

Sobre los repositorios de datos y los sistemas de información de investigación actuales

Marc Goovaerts,
Hasselt University Library, Belgium
Hatun Tinkuy, Cusco – Agosto 12, 2022
marc.goovaerts@uhasselt.be



Resumen

- **Repositorios de datos**

- ¿Por qué compartir datos de investigación?
- ¿Qué es un repositorio de datos?
- ¿Cómo debo elegir un repositorio de datos?
- Seis repositorios para compartir datos de investigación
- Motores de búsqueda para repositorios de datos
- Tres soluciones de software de código abierto para repositorios de datos

- **Sistema de información sobre la investigación actual (CRIS)**

- Definición
- Central en la gestión de la información de investigación de una organización de investigación
- Sistemas FOSS - Tres ejemplos

¿Por qué compartir datos de investigación?

- **Compartir la información estimula la ciencia.** Cuando los investigadores deciden poner sus datos a disposición del público, están permitiendo que su trabajo contribuya mucho más allá de sus hallazgos originales.
- Los beneficios de compartir datos son inmensos:
 - Cuando los investigadores hacen públicos sus datos, aumentan la transparencia y la confianza en su trabajo, permiten que otros reproduzcan y validen sus hallazgos y, en última instancia, contribuyen al ritmo de los descubrimientos científicos al permitir que otros reutilicen sus datos y se basen en ellos.

"If I have seen further it is by standing on the shoulders of Giants."

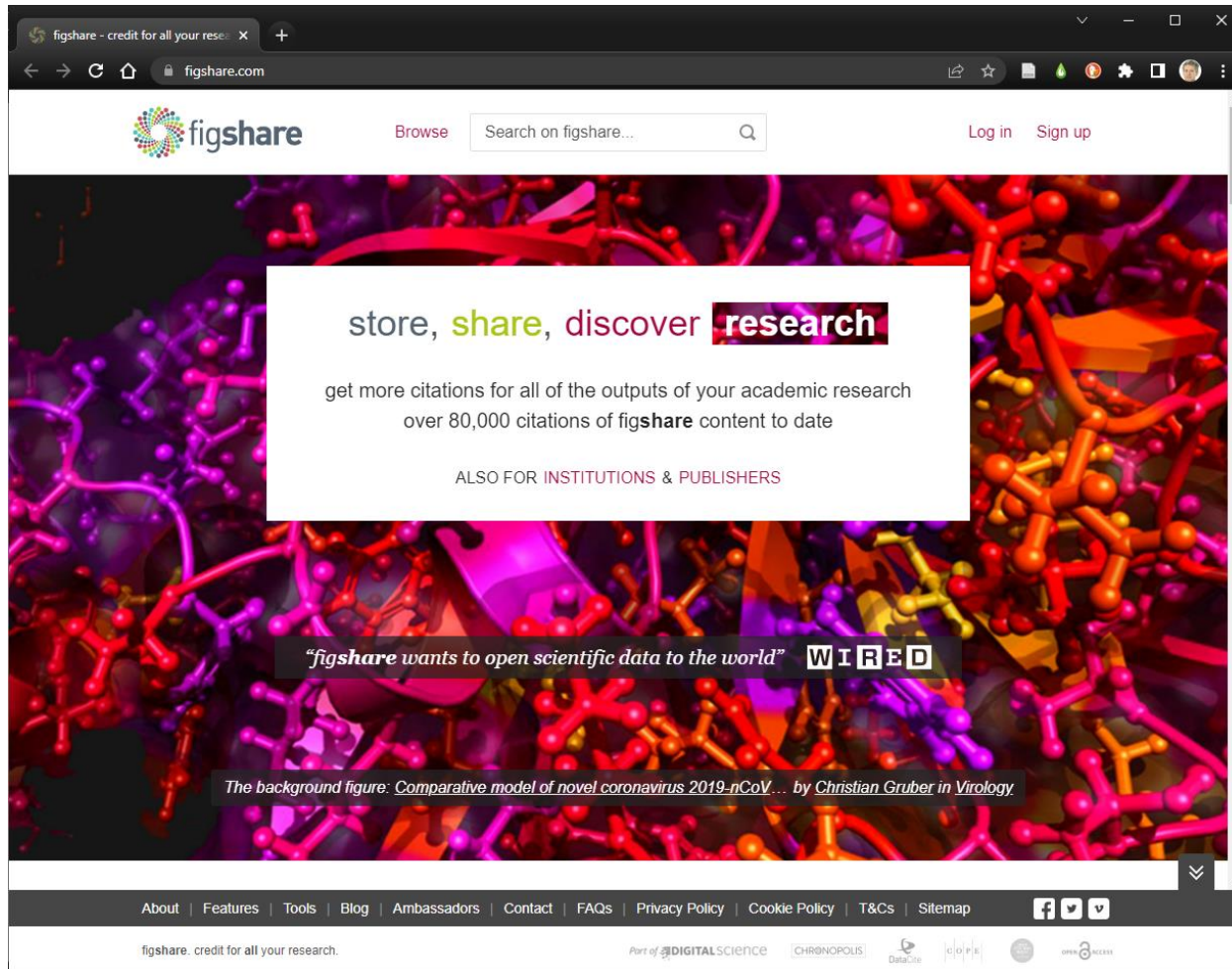
Isaac Newton, 1675.

¿Qué es un repositorio de datos?

- **Los repositorios de datos son un lugar para organizar, preservar, compartir y reutilizar datos**
- La mejor manera de publicar y compartir datos de investigación es con un repositorio de datos de investigación. Un repositorio es una base de datos en línea que permite conservar los datos de investigación a lo largo del tiempo y ayuda a otros a encontrarlos.
- Los datos de investigación tienen una vida más larga que cualquier proyecto. Tendrás que planificar cómo gestionar los datos antes, durante y después del proyecto
- Además de archivar los datos de la investigación, un repositorio asignará un DOI a cada objeto cargado y proporcionará una página web que indique qué es, cómo citarlo y cuántas veces otros investigadores han citado o descargado ese objeto.

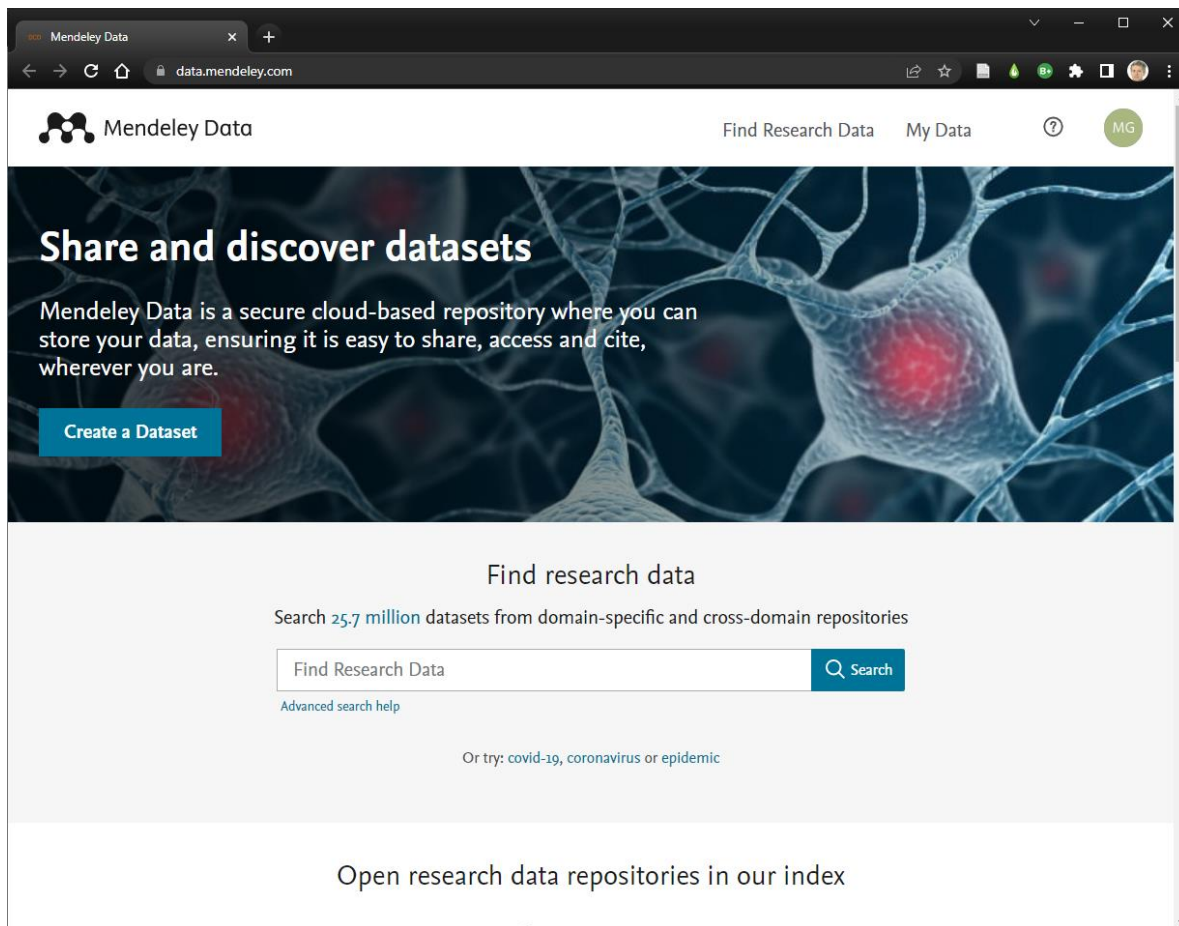
Seis repositorios para compartir datos de investigación

1. Figshare



- Figshare es un repositorio comercial de datos de acceso abierto, lanzado en 2011, en el que los investigadores pueden preservar sus resultados de investigación, como conjuntos de datos, imágenes y vídeos, y hacerlos descubribles.
- Figshare permite a los investigadores cargar cualquier formato de archivo y asigna un identificador de objeto digital (**DOI**) para las citas.
- Más de 50 instituciones utilizan ya esta solución.
- Figshare publica cada año "[The State of Open Data](#)" para evaluar el cambiante panorama académico en torno a la investigación abierta.
- **Las cuentas gratuitas** de Figshare pueden cargar **archivos de hasta 5 gb** y obtener **20 gb de almacenamiento gratuito**.

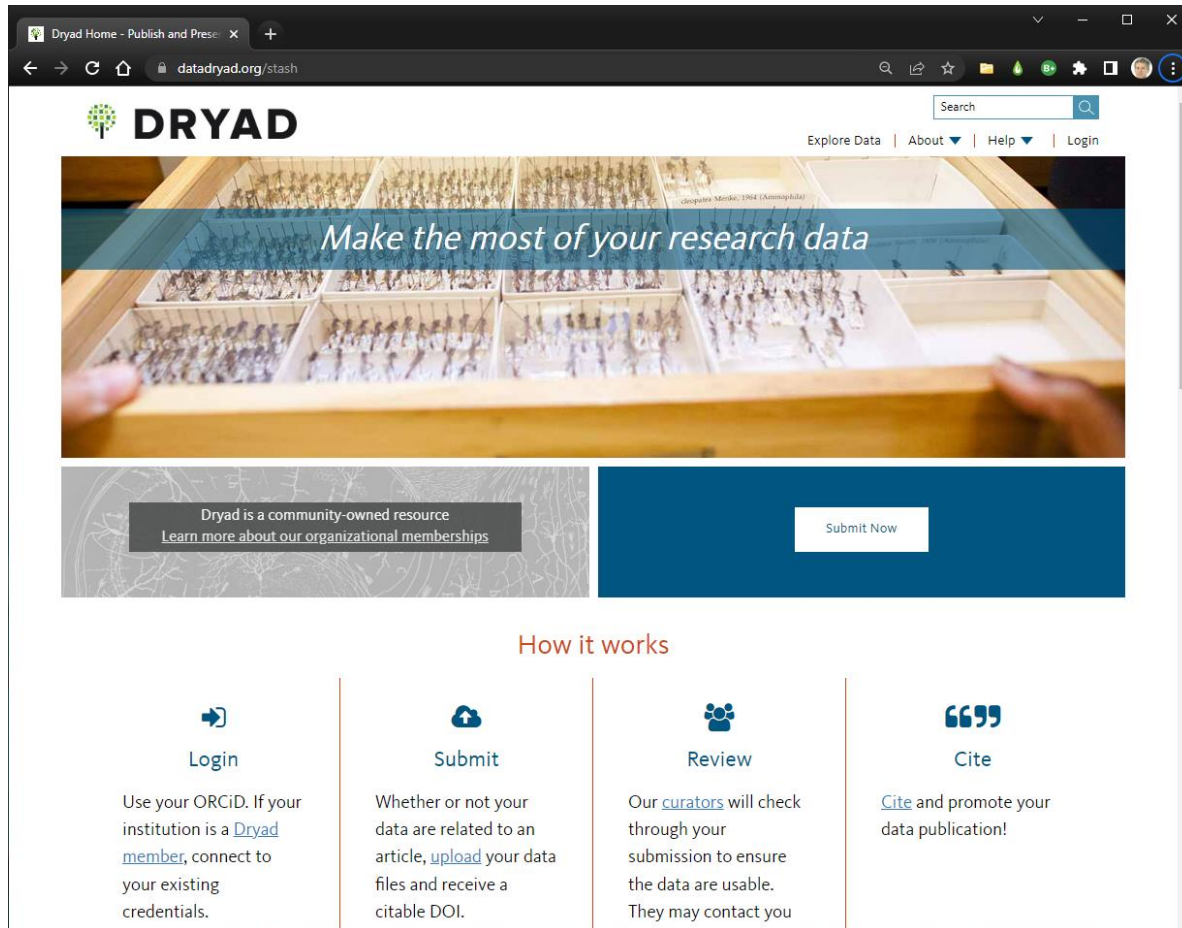
2. Mendeley Data (Elsevier)



The screenshot shows the Mendeley Data website. At the top, there is a navigation bar with the Mendeley logo, the text "Mendeley Data", and links for "Find Research Data" and "My Data". Below the navigation bar is a large banner with a microscopic image of cells. The banner contains the text "Share and discover datasets" and "Mendeley Data is a secure cloud-based repository where you can store your data, ensuring it is easy to share, access and cite, wherever you are." A blue button labeled "Create a Dataset" is positioned on the left side of the banner. Below the banner is a search section titled "Find research data" with the text "Search 25,7 million datasets from domain-specific and cross-domain repositories". A search input field contains the text "Find Research Data" and a blue "Search" button. Below the search field is a link for "Advanced search help" and a suggestion "Or try: covid-19, coronavirus or epidemic". At the bottom of the search section, there is a link "Open research data repositories in our index".

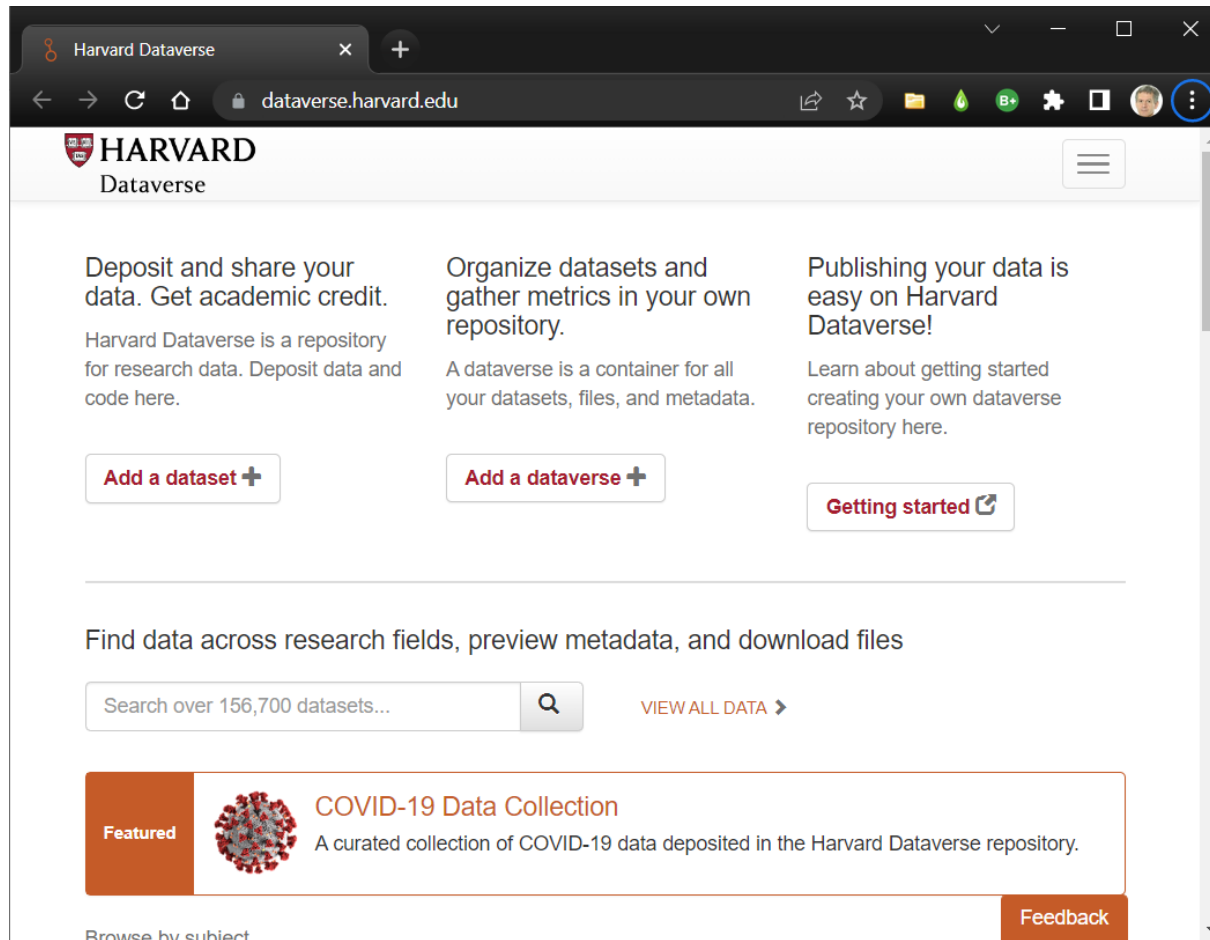
- Mendeley Data es un repositorio de datos de investigación abierto, donde los investigadores pueden almacenar y compartir sus datos. Los conjuntos de datos pueden ser compartidos de forma privada entre individuos, así como públicamente con el mundo.
- La misión de Mendeley es facilitar el intercambio de datos. En sus propias palabras, "cuando los datos de la investigación se ponen a disposición del público, la ciencia se beneficia:
 - los resultados pueden ser verificados y reproducidos - los datos pueden ser reutilizados de nuevas maneras
 - se facilita el descubrimiento de investigaciones relevantes
 - los financiadores obtienen más valor de su inversión en financiación".
- **Los investigadores pueden cargar y almacenar sus trabajos de forma gratuita en Mendeley Data.**

3. Dryad Digital Repository



- Dryad es un repositorio de propósito general que hace que los datos sean descubribles, reutilizables libremente y citables.
- Acceso gratuito para particulares e instituciones
- Se pueden enviar la mayoría de los tipos de archivos (por ejemplo, texto, hojas de cálculo, vídeo, fotografías, código de software), incluidos los archivos comprimidos de múltiples archivos.
- Dryad financia su funcionamiento cobrando **una cuota de 120 dólares cada vez que se publican datos.**

4. Harvard Dataverse



The screenshot shows the Harvard Dataverse website homepage. At the top, there is a navigation bar with the Harvard logo and the text "HARVARD Dataverse". Below this, there are three columns of text and buttons:

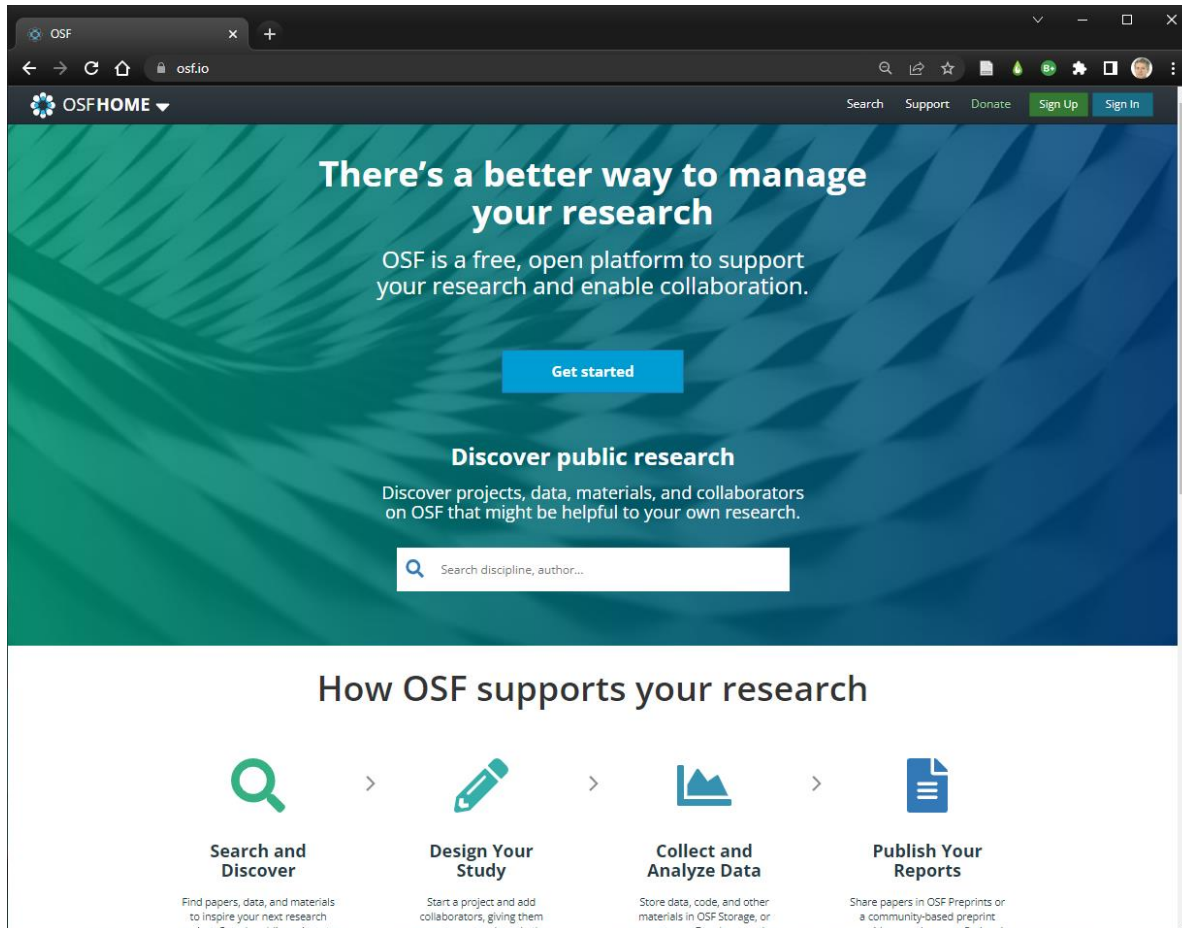
- Column 1:** "Deposit and share your data. Get academic credit. Harvard Dataverse is a repository for research data. Deposit data and code here." Below this is a button labeled "Add a dataset +".
- Column 2:** "Organize datasets and gather metrics in your own repository. A dataverse is a container for all your datasets, files, and metadata." Below this is a button labeled "Add a dataverse +".
- Column 3:** "Publishing your data is easy on Harvard Dataverse! Learn about getting started creating your own dataverse repository here." Below this is a button labeled "Getting started" with an external link icon.

Below these columns, there is a search bar with the text "Search over 156,700 datasets..." and a magnifying glass icon. To the right of the search bar is a link labeled "VIEW ALL DATA >".

At the bottom, there is a featured section with a red and white circular graphic and the text "Featured COVID-19 Data Collection A curated collection of COVID-19 data deposited in the Harvard Dataverse repository." Below this is a "Feedback" button.

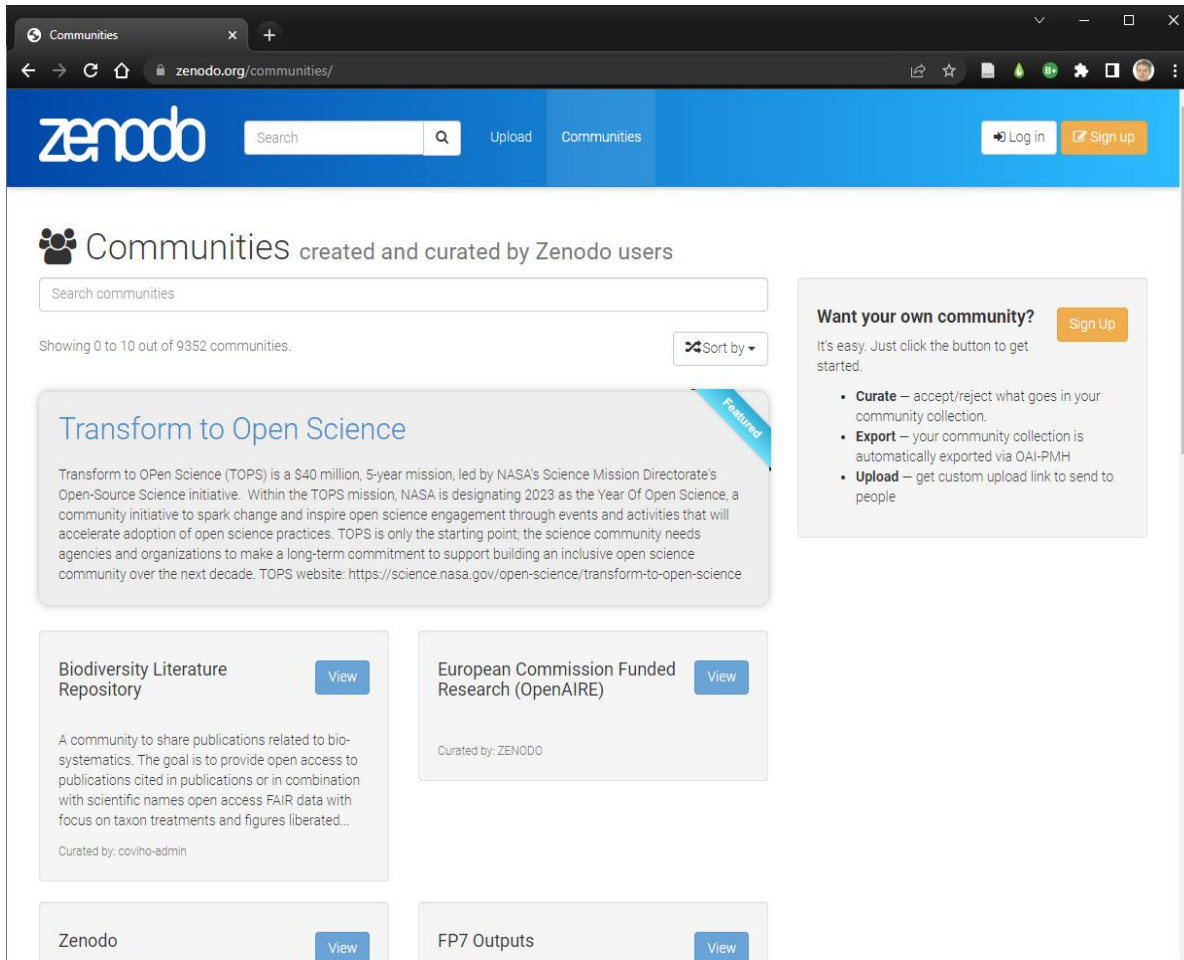
- Harvard Dataverse es un repositorio de datos en línea donde los científicos pueden preservar, compartir, citar y explorar los datos de investigación.
- El repositorio Harvard Dataverse funciona con **la aplicación web de código abierto Dataverse**, desarrollada por el Instituto de Ciencias Sociales Cuantitativas de Harvard.
- **Harvard Dataverse es gratuito y tiene un límite de 2,5 GB por archivo y 10 GB por conjunto de datos.**

5. Open Science Framework



- OSF es una herramienta de colaboración y gestión de la investigación, de código abierto, diseñada para ayudar a los investigadores a documentar el ciclo de vida de sus proyectos y archivar materiales.
- A cada usuario, proyecto, componente y archivo se le asigna un localizador uniforme de recursos (**LOD**) único y persistente para facilitar el intercambio y promover la atribución.
- A los proyectos también se les pueden asignar identificadores de objetos digitales (**DOI**) si se ponen a disposición del público.
- Está construida y mantenida por la organización sin ánimo de lucro Center for Open Science. OSF es un **servicio gratuito**.

6. Zenodo



- Zenodo es un repositorio de acceso abierto de propósito general desarrollado bajo el programa europeo OpenAIRE y operado por el CERN (*originalmente repositorio de registros huérfanos de OpenAire*),
- Zenodo anima a los usuarios a **subir** sus resultados de investigación **en las primeras etapas de su ciclo de vida** permitiendo que sean privados. Una vez que se publica un artículo asociado, los conjuntos de datos se convierten automáticamente en abiertos.
- **Zenodo no restringe el tipo de archivo que los investigadores pueden cargar y acepta conjuntos de datos de hasta 50 GB.**

Motores de búsqueda para repositorios de datos

- Utilice [FAIRsharing](https://fairsharing.org) y re3data.org para buscar un repositorio adecuado: ambos ofrecen una lista de repositorios de datos certificados.

Terminology Artifact	830
Model/Format	510
Reporting Guideline	234
Identifier Schema	23

Repositories	967
Knowledgebases	794
Knowledgebase/Repositories	124

Journal	93
Funder	23
Society	13
Project	13



Mapping the global repository landscape

International collaboration has become very common in the production of scientific knowledge (Coccia and Wang, 2016; Monasterky & van Noorden, 2019) and contributes substantially to the research output of several countries

How open are repositories in re3data?

Openness may sound self-evident, but in fact it can mean different things, even within the Open Science community (Pasquetto, Sands & Borgman, 2015). To gain understanding and to ensure common goals in the transition to Open

Reviewing the subject classification in re3data

When looking for a suitable repository, the discipline(s) a repository caters to is one of the most important pieces of information. The re3data Metadata Schema reflects this priority by requiring the element subject for all indexed repositories. However...

Tres soluciones de software de código abierto para repositorios de datos

Soluciones de software de código abierto para repositorios de datos

- El desarrollo de un repositorio de datos por parte de la propia institución es un reto La Universidad de Hasselt sólo ofrece, por ejemplo, un [repositorio de metadatos](#) para los datos de investigación. Se desarrollará una solución a nivel de las universidades de Flandes para almacenar los conjuntos de datos .
- FOSS soluciones :
 - Dspace/DSpace-CRIS
 - Dataverse
 - Invenio RDM

1. DSpace / DSpaceCRIS

The screenshot shows the DSpace website homepage. At the top, there is a navigation menu with the LYRASIS logo and links for About DuraSpace, Projects, Services, Community, Membership, Support, and News & Events. Below the menu, there is a large banner featuring the DSpace logo and a green button that says "FREE DOWNLOAD". The banner also includes the text: "THE SOFTWARE OF CHOICE FOR ACADEMIC, NON-PROFIT & COMMERCIAL ORGANIZATIONS BUILDING OPEN DIGITAL REPOSITORIES". Below the banner, there are four columns of content, each with an icon and a title: "ABOUT" (with a clipboard icon), "FEATURES" (with a gear icon), "COMMUNITY" (with a speech bubble icon), and "RESOURCES" (with a book icon). Each column contains a brief description of the respective section.

The screenshot shows the DSpaceCRIS interface. At the top, there is a navigation menu with the UHASSELT logo and links for Home, Research Publications, Artistic/Designerly Creations, Datasets, Theses, and Administration. Below the menu, there is a search bar and a "Sign in" button. The main content area is divided into two sections: "Browse" and "Search Datasets". The "Browse" section has a red sidebar with filters for Contributor, Title, Issue Date, and Research Discipline. The "Search Datasets" section has a search form with dropdown menus for "All" and "AND" options, and a "Search" button. Below the search form, there are two sections: "Recent Submissions" and "Top Viewed". The "Recent Submissions" section shows two entries: "STRIDE v1.1.0 (household bubbles)" and "STRIDE universal testing COVID19". The "Top Viewed" section shows two entries: "SLF4Web - MPEG-DASH datasets of static light fields" and "Temporally consistent species differences in parasite infection but no evidence for rapid parasite-mediated speciation in Lake Victoria cichlid fish". At the bottom, there is a "Discover" section with a table of results.

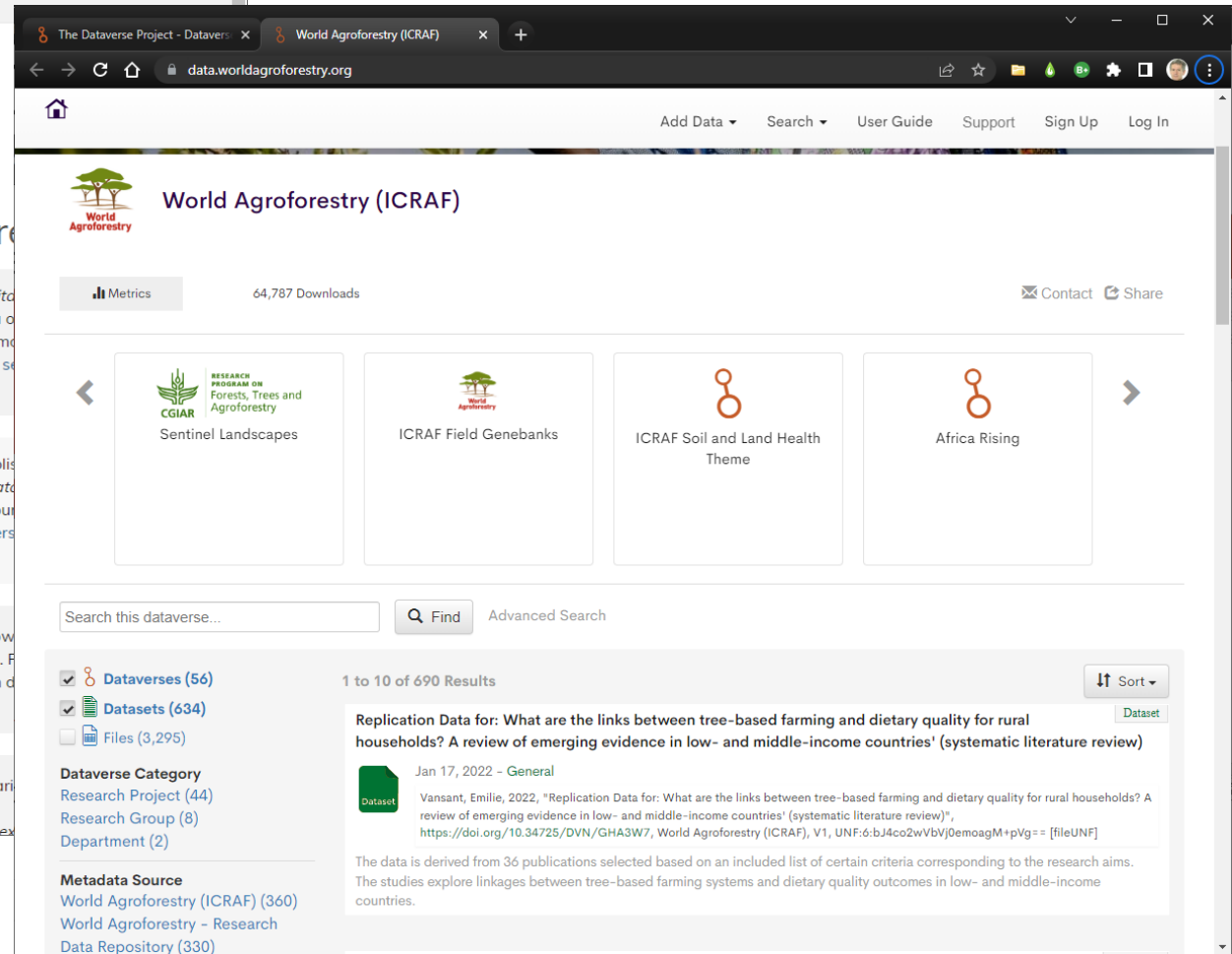
Contributor	Research Discipline	Date issued
HENS, Niel	Epidemiology	2022

2. Dataverse



The screenshot shows the homepage of the Dataverse Project. The header includes the logo and navigation links: About, Community, Best Practices, Software, and Contact. The main heading reads "The Dataverse Project" with the tagline "Open source research data repository software". Below this, there are four key features highlighted in grey boxes:

- Researchers:** Enjoy full control over your data. Receive *web visibility, academic credit, and increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on a personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up a personal Dataverse collection?
- Journals:** Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal's policy or list of repository recommendations. Want to find out more about journal Dataverse collections?
- Institutions:** Establish a research data management solution for your community. Federate with a growing network of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?
- Community:** Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and*



The screenshot shows the World Agroforestry (ICRAF) Dataverse repository page. The header includes the ICRAF logo and navigation links: Add Data, Search, User Guide, Support, Sign Up, and Log In. The main heading reads "World Agroforestry (ICRAF)" with a sub-heading "64,787 Downloads" and "Contact Share" links. Below this, there are four featured datasets in a carousel:

- CGIAR Research Program on Forests, Trees and Agroforestry: Sentinel Landscapes
- ICRAF Field Genebanks
- ICRAF Soil and Land Health Theme
- Africa Rising

Below the carousel is a search bar with the text "Search this dataverse..." and a "Find" button. To the right of the search bar is a link to "Advanced Search". Below the search bar, there are filters for "Dataverses (56)", "Datasets (634)", and "Files (3,295)". The main content area shows "1 to 10 of 690 Results" and a "Sort" dropdown menu. The first result is titled "Replication Data for: What are the links between tree-based farming and dietary quality for rural households? A review of emerging evidence in low- and middle-income countries' (systematic literature review)". The result includes the date "Jan 17, 2022 - General", the author "Vansant, Emilie, 2022", and the DOI "https://doi.org/10.34725/DVN/GHA3W7". Below the result is a brief description: "The data is derived from 36 publications selected based on an included list of certain criteria corresponding to the research aims. The studies explore linkages between tree-based farming systems and dietary quality outcomes in low- and middle-income countries."

3. Invenio RDM

The screenshot shows the homepage of Invenio Software. The header includes the Invenio logo and navigation links for Home, Products, Examples, and Community. The main content area features the Invenio logo with the tagline "Powering Open Science". Below this, there are two highlighted sections: "Framework" described as an "Open Source framework for large-scale digital repositories" and "RDM" described as a "Turn-key Research Data Management repository". At the bottom, there is a link to "New on the blog" with a right-pointing arrow.

The screenshot shows the homepage of HEPData. The header includes navigation links for About, Submission Help, File Formats, and Sign in. The main content area features the HEPData logo and the tagline "Repository for publication-related High-Energy Physics data". Below this, there is a search bar with the text "Search on 9720 publications and 108676 data tables." and a search button. A search example is provided: "e.g. reaction P P -> LQ LQ X, title has 'photon collisions', collaboration is LHCf or D0." Below the search bar, there is a section titled "Data from the LHC" with four sub-sections: ATLAS, ALICE, CMS, and LHCb, each with a "View Data" button. At the bottom, there is a section for "Recently Updated Submissions" with a "View all" link. A footer contains a cookie notice: "This website uses cookies to ensure you get the best experience on our website More info" and a "Got it!" button.

Sistema de información sobre la investigación actual

Current Research Information System

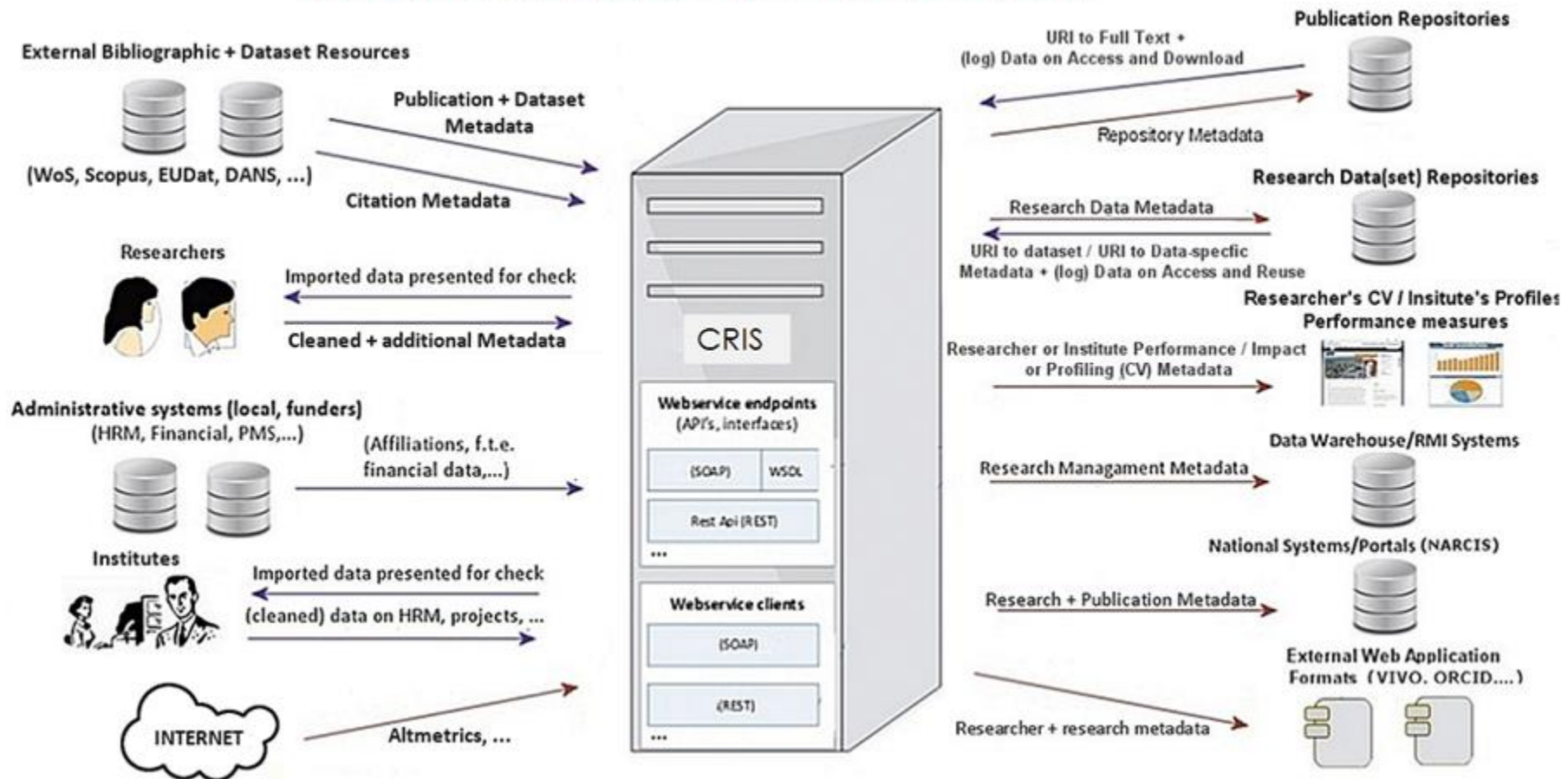
CRIS

CRIS Definición

- Un **sistema de información de investigación actual (CRIS)** es una base de datos u otro sistema de información para almacenar, gestionar e intercambiar metadatos contextuales para la actividad de investigación financiada por un financiador de investigación o realizada en una organización que realiza investigación (o su agregación).
- Los sistemas CRIS también se conocen como 'Research Information Management' or 'RIM Systems' (RIMS) (= sistemas de **gestión** de la información de investigación)

CRIS: Central en la gestión de la información de investigación de una organización de investigación

RI Infrastructure on an institutional level: Central position of CRIS



Ejemplo VIVO

The screenshot shows the AgriPerfiles website. The header includes the logo and the tagline "Conectando especialistas agrícolas de las Américas". A navigation menu contains "Inicio", "Personas", "Organizaciones", "Publicaciones", "Eventos", and "Mapa de capacidades". The main content area features a welcome message, a login section with fields for "Correo electrónico" and "Contraseña", and a search bar labeled "Buscar AgriPerfiles". Below the search bar, there are three columns: "Investigación" with counts for "Artículo académicos" (4,028), "Libros" (1,558), "Capítulos" (863), and "Paper de conferencias" (976); "Perfiles" listing individuals like "Fernández, Fernando"; and "Departamentos" listing institutions like "Estación Experimental Agropecuaria Cerro Azul INTA".

The screenshot shows the TIB VIVO website. The header features the "TIB VIVO" logo and navigation links for "Home", "People", "Organizations", "Research", "Events", and "Capability Map". The main heading is "Welcome to VIVO", followed by a search bar and a "Search" button. Below this, a paragraph describes VIVO as a research-focused discovery tool. At the bottom, three large icons represent "Who?", "Where?", and "What?". A footer note states, "This website uses cookies to ensure you get the best experience on our website. [More info.](#)" with a "Got it" button.

Ejemplo DSpaceCRIS

The screenshot shows the homepage of the HKU Scholars Hub. At the top, there is a navigation bar with the HKU logo and the text "The University of Hong Kong The HKU Scholars Hub 香港大學學術庫". Below this is a secondary navigation bar with links for Home, Publications, Researchers, Organizations, Grants, Datasets, Theses, Patents, and Community Service. The main content area features a large banner with the text "THE HKU SCHOLARS HUB AT THE CENTRE OF HKU" and a description of the hub's purpose. Below the banner are three columns: "Featured Scholar" featuring Professor Lam, Tai Hing; "Hub News" with a list of recent updates; and "Relevant Links" with a list of useful resources.

HKU Scholars Hub: Home

hub.hku.hk

The University of Hong Kong

The HKU Scholars Hub 香港大學學術庫

Home Publications Researchers Organizations Grants Datasets Theses Patents Community Service

HELP HKU Login Guest Login

THE HKU SCHOLARS HUB AT THE CENTRE OF HKU

The HKU Scholars Hub is the current research information system of The University of H... As a key vehicle of HKU's Knowledge Exchange Initiative, The Hub aims to enhance the of HKU authors and their research, and to foster opportunities for collaboration.

Quick Search ^{BETA} Research Collaborations Thesis Supervisors Media Commentators

Search for Everything...

Featured Scholar

Professor Lam, Tai Hing

- Robert Kotewall Professor in Public Health, School of Public Health
- Professor (Clinical)
- Sir Robert Kotewall Professor in Public Health
- Professor: Chair of Community Medicine

Research Interests

Hub News

- Mar 2022: HKU Theses On Amazon and Other Online Retailers (Update).
- Dec 2021: HKU Theses On Amazon and Other Online Retailers (Update).
- Sep 2021: HKU Theses On Amazon and Other Online Retailers (Update).
- Sep 2021: Announcing the 2021 HKU's Top 1% Scientists.

Relevant Links

- HKU's Top 1% Scientists
- Open Access @HKU
- Usage Stats & Downloads
- HKU research on Web of Science
- Research @HKU

The screenshot shows the homepage of the CONCYTEC Institutional Repository. The header includes the CONCYTEC logo and a navigation menu with links for Inicio, Comunidades y colecciones, Publicaciones, Investigadores, Organizaciones, Financiamientos, and Eventos. A prominent search bar is located in the center, with a "Término de búsqueda" input field and a "Busqueda Avanzada" dropdown. Below the search bar, there are three large circular icons representing different categories: Publications (2960), Financiamientos (3385), and Otras entidades (113). Each icon has a corresponding "Explorar" button below it.

Página de inicio | Repositorio CC

Niet beveiligd | repositorio.concytec.gob.pe

CONCYTEC CONSEJO NACIONAL DE CIENCIA, TECNOLOGÍA Y INNOVACIÓN TECNOLÓGICA

Iniciar sesión

Inicio Comunidades y colecciones Publicaciones Investigadores Organizaciones Financiamientos Eventos

Bienvenidos a CONCYTEC

Término de búsqueda Busqueda Avanzada

Bienvenido al Repositorio Institucional del Concytec

El Repositorio Institucional del Concytec tiene como objetivo permitir el libre acceso a la producción científica institucional, optimizando su visibilidad; así mismo garantizar la preservación y conservación de la información relacionada a la ciencia, tecnología e innovación.

Publicaciones, libros, artículos ... 2960

Financiamientos 3385

Otras entidades 113

Explorar publicaciones

Explorar financiamientos

Explorar otras entidades

Ejemplo IRINS

Faculty / Scientist
617

- Director: 1
- Professor: 252
- Associate Professor: 143
- Assistant Professor: 162
- Chair Professor: 44

Scholarly Resources
Publications: 29747
Patents: 966

Access Types:
Closed Access: 19045
Green OA: 16260
Gold OA: 1517
Hybrid OA: 906
Bronze OA: 705

Journal Articles: 20846
Conference / In Proceedings: 7254

Resource Impact: 39100 Citations

Faculty / Scientist
366

- Vice Chancellor: 1
- Professor: 128
- Associate Professor: 65
- Assistant Professor: 162
- Research Scholar: 2

Scholarly Resources
Publications: 7541
Patents: 19

Access Types:
Closed Access: 3345
Gold OA: 468
Green OA: 181
Bronze OA: 164

Journal Articles: 5634
Conference / In Proceedings: 1110

Resources Impact: 79334 Citations, 84706 Citations

<https://irins.org/irins/>

Observaciones finales

- Los datos de la investigación pueden salvar vidas, ayudar a desarrollar soluciones y maximizar nuestros conocimientos. Promover la colaboración y la cooperación entre una comunidad investigadora mundial es el primer paso para reducir la carga de la investigación desperdiciada.
- La presión para reducir la carga de la investigación desperdiciada está empujando a las revistas, a los financiadores y a las instituciones académicas a hacer del archivo y intercambio de datos un requisito estricto.
- El acceso abierto, los datos abiertos y la ciencia abierta se integran ahora en un enfoque de gestión de la investigación. Los sistemas CRIS son el exponente de esta evolución.