Koninklijk Meteorologisch Instituut

Institut Royal Météorologique

Königliches Meteorologisches Institut

Royal Meteorological Institute

Open Science: an overview

Pierre de Buyl

2022/03/10

Outline

Licensing

- This presentation: Pierre de Buyl (ORCID: 0000-0002-6640-6463) License CC BY
- Reused materials:
  - Open-Science building blocks figure by Gema Bueno de la Fuente License CC BY
  - XKCD cartoon https://xkcd.com/2347/ License CC BY-NC
  - Open Access schema from https://openaccess.be/ adapted from https://www.fosteropenscience.eu/ License CC BY

- ▶ Open science: Open Access, Open Source Software, Open Data, Open Educational Resources
- ▶ UNESCO open science recommendation
- ▶ Open Source & Open Data recognized by EU as a competitive advantage for the economy
- ▶ INSPIRE directive
- ▶ Open Science European Conference
- ▶ Glossary of Open Science

Building blocks
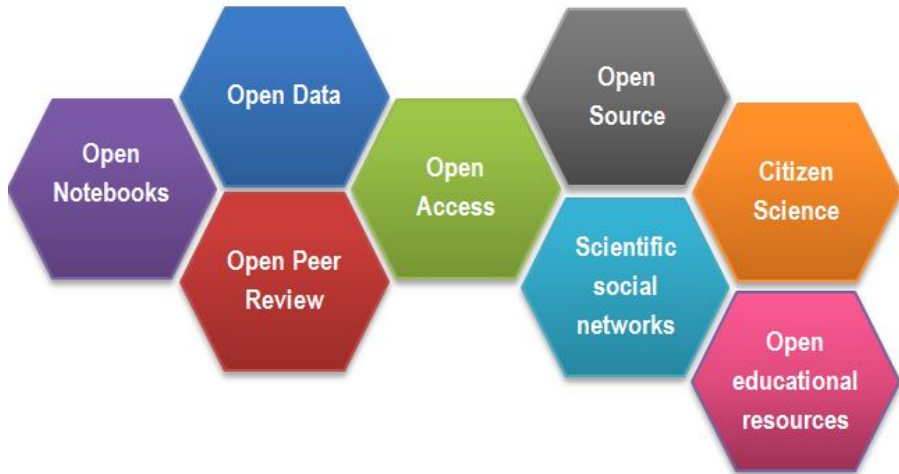
▶ Define Open Science

▶ Present tools for the practice of Open Science

▶ Publishing Open Access articles and Open Data

▶ A look at RMI & Belgium for Open Science

- ▶ Citizen Science - crowdsourcing: WOW and VLINDER
- ▶ Publications
- ▶ Open Data portal https://opendata.meteo.be/
- ▶ Some programs are open source
- ▶ Other initiatives I might not be aware of.

Flemish Open Science Board

Vlaamse beleidsplan Open Science en op Open Science Board goedgekeurd

📋 **Nieuws**

6 januari 2020

Vlaanderen

Op initiatief van Vlaams minister Hilde Crevits keurde de Vlaamse Regering voor het kerstverlof het Vlaams beleidsplan Open Science en de oprichting Open Science Board goed (*het document kan onderaan de pagina gedownload worden*). Vlaanderen zal 5 miljoen euro recurrent investeren in Open Science om invulling te geven aan het Europese engagement.

Europa

Open Science, inclusief Open Data, vormt een belangrijke motor van verandering in de wetenschap. Een nieuwe wijze van onderzoek vraagt steeds meer om het delen en open

Open Science Fund

Met het Open Science Fund wil NWO onderzoekers ondersteunen bij het ontwikkelen, testen en implementeren van innovatieve manieren om onderzoek open, vrij toegankelijk, transparant en herbruikbaar te maken. Hiermee zet NWO een stap voorwaarts om de manier waarop wetenschappers in Nederland worden erkend en gewaardeerd, te veranderen.

Doel en doelstellingen ⊕

Budget en doorlooptijd ⊖

NWO stelt 500.000 euro per jaar beschikbaar voor het Open Science Fund, met een maximum van 50.000 euro per project. Dit programma loopt tot en met 2023.

Chan Zuckerberg Initiative ⊕

ABOUT US   OUR INITIATIVES   HOW WE WORK

CLOSED

REQUEST FOR APPLICATIONS

OPEN SCIENCE

Essential Open Source Software for Science (Cycle 4)

This RFA supports the maintenance, growth, development, and community engagement of open source software projects to help make the computational foundations of biological research more usable and robust.

Total Grant Resources Cycle 4 - $11.1 Million

Commission makes software available to all to benefit businesses, innovators and areas of public interest

Some extra motivation for Open Science

- Resources accessible to all in digital form
- Transparency / reproducibility
- Pooling of development/research resources
- Funder requirement (FWO, FNRS, BELSPO)
- Get rid of "Data available upon request"
- Multiple p.o.v.: technical, cultural, legal, scientific, ethics
- Create communities (SciPy, Wikipedia, Linux)
- COVID19 response: sharing of RNA sequences, preprints, epidemiologic data

- Linux & Android
- FreeBSD: foundation for Mac OS, Netflix CDN
- The Apache web server & PHP, Perl, MySQL/MariaDB
- git distributed version control
- OpenVPN, Firefox, Chromium, LibreOffice
- The Belgium e-ID tools
- Many scientific tools

Linux

- ▶ The Linux Foundation
- ▶ Sponsored by the industry
- ▶ Several M$/year of donations by foundation members
- ▶ Thousands of contributors
- ▶ Windows provides "Linux"
- ▶ Linux on Mars! (also Python)

From XKCD https://xkcd.com/2347/ CC-BY-NC
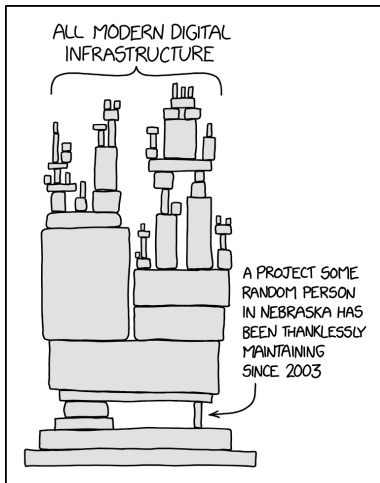
OpenSSL

- Before the Heartbleed vulnerability, donations of about 2k$/y
- Less than twenty developers
- Used by most web servers
- Since Heartbleed, the situation improved

Purpose of an OSS license

- ▶ The license is a contract between the software author and the users.
- ▶ The OSS license enables re-use and community.
- ▶ Publishing code on a webpage or sending it via email without a license places the user in a unclear situation.
- ▶ Pick a known OSI license.

OSS licenses

- **OSI licenses** Two major families:
  - GNU GPL: derivative work must be made available under the GPL. Linux, GCC, Emacs, Octave, GSL.
  - BSD/MIT: derivative work must only include license info. Python, GotoBLAS/OpenBLAS, NumPy, scikit-learn
- License choice best left to authors.
- See matplotlib's documentation for an interesting discussion

- ▶ git / GitHub / GitLab: version control and collaboration
- ▶ Institutional repositories (see Orfeo) / forges (see gitlab)
- ▶ Zenodo: archive articles, data, software
- ▶ ORCID: unique ID for scientists
- ▶ Notebooks

- No licensing trouble for install (how many Matlab licenses do we have?)
- Efficient sharing of code $\rightarrow$ collaboration
- Facilitate reproducible research / transparency (i.e. also document the code for you in the future)
- Documentation
- Recognize scientific software as legitimate output of a scientist

Introduction

Open-source Software

Open Access & Open Data

Open science in weather & climate science

Concluding remarks

- ▶ Findable: identifier / DOI / indexed
- ▶ Accessible: retrievable
- ▶ Interoperable: formal, accessible, shared, and broadly applicable language
- ▶ Reusable: provenance / license

(1. Personal "translation". 2. The above also applies to software.)

- ▶ EU directive to mandate Open Data
- ▶ Members state put into law
- ▶ See https://data.gov.be/en/info-faq
- ▶ Includes links to meteo.be !
- ▶ The governement focused on administrative data (ex: share of renewable energy, population)
- ▶ Best practice document plenty of good intentions, reality more random: invalid links or bad files

- ▶ Open access journals:
  - ▶ Models: Gold, Green, Hybrid, Bronze, Diamond (see Wikipedia).
  - ▶ Author processing charge (APC)
- ▶ Preprint servers by field (e.g. arXiv) or by institution (e.g. orfeo).
- ▶ Publishing a pre/post-print is Open Access (Green).
- ▶ See Open Access at BELSPO, Open Science in Belgium and the Belgian law on Green OA
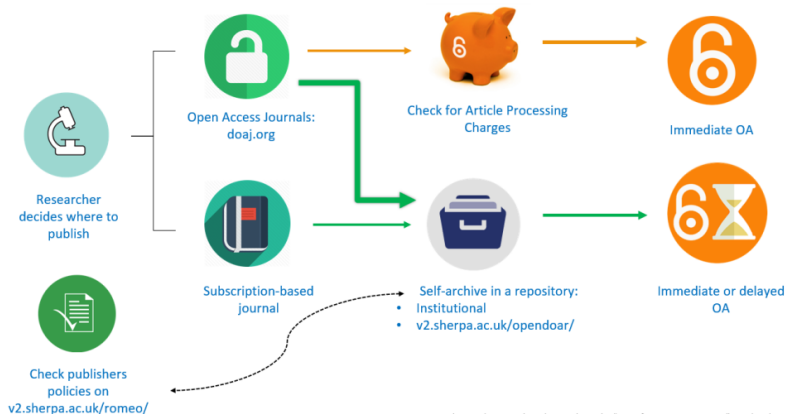
This is a derivative based on a scheme by "www.fosteropenscience.eu", used under CC BY.

From https://openaccess.be/ adapted from
https://www.fosteropenscience.eu/ CC-BY

- Open Access is a liberating opportunity for scientific communities (e.g. journal Algebraic Combinatorics that "fled" from Elsevier).
- The publishing industry is looking at ways to grow their revenue by pushing for Gold OA and large author fees (9500€ for Nature).
- Beware of predatory publishers.

Orfeo

- Orfeo is an institutional repository.
  https://orfeo.belnet.be/
- Part of the OA strategy of BELSPO.
- Scientists must deposit article metadata (including DOI etc) and (embargoed) full-text pdf (author typeset version).
- After maximum one year, the article must be readable via Orfeo.

ORCID

▶ Open Researcher and Contributor ID
▶ Non-profit org in which publishers, funders, etc are members
▶ Solves the identification problem due to growing number of authors
▶ Infrastructure for open science (can be used to log in to zenodo for instance)
▶ https://orcid.org/0000-0002-6640-6463
▶ https://orcid.org/orcid-search/search?searchQuery=GRID:grid.424737.1

- **Creative Commons** offers a modular set of license
- BY: by attribution
- NC: Non-Commercial
- SA: Share-Alike

CC - examples

- Wikipedia pages are under the CC-BY-SA
- Articles in *Remote Sensing* are under the CC-BY
- RMI open data portal: In progress (synoptic under CC-BY for instance).
- Belspo recommends CC0 or CC-BY as default for Open Data.
- XKCD comics are under the CC-BY-NC
- ECMWF Data Store (ECPDS): CC-BY (+dubious disclaimers)

Examples of open-source programs

- The Weather Research and Forecasting (WRF) model is under the public domain
- ADAGUC: open-source (Apache 2.0 for server) GIS viewer developed at KNMI. Viewer for opendata.meteo.be
- RMI: qgs: A 2-layer quasi-geostrophic atmospheric model in Python. Demaeyer, De Cruz and Vannitsem. J. Open Source Software, 5(56), 2597, DOI:10.21105/joss.02597.

▶ Open Science at the Finnish Meteorological Institute
▶ GitHub profiles: FMI, KNMI, DWD
▶ EUMETSAT: EUMETSAT open-source resources , CM SAF GitHub page
▶ ECMWF Data Store (ECPDS)
▶ NASA's Open Data, Services, and Software Policies 2023 is year of Open Science at NASA!

Finnish Meteorological Institute

Open Science at FMI

Openness is one of the defining characteristics of modern science. Open science crusade has been unfolding extensively over the past years to improve the credibility and reproducibility of science. Key domains of open practices include publishing open access and publicly sharing data and software.

Finnish Meteorological Institute (FMI) is committed to promoting the Open Science principles and good practices at our institute, opening, whenever is possible, the entire research process by enabling the reuse of all publicly funded FMI research output.

FMI encourages its researchers to seek and improve their knowledge in Open Science practices. To support this development, FMI provides its researchers the tools and proper training to build appropriate skill sets on Open Science. Researchers – both young and venerable - need infrastructure, support, and time to implement Open Science principles into their work.

On this page we describe the FMI's services for open access to data, publications and software developed, based on the national and EU recommendations as well as funding agencies.

- http://pytroll.github.io/
- Five year Memorandom of Understanding between Swedish Meteorological and Hydrological Institute, Finnish Meteorological Institute, KNMI, DWD, and MeteoSwiss
- Development of satpy and pyresample

ecCodes installation

Created by Daniel Varela Santoalla, last modified by Shahram Najm on Jan 21, 2021

Overview

ecCodes uses **CMake** for compilation and installation. This is a first step towards an homogenisation of the installation procedures for a

Like autotools, CMake will run some tests on the user's system to find out if required third-party software libraries are available and note

CMake is a cross-platform free software program for managing the build process of software using a compiler-independent method.

Generating the Makefiles with CMake

One nice and highly recommended feature of CMake is the ability to do **out of source** builds. In this way you can make all your ".o" files, pointing it to the source directory and using your own options.

The command gives feedback on what requirements are fulfilled and what software is still required. The following table gives an overvie Note: When an option is ON by default, it is enabled even if not explicitly requested. Features with default OFF need to be explicitly enabl

cmake options	doc
CMAKE_INSTALL_PREFIX	where you want to install your ecCodes
CMAKE_BUILD_TYPE	to select the type of compilation: - Debug - RelWithDebInfo

- ► `conda install -c conda-forge eccodes`
    - ► eccodes Open-source (Apache 2.0) code from ECMWF
    - ► `conda`: Open-source Python (+ compiled ext.) package manager
    - ► `conda-forge`: A community led collection of recipes, build infrastructure and distributions for conda. Supported by NumFOCUS org.
    - ► Used in statistical post-processing of forecasts at RMI

Overview

Concluding remarks

- Broad recognition *in principle* by politics
- Open Science enables personal or team initiatives (free).
- Also a matter of institutional policy (CV assessment, intellectual property, funding)
- "Open" is not equal to "zero cost"
- When there is no motivation or funding, open science does not fulfill its promise.
- Ideally, start a project with openness in mind.
- Reproducibility would require another seminar.