USER PERSPECTIVE TO OPEN DATA

Urmas Kõljalg
University of Tartu

Open Forest Science, 7 May 2018, Helsinki

Not just Open Data but FAIR Data are needed today.

FAIR DATA
Findable, Accessible, Interoperable and Re-usable
Common way to the Open Data

Often this means no FAIR Data because data are stored in repositories as files in different formats. Additional effort needed to recover data from files.

Creating datasets
Datasets are created with spreadsheets, local database, etc.

Analysing datasets
Datasets are analysed with great variety of software. Data must be usually converted first into suitable format.

Publishing paper
Results of the analyses are presented in the paper. Datasets are sometimes attached as a Supplementary material. Often it means that datasets are in different file formats and are not machine-readable. But they are Open Data s.l.

Some journals ask to lodge datasets in public repository like DataONE, etc. In this case again files in different formats are uploaded. Files and metadata are published with DOI.

Archiving

Data Standards (biodiversity)

TDWG
Biodiversity Information Standards
http://www.tdwg.org
development of standards for the exchange of biological/biodiversity data, most important Darwin Core and ABCD

GSC
Genomic Standard Consortium
http://gensc.org
The aim of the GSC is making genomic data discoverable
NCBI (GenBank) implemented MiXS for sample data

EML
Ecological Metadata Language
https://knb.ecoinformatics.org/#external/emlpars/docs/index.html
Machine readable FAIR data

GBIF
Global Biodiversity Information Facility
https://www.gbif.org

INSDC
International Nucleotide Sequence Database Collaboration
http://www.insdc.org

GBIF
Free and open access to biodiversity data

Occurrence records: 978,494,334
Datasets: 38,893
Publishing institutions: 1,176
Species: Learn more about the number of species covered by data in GBIF.org
PlutoF - online data management platform which secure FAIR data

Full data lifecycle from Data Management Plan to the publishing and archiving your datasets

**Data Lifecycle with PlutoF**

Registration and login at https://plutof.ut.ee

**Create**

Datasets are created in PlutoF either by importing data with template csv files or by uploading data manually via online workbench. Mobile devices can be used for the fast and simple data upload via application Legulus (https://legulus.tools/#/).

**Manage/Curate**

Data management and curation is carried in online workbench. sit amet, at pede luctus vel gravida, nonummy massa rutrum lorem volutpat, a nunc ornare, etiam eleifend, fusce id mauris phasellus fusce ultriciesi.

**Publish**

There are several ways you may publish your data via PlutoF. You may export data in different formats and publish as a files. You may ask DOI for your dataset or publish data in GBIF. Then DOI is issued by GBIF. Or you can send your data directly to the Pensoft data journal.

**Archive**

PlutoF serves as an archive as well because you may store your datasets here for the unlimited time. Your datasets can be stored as such or they can be published with DOI. Dataset without DOI can be managed and changed if needed in future.

**Share**

In PlutoF you can share your datasets with any other user or workgroup. You can give simple access to your data or even right to edit your data. In this way you may collaborate with other scientists, students and citizen scientists.
Example of making data FAIR
Global Dataset of Oligochaeta
Dataset developed by Tarmo Timm
Free and open access to biodiversity data

Occurrence records 856,072,739
Datasets 36,831
Publishing institutions 1,118
Species Learn more about the number of species covered by data in GBIF.org.

Administrative job opportunity at the GBIF Secretariat
22 October 2017

Exploring patterns and gaps in plant diversity of tropical Africa
20 October 2017

Tanya Abrahamse elected chair of GBIF Governing Board
12 October 2017

Thomas Ormel elected chair of GBIF Science Committee
16 October 2017

World distribution of the aquatic Oligochaeta
Published by PlutoF
Tamis Tamm + Keezy Abamenkov

This is a database of published findings of aquatic Oligochaeta (Annelida, Clitellata) species, including the literature sources and distribution maps. Data on the Aphanoezus (Annelida, Polychaeta) and Branchiobdellida (Annelida, Clitellata) are included as well, since these worms have often been listed as oligochaetes. The database is still in progress. Registration of findings from territory of the former USSR will be completed in 2017. In the subsequent years, data from the remaining world will be treated in alphabetical order of the authors of publications. Countess of the species occurrences in water bodies are found by author from literature or by the aid of internet.

32,876 Occurrences 100% with taxon match 92% with coordinates 0.01% with year

38,351 GIDEREFERENCED RECORDS

PlutoF
Last Modified 22 October 2017
License: CC BY 4.0
How to cite: 10.15468/2ymiu

Nearly 500 000 DOIs (Digital Object Identifier) released. Second largest data center in the world in number of released DOIs (DataCite consortium)

PlutoF statistics