Werkgroep datapublicatie en datacitatie

Inventariseerde
Publicatieplatforms
Richtlijnen m.b.t. citeren

maar…

het werd meer
want zoveel nieuwe ontwikkelingen
(Zie rapport)

Waarom data publiceren?

Het is goed !

Research data remains accurate, authentic, reliable and complete.

Duplication of effort is kept to a minimum.

Research data keeps its integrity and research results may be replicated.

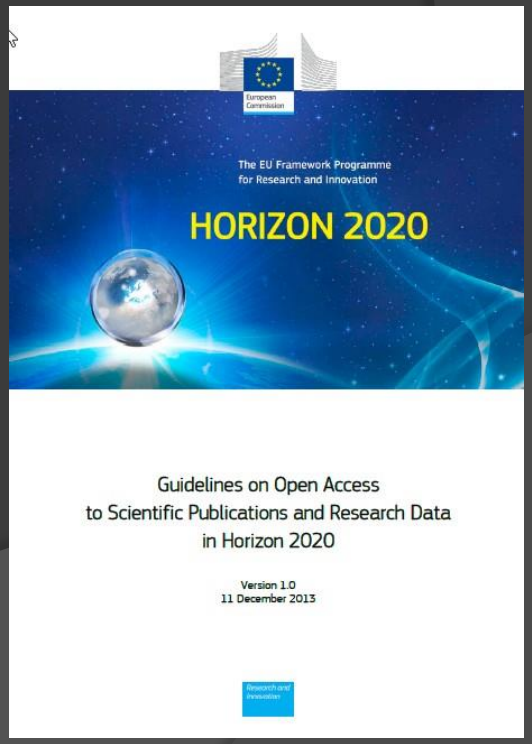
Data security is enhanced, thus minimising the risk of data loss.

(Oxford)

Het moet !

Tijdschriften:
PLOS ONE: “Authors must make all data publicly available, without restriction, immediately upon publication of the article.”

Onderzoeksfinancierders o.a. EU Horizon 2020
EUDAT



Waar zet je de data?

Data journals

Nature Scientific Data
Database: The Journal of Biological Databases and Curation

Repositories/archives

institutionele, nationale, internationale, disciplinaire, en multidisciplinaire

Verzamelingen van repositories:

Databib, DataCite, BioSharing en re3data

Artikelen linken aan data in repositories, voorbeeld Elsevier

Kwaliteit repositories

- DANS Data Seal of Approval
- Trusted Repositories Audit & Certification (TRAC).

Rankings and statistics:

Bijvoorbeeld DataCite

Data in open access?

OpenAIRE, Open Access Infrastructure for Research in Europe

- “Policy RECommendations for Open Access to Research Data in Europe (RECODE)
- Met CERN samen Zenedo repository

Met LIBER en COAR Horizon 2020 Pilot Open Research Data

COAR Confederation of Open Access Repositories

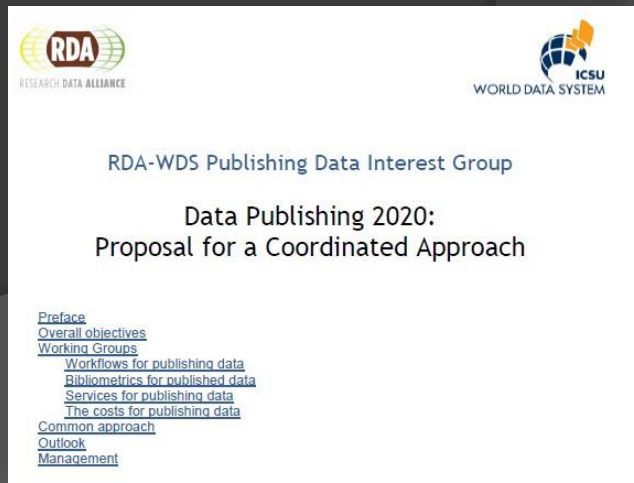
LIBER Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

Wat komt er bij kijken?

Verschillende initiatieven om het publiceren eenvoudiger te maken. Bijvoorbeeld bij de biodiversiteit

The Integrated Publishing Toolkit (IPT) is a free open source software tool written in Java that is used to publish and share biodiversity datasets through the GBIF network. Designed for interoperability, it enables the publishing of content in databases or text files using open standards namely the Darwin Core and the Ecological Metadata Language. You can also use a 'one-click' service to convert your metadata into a draft data paper manuscript for submission to a peer-reviewed journal.

RDA zoekt met werkgroep generieke oplossing

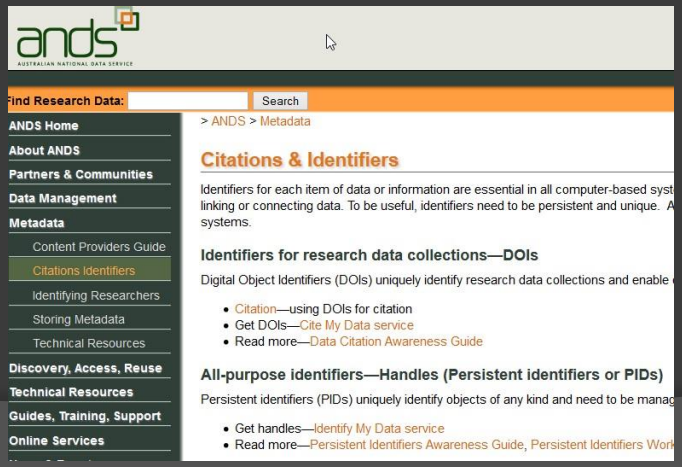


Standaarden, metadata en PID's

Wens is: Data delen en hergebruiken op een standard conforme manier

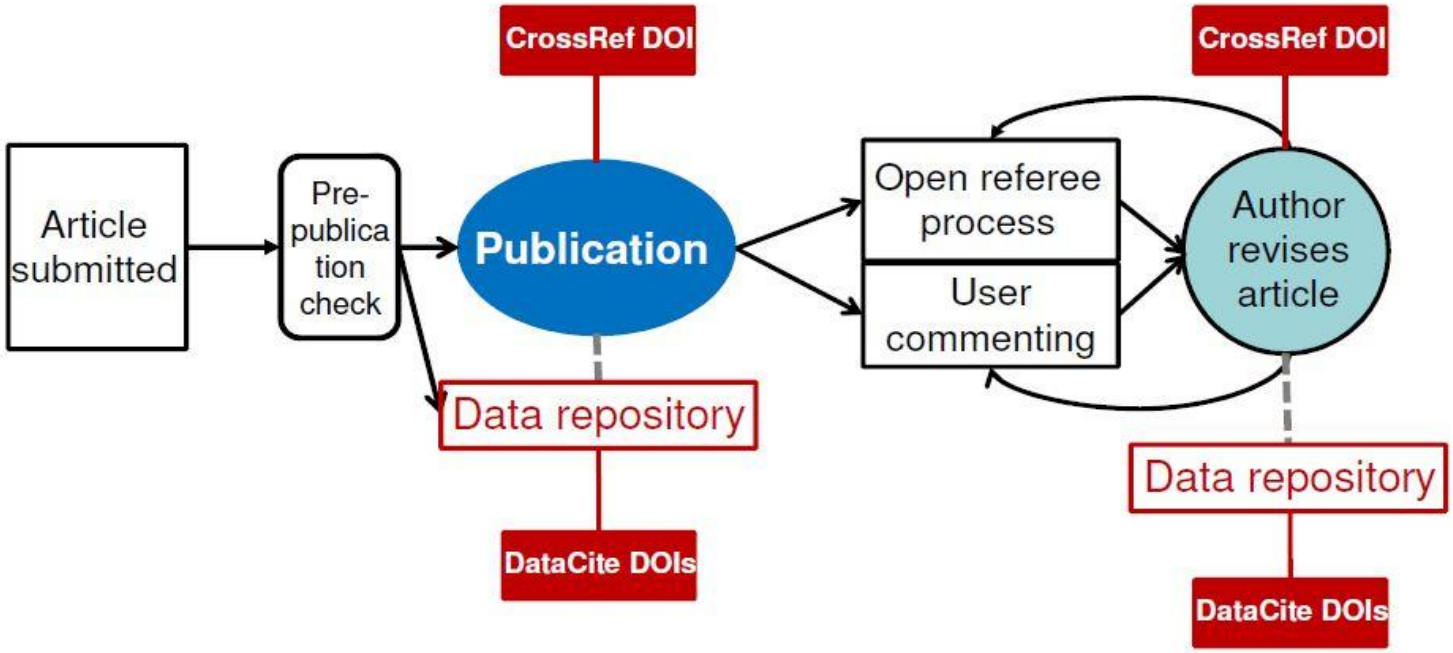
BioSharing bijvoorbeeld werkt aan een data standard voor de life sciences

Het Britse Digital Curation Centre geeft een overzicht van metadata voor de verschillende disciplines.



F1000

THE PUBLISHING PROCESS



F1000 FACULTY of 1000

Nog meer aspecten

- Data processing charges (dpc's)
  - Dryad bijvoorbeeld
- Peer reviewing
  - PREPARDE Peer REview for Publication & Accreditation of Research data in the Earth sciences.
- Services (cross referencing)
- Cost for publishing data
- Licensing, usage rights, and Privacy

Citeren van data

“Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice.”

Draft Declaration of Data Citation Principles van de

Data Citation Synthesis Group

Data citation principles

Veel guidelines

Veel organisaties CODATA-ICSTI, RDA, ICSU, W3C, DataCite, OpenAire en vele anderen

Force11 Group Amsterdam Manifesto

juli 2013 de “Data Citation Synthesis Group”.

Joint Declaration of Data Citation Principles

“…..to produce a consolidated set of data citation principles…

De declaratie kent de volgende principes:

1. Importance
Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.

2. Credit and Attribution
Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

3. Evidence
In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

4. Unique Identification
A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.

5. Access
Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

6. Persistence
Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.

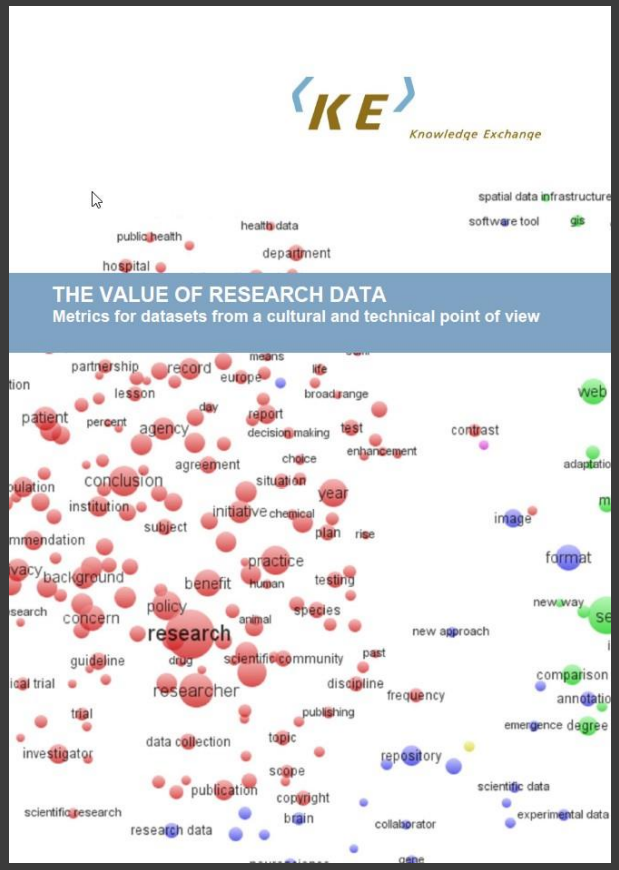
7. Specificity and Verifiability
Data citations should facilitate identification of, access to, and verfication of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verfiying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.

8. Interoperability and flexibility
Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities.

Complexiteit van het citeren

deep citation (welke subset?)
granularity (onderverdeling van velden)
dynamic data (versions)

Data metrics



Bibliometrie voor gepubliceerde data

Andere indicatoren?

“At present there is insufficient experience with alternative metrics and therefore it is hard to judge their value”.

- CWTS rapport
- The Data Citation Index on the Web of Science™ platform

Alles verbinden

ODIN
ORCID and DataCite Interoperability Network

Het ODIN project:

“ODIN will build on the ORCID and DataCite initiatives to uniquely identify scientists and data sets and connect this information across multiple services and infrastructures for scholarly communication. It will address some of the critical open questions in the area:

- Referencing a data object
- Tracking of use and re-use
- Links between a data object, subsets, articles, rights statements and every person involved in its life-cycle.