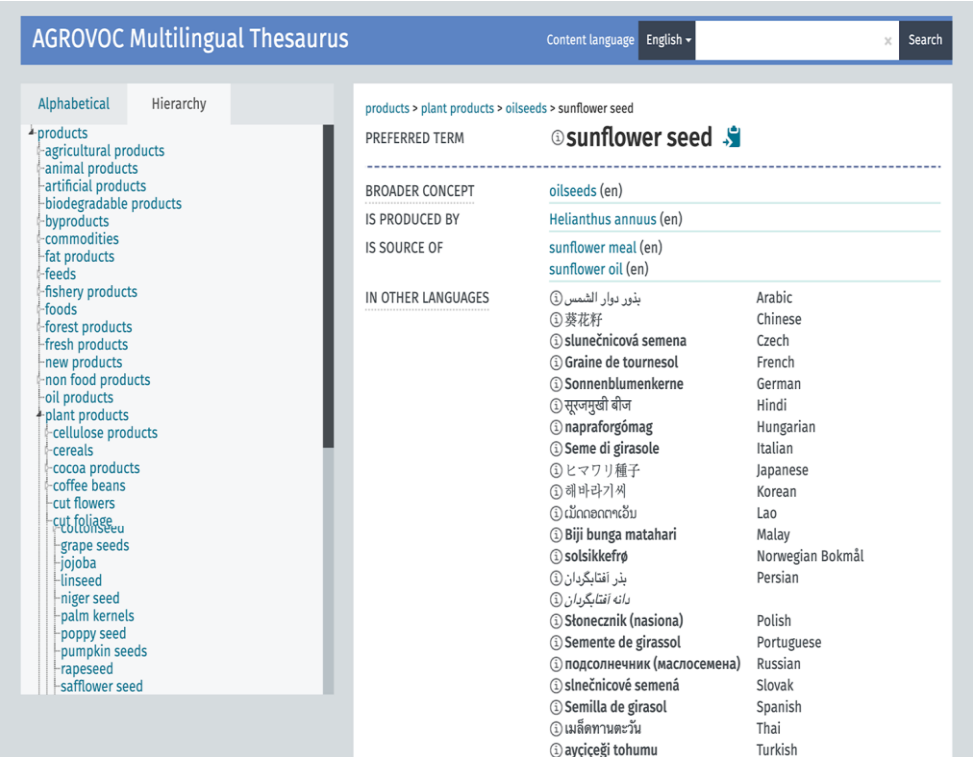
AGROVOC está ahora en línea y vinculado a otros sistemas multilingües de organización del conocimiento, construyendo puentes entre conjuntos de datos.

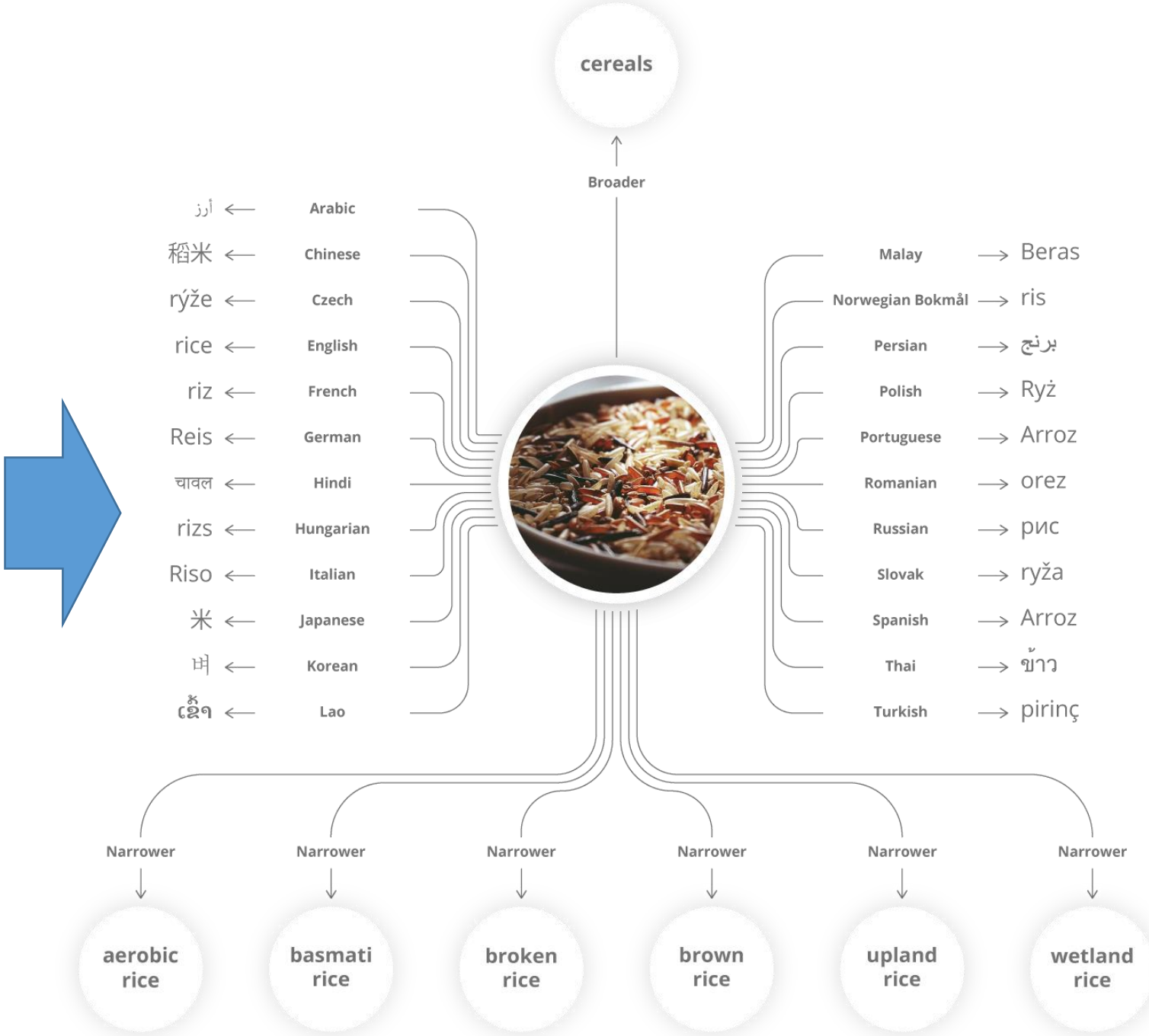
Distribución **CC BY IGO 3.0**

Basado en un modelo conceptual:SKOS



37.000+ conceptos
750.000 términos
alineado con más de 20 conjuntos de datos
hasta 40 idiomas


AGROVOC Multilingual Thesaurus
Content language
English
Search
Alphabetical
Hierarchy
products
agricultural products
animal products
artificial products
biodegradable products
byproducts
commodities
fat products
feeds
fishery products
foods
forest products
fresh products
new products
non food products
oil products
plant products
cellulose products
cereals
cocoa products
coffee beans
cut flowers
cut foliage
cottonseed
grape seeds
jojoba
linseed
niger seed
palm kernels
poppy seed
pumpkin seeds
rapeseed
safflower seed
products > plant products > oilseeds > sunflower seed
PREFERRED TERM
sunflower seed
BROADER CONCEPT
oilseeds (en)
IS PRODUCED BY
Helianthus annuus (en)
IS SOURCE OF
sunflower meal (en)
sunflower oil (en)
IN OTHER LANGUAGES
بذور دوار الشمس Arabic
葵花籽 Chinese
slunečnicová semena Czech
Graine de tournesol French
Sonnenblumenkerne German
सूरजमुखी बीज Hindi
napraforgómag Hungarian
Seme di girasole Italian
ヒマワリ種子 Japanese
해바라기씨 Korean
ເມັດດອກຕາເວັນ Lao
Biji bunga matahari Malay
solsikkefrø Norwegian Bokmål
بذر آفتابگردان Persian
دانه آفتابگردان
Słonecznik (nasiona) Polish
Semente de girassol Portuguese
подсолнечник (маслосемена) Russian
slnečnicové semená Slovak
Semilla de girasol Spanish
เมล็ดทานตะวัน Thai
ayçiçeği tohumu Turkish


cereals
Broader
أرز Arabic
稻米 Chinese
rýže Czech
rice English
riz French
Reis German
चावल Hindi
rizs Hungarian
Riso Italian
米 Japanese
벼 Korean
ເຂົ້າ Lao
Malay Beras
Norwegian Bokmål ris
Persian برنج
Polish Ryż
Portuguese Arroz
Romanian orez
Russian рис
Slovak ryža
Spanish Arroz
Thai ข้าว
Turkish pirinç
Narrower
aerobic rice
Narrower
basmati rice
Narrower
broken rice
Narrower
brown rice
Narrower
upland rice
Narrower
wetland rice

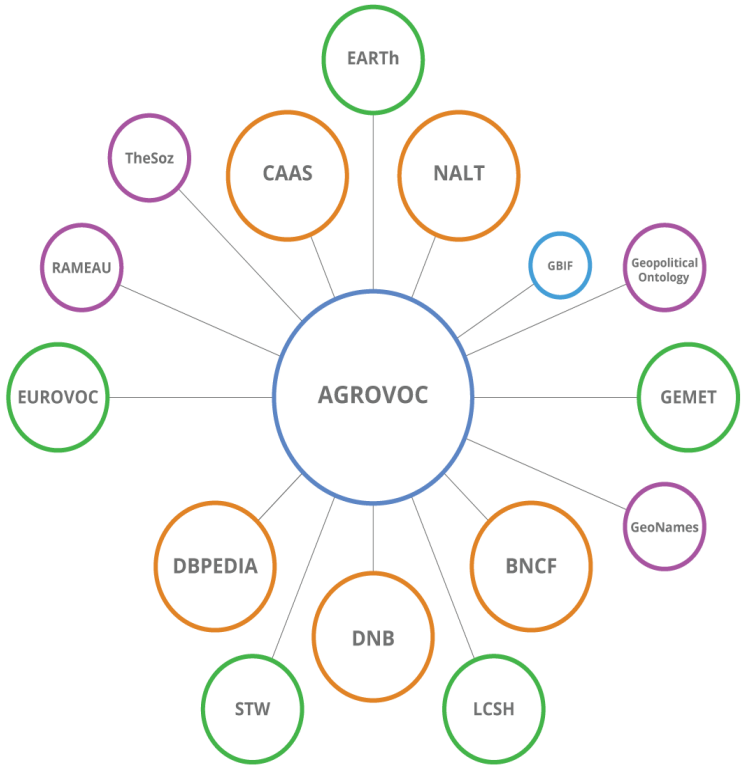
Organizaciones contribuyendo a AGROVOC

30 activos

25 paises

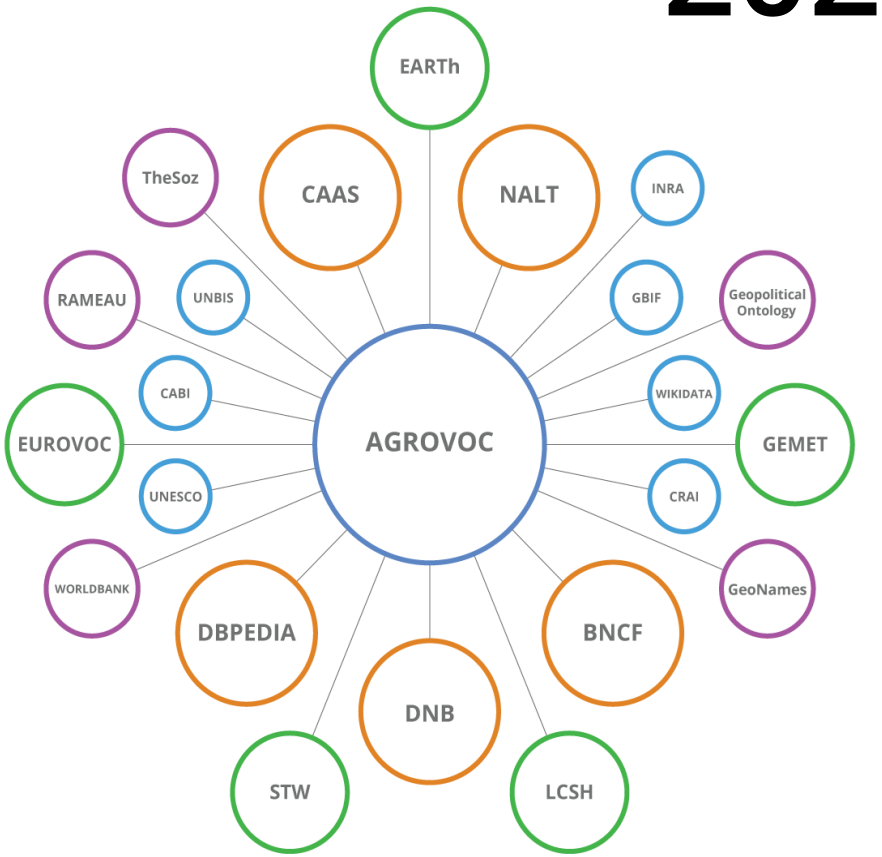
AGROVOC enlaces

2017



Número de enlaces: **61063**

2020



Número de enlaces: **63477**

¿Qué hacer con los datos?

Gestión por diseño

- Será la clave del éxito
- Caso de los planes de gestión de datos (=transparencia de los gobiernos)

Beneficios de planificar la gestión de datos

Demuestra la **viabilidad** de la investigación (colaboraciones y recursos institucionales)

- Planificar la gestión demanda una reflexión previa que afecta al diseño experimental, aporta robustez al método científico
- Buscar datos existentes evita duplicar trabajos recolectándolos de nuevo y promociona la colaboración con otros grupos de investigación y la cita de datos, tanto los propios como los ajenos.
- Convierte al investigador en consumidor de otros datos, para lo que debe conocer, por ejemplo, las fuentes
- Documentar los datos que genera un proyecto lo dota de sostenibilidad si los recursos humanos cambian

(Jones, 2011)

Exigidos por financiadores como Horizon

"***A goal without a plan is just a wish.***" Antoine de Saint-Exupery (1900 -1944)

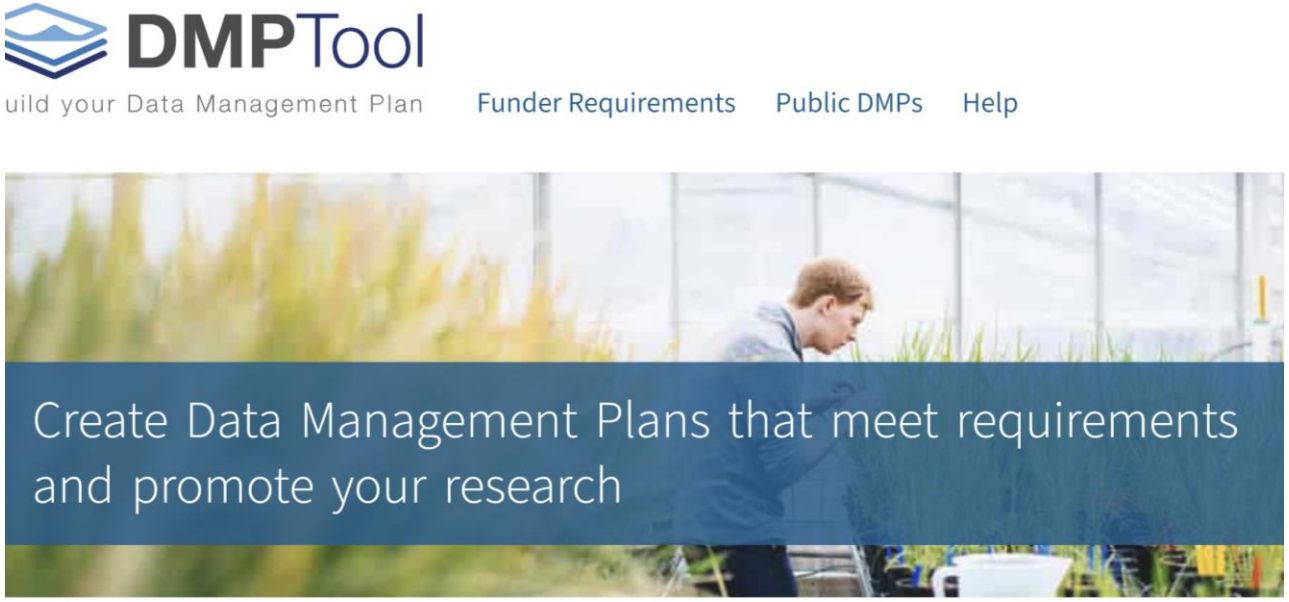
Información sobre:

- el tratamiento de los datos de la investigación durante y después de la finalización del proyecto
- qué datos se recogerán, procesarán y/o generarán
- qué metodología y estándares se aplicarán
- si los datos serán compartidos o de acceso abierto y
- cómo se curarán y conservarán los datos (incluso después del final del proyecto).
- Open // As open as posible, as closed as necessary // Gestión desde el diseño

FAQs https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;categories=;programme=null;actions=;keyword=ordH2020

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Herramientas



75,843 Users

335 Participating Institutions

73,855 Plans

DMP(UK). *Postgraduate Research Project*

Definir los datos

- ¿Cómo los obtienes e instrumentos?
- ¿Cuánto se actualizan?
- ¿Cuántos generas y en qué formatos?
- ¿Cuántas versiones almacenas?

Controlar

- ¿Cuánta información adicional es necesaria para entenderlos?
- ¿Dónde los almacenas?
- Directorios y nombres de archivos
- Copias de seguridad ¿cómo y cuándo? ¿testeas?

Compartir

- ¿De quién es la propiedad?
- ¿Quién puede usarlos y quién podría?
- ¿Qué compartes y qué no? ¿por?

Archivo de datos

- ¿Qué debe ser archivado?
- ¿Por cuánto tiempo y dónde?
- ¿Cuándo pasan al estado “archivo”?
- ¿Quién es el responsable de moverlos?
- ¿Quién tendrá acceso? Condiciones

Supervisión del plan

- ¿Quién es responsable?
- ¿Con qué frecuencia se actualizará?
- …

Conclusion

La **interoperabilidad** de la información y los datos es hoy día una realidad

Si los datos pueden ser utilizados y distribuidos libremente por cualquier persona / servicio la **innovación** se dispara

La Inteligencia artificial necesita no sólo los modelos matemáticos que la sustentan sino datos de **calidad**, CONFIABLES, para que su capacidad de predicción también los sea

DATOS RESPONSABLES

Esperamos que haya sido de su interés

Boulton (2015) La ciencia es una actividad internacional, pero que se realiza inmersa en una cultura de trabajo nacional por lo que necesita que las estrategias nacionales estén ajustadas al marco común internacional.

Peset, F. (2014). Datause. https://media.upv.es/#/portal/video/f061ca60-69e5-11eb-9339-cbab4b322e5a
@datausedatasea