

**Data Sharing  
як елемент сервісів  
із управління даними  
досліджень  
в університетській бібліотеці**

Тетяна Борисова



# Управління даними

**досліджень** (Research Data Management) – частина дослідницького процесу, метою якої є організація даних від моменту їх залучення у дослідницький цикл до поширення та архівування результатів



# Складові управління даними

Створення даних та плану для їх використання

Організація, структурування та ідентифікація даних

Зберігання – захист даних, забезпечення до них доступу, збереження та копіювання

Поширення даних



# Які дані поширюють

Не всі дані, що створюються під час дослідження, потребують поширення. На етапі планування це потрібно врахувати заздалегідь.

Можливе поширення як неструктурованих (первинних даних), так і структурованих.

Дані можуть бути як відкритого, так і закритого чи обмеженого доступу.





# З якою метою поширюють дані

Убезпечити від втрати

Дати можливість іншим перевірити власні  
результати

Підтримка відкритого доступу

Надати можливість повторного використання

Збільшення цитувань

Політики університетів

**Вимоги грантодавців**

**Журнальні політики**



# Горизонт2020

“The European Research Council supports the basic principle of Open Access to research data. It therefore **recommends** to all its funded researchers that they follow best practice by retaining files of all the research data they have produced and used during the course of their work, and that they be prepared to share these data with other researchers whenever they are not bound by copyright restrictions, confidentiality requirements, or contractual clauses.”

[http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf)

European Research Council  
Executive Agency

Established by the European Commission



Horizon 2020  
European Union Funding  
for Research & Innovation

European Research Council (ERC)

Guidelines on

Implementation of Open Access to Scientific  
Publications and Research Data

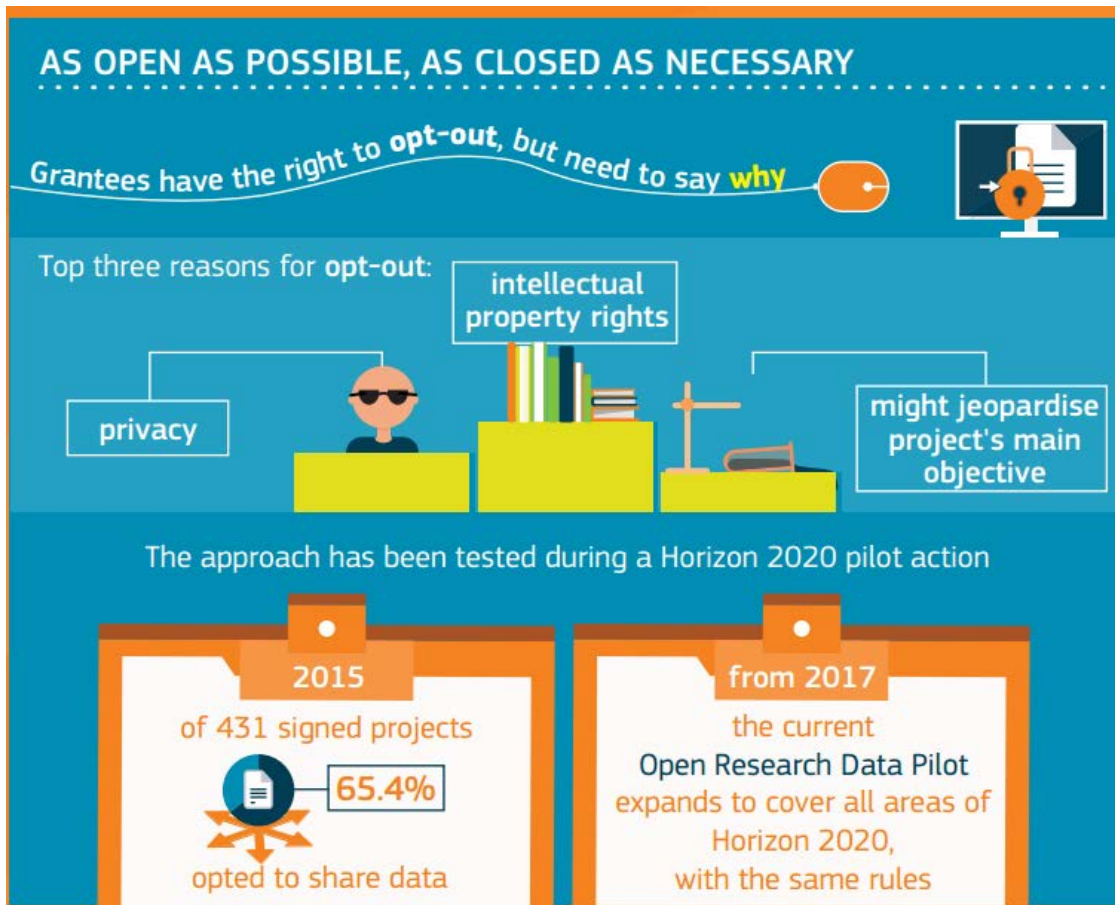
in projects supported by the European Research Council under Horizon 2020

Version 1.1  
21 April 2017

# Горизонт2020

Грантоотримувачі можуть не відкривати дані, проте мають обгрунтувати **ЧОМУ**.

[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf)





# Nature

Політика журналу **вимагає**,  
щоб усі матеріали з  
дослідження, код, дані та  
протоколи були доступними  
для читачів. Для великих  
наборів даних пропонується  
використання репозитаріїв.

The screenshot shows the 'authors & referees' section of the Nature website. At the top, there are navigation links for 'nature.com', 'Publications A-Z index', and 'Browse by subject'. On the right side, there are buttons for 'My account', 'Submit manuscript', 'Register', and 'Subscribe'. Below these, there are 'Login' and 'Cart' icons. The main header area features the text 'authors & referees' and a search bar with a 'go' button and a link to 'Advanced search'. The breadcrumb trail reads 'authors & referees > Policies > Availability of data, material and methods'. On the left, a 'Site content' sidebar lists various pages, with 'Availability of data & materials' highlighted. The main content area is titled 'Availability of data, material and methods' and contains the following text: 'An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must **also** be disclosed in the submitted manuscript. After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings. See sections below for details on: 

- reporting requirements
- availability of data
- availability of materials
- availability of computer code
- experimental protocols
- clinical trials
- further reading

 A 'SUBMIT NOW' button is visible on the right side of the page.

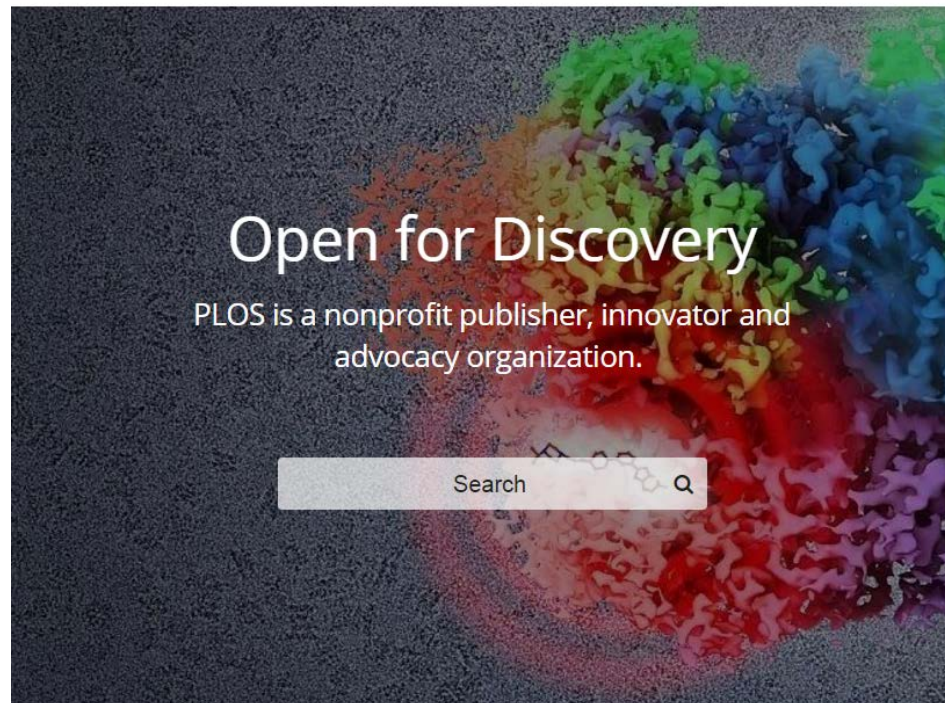
<http://www.nature.com/authors/policies/availability.html#data>

# PLOS

«Дані, що лежать в основі дослідження, **повинні бути** оприлюднені. Інші випадки мають узгоджуватися із редактором. Дані мають бути де-ідентифіковані»



Submit



<https://www.plos.org/editorial-publishing-policies>

# Scientific Data

Рецензований журнал, що публікує описи та аналіз науково важливих наборів даних. Статті називаються **Data Descriptors**, поєднують нарративну частину та структуровану (метадані), містить опис методології збору даних та технічного аналізу.



The screenshot shows the top navigation bar of the Scientific Data journal website. The header is dark blue with the journal title 'SCIENTIFIC DATA' and a logo. To the right are icons for Search, E-alert, Submit, and Login. Below the header is a light gray section titled 'Principles'. A dark gray navigation bar contains links for 'About', 'Principles', 'Open Access', 'FAQ', and 'Editors, Advisory Panel & Editorial Board'. The main content area has a white background with the following text:

Scientific Data is an open-access, online-only journal for descriptions of scientifically valuable datasets. Our articles, known as Data Descriptors, combine traditional narrative content with curated, structured descriptions (metadata) of the published data to provide a new framework for data-sharing and -reuse that we believe will ultimately accelerate the pace of scientific discovery. These principles are designed to align with and support the [FAIR Principles](#) for scientific data management and stewardship, which declare that research data should be **Findable, Accessible, Interoperable** and **Reusable**.

<https://www.nature.com/sdata/about/principles>

[Data Descriptor](#) | [OPEN](#)

# A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations

Jörg Franke , Stefan Brönnimann, Jonas Bhend & Yuri Brugnara*Scientific Data* **4**, Article number: 170076 (2017)

doi:10.1038/sdata.2017.76

[Download Citation](#)[Atmospheric dynamics](#) [Palaeoclimate](#)

Received: 07 October 2016

Accepted: 23 March 2017

Published online: 06 June 2017

## Abstract

Climatic variations at decadal scales such as phases of accelerated warming or weak monsoons have profound effects on society and economy. Studying these variations requires insights from the past. However, most current reconstructions provide either time series or fields of regional surface climate, which limit our understanding of the

[Abstract](#)[Background & Summary](#)[Methods](#)[Data Records](#)[Technical Validation](#)[Usage Notes](#)[Additional Information](#)[References](#)[Data Citations](#)[Acknowledgements](#)[Author information](#)

## Data Citations

1. Franke, J., Brönnimann, S., Bhend, J. & Brugnara, Y. *World Data Center for Climate at Deutsches Klimarechenzentrum* [http://dx.doi.org/10.1594/WDCC/EKF400\\_v1](http://dx.doi.org/10.1594/WDCC/EKF400_v1) (2017).

<https://www.nature.com/articles/sdata201776>

# Способи поширення даних

- депонування їх в банк даних чи спеціалізований архів
- подання в журнал разом із публікацією
- розміщення в інституційному репозитарії
- розміщення на сайті проекту чи інституційному сайті
- неофіційне поширення (персонально, на запит інших дослідників)

<https://www.nature.com/sdata/about/principles>

# На що звернути увагу, обираючи сховище для розміщення даних

Які формати підтримує, чи змінює якість даних під час розміщення?

Які метадані присвоює?

Чи є політики авторства, чи підтримує ліцензії?

Чи робить бекапи і як регулярно?

Чи присвоює ідентифікатор?

Чи рецензується?

Які можливості доступу передбачає?



# Re3data

Глобальний реєстр репозитаріїв даних досліджень з усіх дисциплін.

Запущений у травні 2013 року. На сьогодні надає інформацію про близько 2000 репозитаріїв.





The research data repository provides additional information on its service.



The research data repository was reviewed by the re3data.org Team.



The research data repository is either certified or supports a repository standard.



The research data repository provides a policy.



The research data repository provides open/restricted/closed access to its data.



The terms of use and licenses of the data are provided by the research data repository.



The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.





# Склад Re3data (за темою)

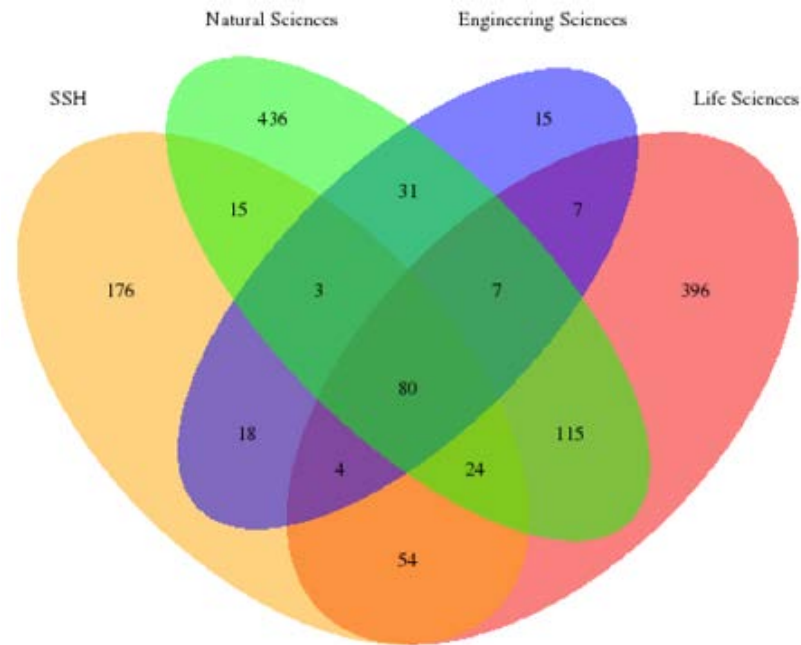
4 основні предметні категорії, на які розподіляються репозитарії в реєстрі:

Natural Sciences (51.5 %)

Life Sciences (49.8 %)

Humanities and Social Sciences (27.1 %)

Engineering Sciences (12.0 %)



The Landscape of Research Data Repositories in 2015: A re3data Analysis -

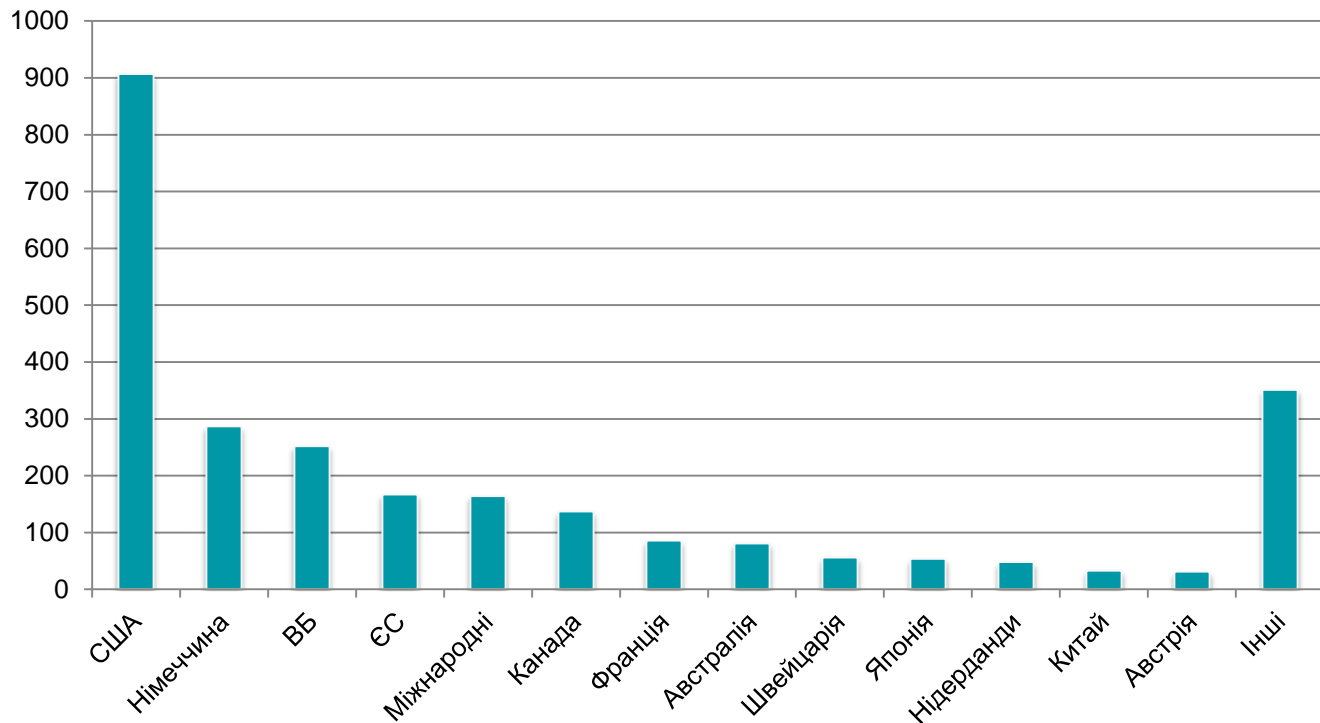
<http://www.dlib.org/dlib/march17/kindling/03kindling.html>

# Склад Re3data (за типом даних)

Типи даних	К-сть репозитаріїв
Наукові та статистичні дані	1174
Стандартні офісні документи	1114
Зображення	920
Звичайний текст	913
Необроблені дані	833
Структуровані графіки	707
Інше	682
Структурований текст	594

Типи даних	К-сть репозитаріїв
Архівовані дані	433
Аудіовізуальні дані	349
Програмні продукти	338
Бази даних	328
Networkbased data	113
Програмний код	85
Дані конфігурації	48

# Склад Re3data (за країною)



# Стандарти метаданих

ABCD - Access to Biological Collection Data (8)  
AVM - Astronomy Visualization Metadata (1)  
CF (Climate and Forecast) Metadata Conventions (23)  
CIF - Crystallographic Information Framework (1)  
CIM - Common Information Model (1)  
DCAT - Data Catalog Vocabulary (4)  
DDI - Data Documentation Initiative (102)  
DIF - Directory Interchange Format (27)  
Darwin Core (16)  
DataCite Metadata Schema (46)  
Dublin Core (135)  
EML - Ecological Metadata Language (19)  
FGDC/CSDGM - Federal Geographic Data Committee  
Content Standard for Digital Geospatial Metadata (60)  
FITS - Flexible Image Transport System (7)

Genome Metadata (1)  
ISA-Tab (4)  
ISO 19115 (98)  
International Virtual Observatory Alliance Technical  
Specifications (1)  
MIBBI - Minimum Information for Biological and Biomedical  
Investigations (5)  
OAI-ORE - Open Archives Initiative Object Reuse and  
Exchange (6)  
PROV (1)  
RDF Data Cube Vocabulary (18)  
Repository-Developed Metadata Schemas (15)  
SDMX - Statistical Data and Metadata Exchange (2)  
other (70)

# Dublin Core

Один із найпопулярніших стандартів, що використовується в репозитаріях.

<http://dublincore.org/>

The Metadata Community — Supporting Innovation in Metadata Design, Implementation & Best Practices



METADATA  
Dublin Core® Metadata Initiative  
INNOVATION



Home Metadata Basics DCMI Specifications Community and Events Join/Support About Us

Enter keyword  Search

### Upcoming DCMI Events

- » **May 24 (EN) & May 31 (PT):** Webinar: "Me4MAP: A method for the development of metadata application profiles" with Mariana Curado Malta [Information & registration](#) (Free to DCMI members)
- » **June 14:** Webinar: "How to Design & Build Semantic Applications with Linked Data" with Dave Clarke [Information & registration](#) (Free to DCMI members)
- » **Oct. 26-29:** [DC-2017](#), Washington, D.C. (Crystal City, Virginia), USA

---

DCMI is a project of



## Announcements

### Webinar: How to Design & Build Semantic Applications with Linked Data

2017-05-19, This webinar, presented by Dave Clarke, co-founder and CEO of the Synaptica® group of companies, will demonstrate how to design and build rich end-user search and discovery applications using Linked Data. The Linked Open Data cloud is a rapidly growing collection of publicly accessible resources, which can be adopted and reused to enrich both internal enterprise projects and public-facing information systems. The webinar will use the Linked Canvas application as its primary use-case. Linked Canvas is an application designed by Synaptica for the cultural heritage community. It enables high-resolution images of artworks and artifacts to be catalogued and subject indexed using Linked Data. The talk will demonstrate how property fields and relational predicates can be adopted from open data ontologies and metadata schemes, such as DCMI, SKOS, IIIF and the Web Annotation Model. Selections of properties and predicates can then be recombined to create Knowledge Organization Systems (KOS) customized for business applications. The demonstration will also illustrate how very-large-scale subject taxonomies and name authority files, such as the Library of Congress Name Authority File, DBpedia, and the Getty Linked Open Data Vocabularies collection, can be used for content enrichment and indexing.

To register and for more information about the webinar and presenter, visit <http://dublincore.org/resources/training/#2017clarke>.

# DDI - Data Documentation Initiative

Міжнародний стандарт, що використовується для опису опитувань, опитувальників, файлів зі статистичними даними та інші набори даних із соціо-економічних досліджень.

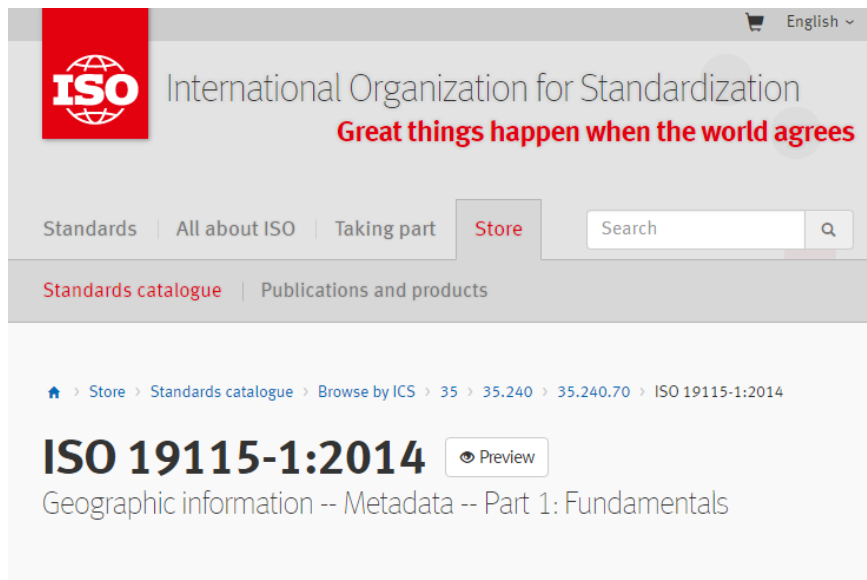
Останнє оновлення – 2014 р.



# ISO 19115

Стандарт, що описує метадані для геоінформаційних та пов'язаних даних.

<https://www.iso.org/standard/53798.html>



The screenshot shows the ISO website interface. At the top left is the ISO logo (a red square with a white globe icon and the letters 'ISO'). To its right is the text 'International Organization for Standardization' and the slogan 'Great things happen when the world agrees'. Below this is a navigation bar with links for 'Standards', 'All about ISO', 'Taking part', and 'Store'. A search bar is also present. The main content area shows the breadcrumb path: 'Home > Store > Standards catalogue > Browse by ICS > 35 > 35.240 > 35.240.70 > ISO 19115-1:2014'. The title 'ISO 19115-1:2014' is prominently displayed with a 'Preview' button. Below the title is the subtitle 'Geographic information -- Metadata -- Part 1: Fundamentals'.

ISO 19115-1:2014 defines the schema required for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services.

ISO 19115-1:2014 is applicable to:

- the cataloguing of all types of resources, clearinghouse activities, and the full description of datasets and services;
- geographic services, geographic datasets, dataset series, and individual geographic features and feature properties.

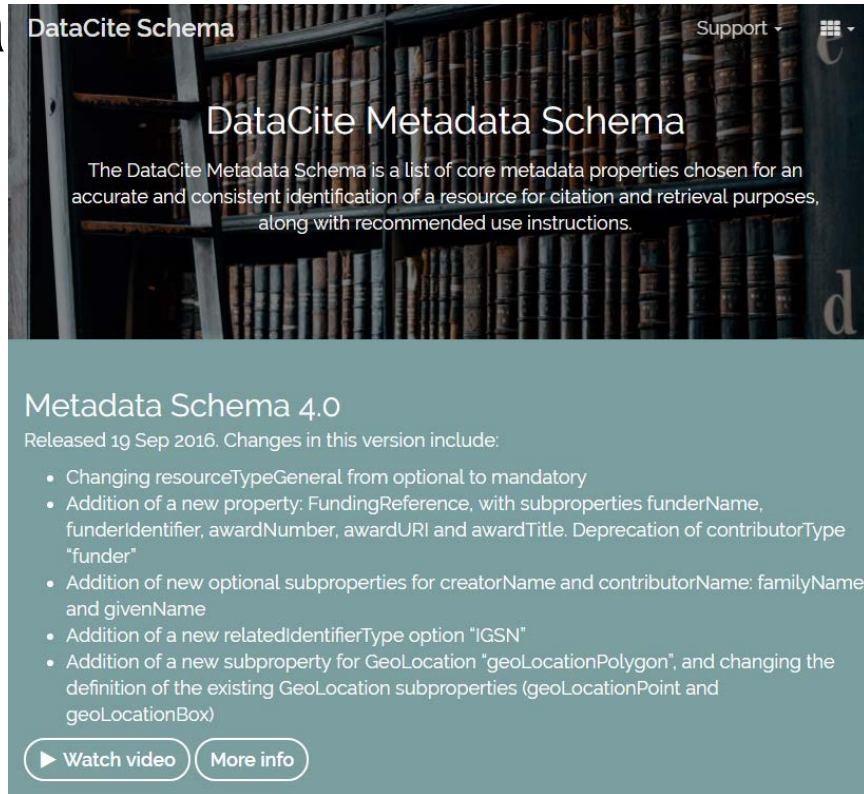
ISO 19115-1:2014 defines:

- mandatory and conditional metadata sections, metadata entities, and metadata elements.

# DataCite Metadata Schema

Набір метаданих, що має на меті  
забезпечити найкращу  
ідентифікацію, пошук та цитування  
для різних типів метаданих.

Інтегрується з іншими форматам  
та стандартами.



DataCite Schema

Support

## DataCite Metadata Schema

The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions.

### Metadata Schema 4.0

Released 19 Sep 2016. Changes in this version include:

- Changing resourceTypeGeneral from optional to mandatory
- Addition of a new property: FundingReference, with subproperties funderName, funderIdentifier, awardNumber, awardURI and awardTitle. Deprecation of contributorType "funder"
- Addition of new optional subproperties for creatorName and contributorName: familyName and givenName
- Addition of a new relatedIdentifierType option "IGSN"
- Addition of a new subproperty for GeoLocation "geoLocationPolygon", and changing the definition of the existing GeoLocation subproperties (geoLocationPoint and geoLocationBox)

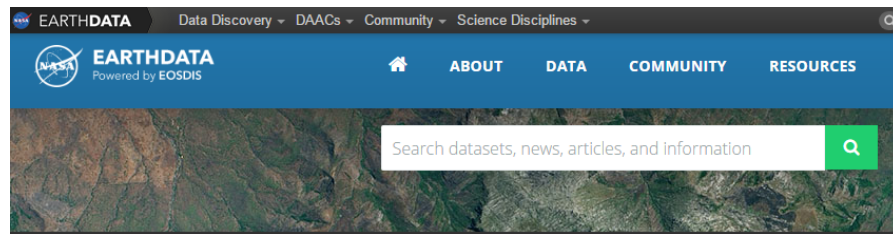
▶ Watch video More info

<https://schema.datacite.org/>



# Directory Interchange Format (DIF)

Формат, рекомендований NASA для представлення наукових даних про Землю: геоінформаційних, аерогеофізичних та ін. Один із найперших стандартів метаданих – виник як Каталог Інтероперабельності в 1987 р.



Standards ▾ Directory Interchange Format (DIF) Standard

ESD-Approved Standards for use in NASA

Earth Science Data Systems

HDF 5

HDF EOS 5

NetCDF Classic

NetCDF-4/HDF5 File Format

## Directory Interchange Format (DIF) Standard



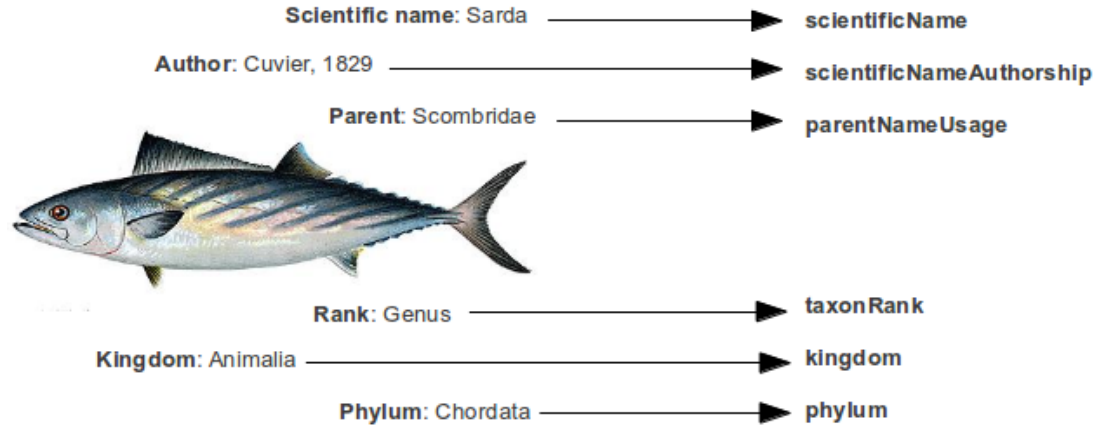
### Summary

This document defines the Directory Interchange Format (DIF) content metadata, a specific set of attributes for describing Earth science data. From its conception in the late 1980's as a way to document and exchange information on scientific data to its implementation in the Global Change Master Directory (GCMD), the DIF has evolved to serve the user community in the discovery, access and use of Earth science and related data.

<https://earthdata.nasa.gov/standards/directory-interchange-format-dif-standard>

# Darwin Core

Метадані, що спеціалізують інформацію про географічне розселення видів та існування зразків у колекції. Darwin Core складова Інформаційних стандартів із Біорізноманіття, що розробляється Міжнародним Союзом Біологічних наук.



# Мультидисциплінарні репозитарії



# Дякую за увагу!

**Тетяна Борисова,**  
Завідувачка відділу електронних ресурсів  
Наукова бібліотека Національного  
університету  
«Києво-Могилянська академія»  
[borysovato@ukma.edu.ua](mailto:borysovato@ukma.edu.ua)



II Міжнародна науково-практична конференція  
«Бібліотеки вищих навчальних закладів: досвід та перспективи»  
м. Одеса, 19-21 червня 2017 р.