

opening a window *into a whole new world*

lennart martens

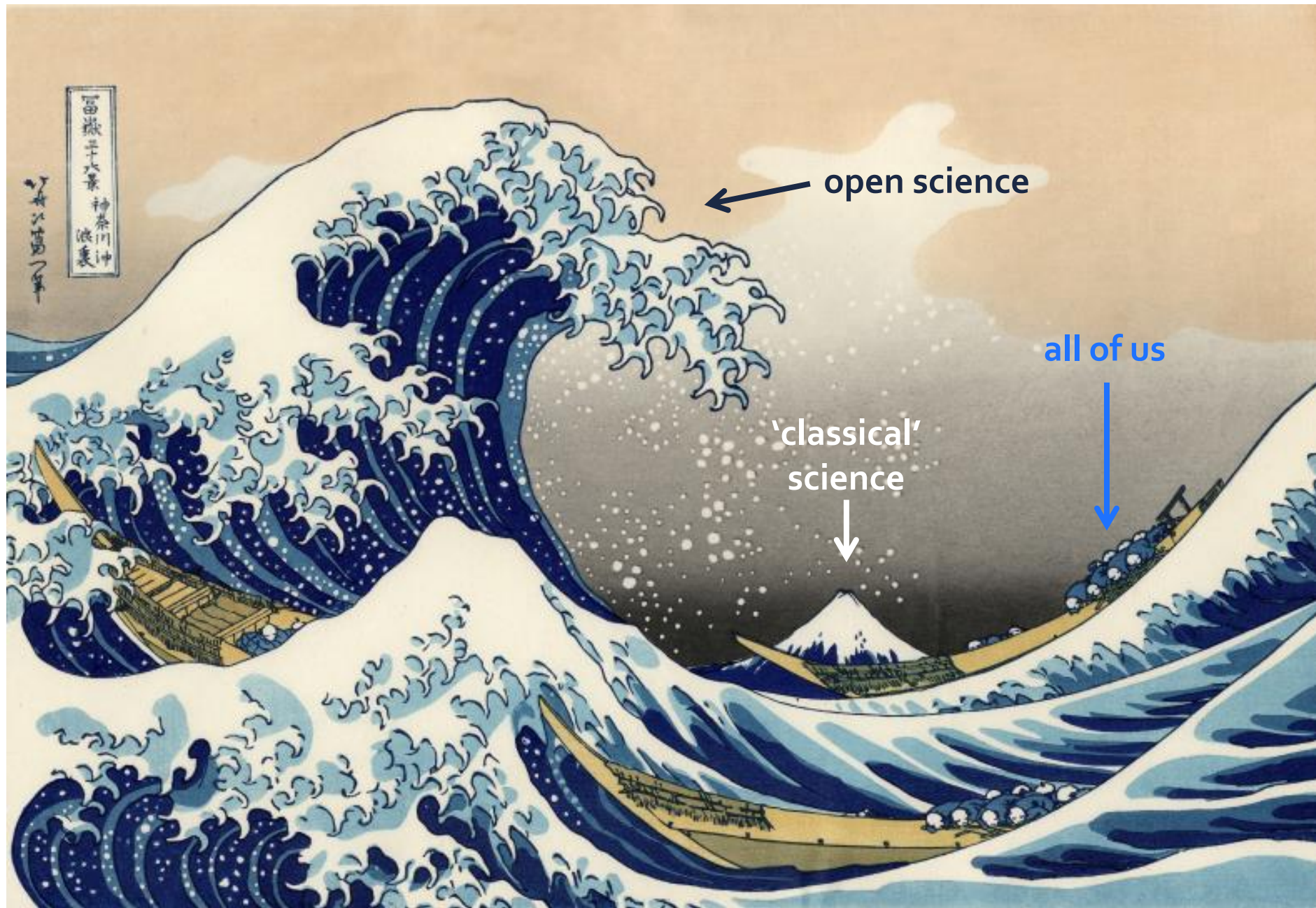
lennart.martens@ugent.be

computational omics and systems biology group

Ghent University and VIB, Ghent, Belgium



CC BY-SA 4.0



"The Great Wave off Kanagawa", by Hokusai, ~1830



open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

Open science is both mandated by funders and journals, but can be a bit of a minefield legally



English

Search

European Research Executive Agency

Home | Funding and grants | Working for REA | About REA | News | Events

European Commission > European Research Executive Agency > Open science

Open science

Q&A on what you should comply with when applying for funding and implementing your project

PAGE CONTENTS

[Open science in Horizon Europe](#)

[Open science and the proposal application](#)

[Open science and project implementation](#)

[Open science and Intellectual Property Rights \(IPR\)](#)

Open science in Horizon Europe

Did you know that open science is a legal obligation under [Horizon Europe](#)? Its purpose is to foster greater transparency and trust for the benefit of scientific research and for the benefit of EU citizens.

Confused or unsure about how to comply with open science principles when applying for EU funding and when implementing your project?

Fear not! REA has prepared an information package and series of Q&As below. This may help you to successfully implement open science practices in your proposals and during your project if your proposal is selected for funding.

Where are the data?

Nature Biotechnology now requires data availability statements to be supplied with research papers.

As the research community embraces data sharing, academic journals can do their bit to help. Starting this month, all research papers published in *Nature Biotechnology*, *Nature* and 11 other Nature titles will include information on whether and how others can access the underlying data.

These statements will report the availability of the 'minimal data set'

PROTECT PERSONAL DATA, PROTECT YOUR BUSINESS

25 May 2018:
new EU data protection
rules apply

Make sure
your business is ready
europa.eu/dataprotection
#GDPR

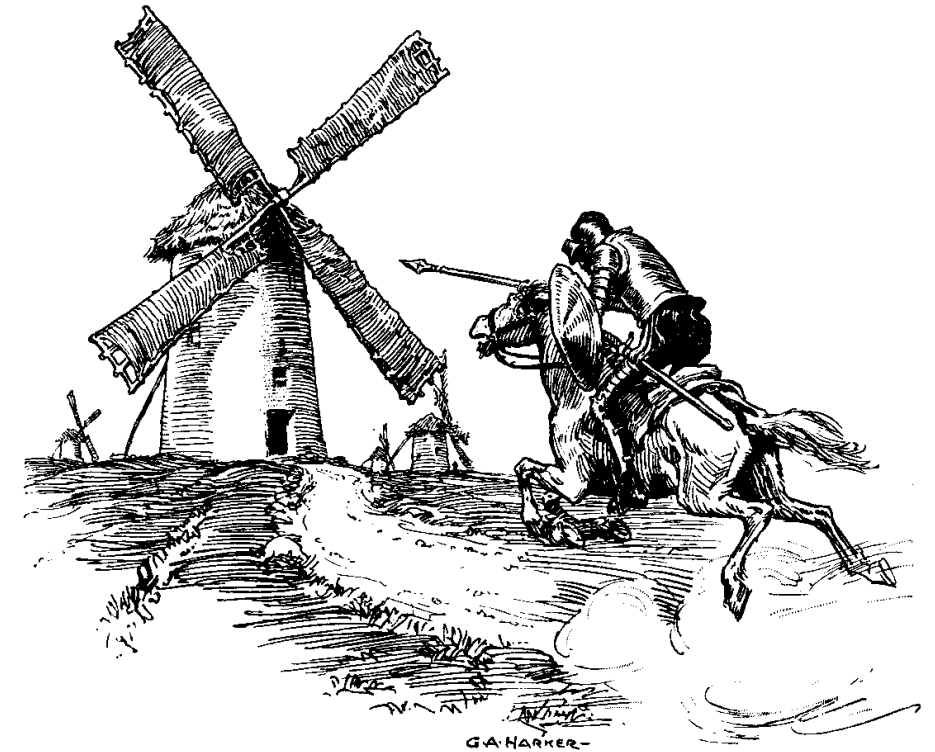


CC BY-SA 4.0

Whatever the context, open science is here today,
here to stay, and will be bigger tomorrow



Saint George and the Dragon
about 1470, Paolo Uccello

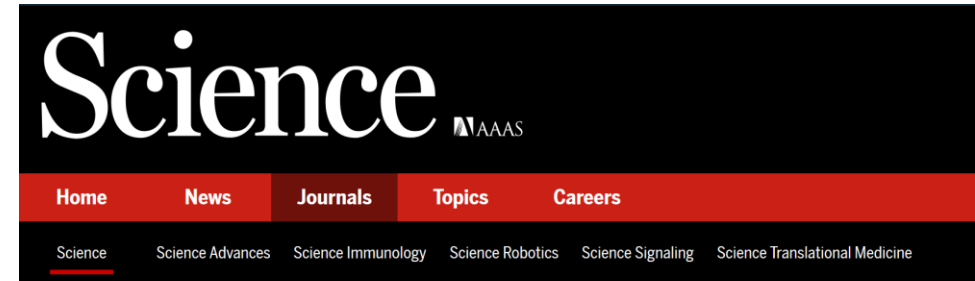


<https://www.flickr.com/photos/scriptingnews/3503448168>
(CC BY-SA 2.0)

Open science is frequently linked to scientific integrity, and absence of data can be a very strong red flag



After an investigation, the Central Ethical Review Board in Sweden has recommended the retraction of the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology,” by Oona M. Lönnstedt and Peter Eklöv, published in *Science* on 3 June 2016 (1). *Science* ran an Editorial Expression of Concern regarding the Report on 1 December 2016 (2). **The Review Board’s report, dated 21 April 2017, cited the following reasons for their recommendation: (i) lack of ethical approval for the experiments; (ii) absence of original data for the experiments reported in the paper; (iii) widespread lack of clarity concerning how the experiments were conducted.** Although the authors have told *Science* that they disagree with elements of the Board’s report, and although Uppsala University has not yet concluded its own investigation, the weight of evidence is that the paper should now be retracted. In light of the Board’s recommendation and a 28 April 2017 request from the authors to retract the paper, ***Science* is retracting the paper in full.**



SHARE LETTERS



0



1

Editorial expression of concern

Jeremy Berg

+ See all authors and affiliations

Science 09 Dec 2016;
Vol. 354, Issue 6317, pp. 1242
DOI: 10.1126/science.aah6990

Article Info & Metrics eLetters PDF

In the 3 June issue, *Science* published the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology” by Oona M. Lönnstedt and Peter Eklöv (1). **The authors have notified *Science* of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository.** *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

Reference

1. O. M. Lönnstedt, P. Eklöv, *Science* **352**, 1213 (2016). [SFX@UGent](#) [Abstract/FREE Full Text](#)



We need to make a choice of how we frame open science

Show me your data, now!

I don't trust you!

I'll find all your mistakes!

This will not end well!



Could I look at your data?

OK, this is pretty cool!

Look what I found in here!

Your data is so useful!

And let us not forget that your data will most likely live a much longer and more useful life than your publication!

open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

A nice guide to practical open science has been written, but it has been a few years ago.

Preprint

NOT PEER-REVIEWED

"PeerJ Preprints" is a venue for early communication or feedback before peer review. Data may be preliminary. [Learn more about preprints](#) or [browse peer-reviewed articles instead](#).

Do you speak open science? Resources and tips to learn the language

Science and Medical Education

Paola Masuzzo^{1,2}, Lennart Martens^{1,2}

January 3, 2017

Author and article information

Abstract

The internet era, large-scale computing and storage resources, mobile devices, social media, and their high uptake among different groups of people, have all deeply changed the way knowledge is created, communicated, and further deployed. These advances have enabled a radical transformation of the practice of science, which is now more open, more global and collaborative, and closer to society than ever. Open science has therefore become an increasingly important topic. Moreover, as open science is actively pursued by several high-profile funders and institutions, it has fast become a crucial matter to all researchers. However, because this widespread interest in open science has emerged relatively recently, its definition and implementation are constantly shifting and evolving, sometimes leaving researchers in doubt about how to adopt open science, and which are the best practices to follow.

Enter your institution

To find colleagues at PeerJ

Enter to search

Download preprint as...

Follow for updates

Tools & info

Citations in Google Scholar

Add feedback 5

Ask questions

Add links

Visitors 3,228 click for details

Views 4,456

Downloads 510

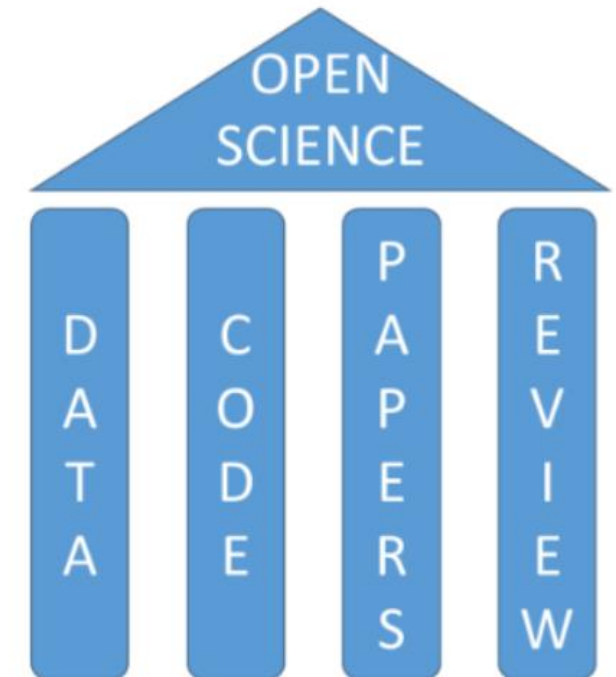
Outline

PeerJ Job Listings [beta]

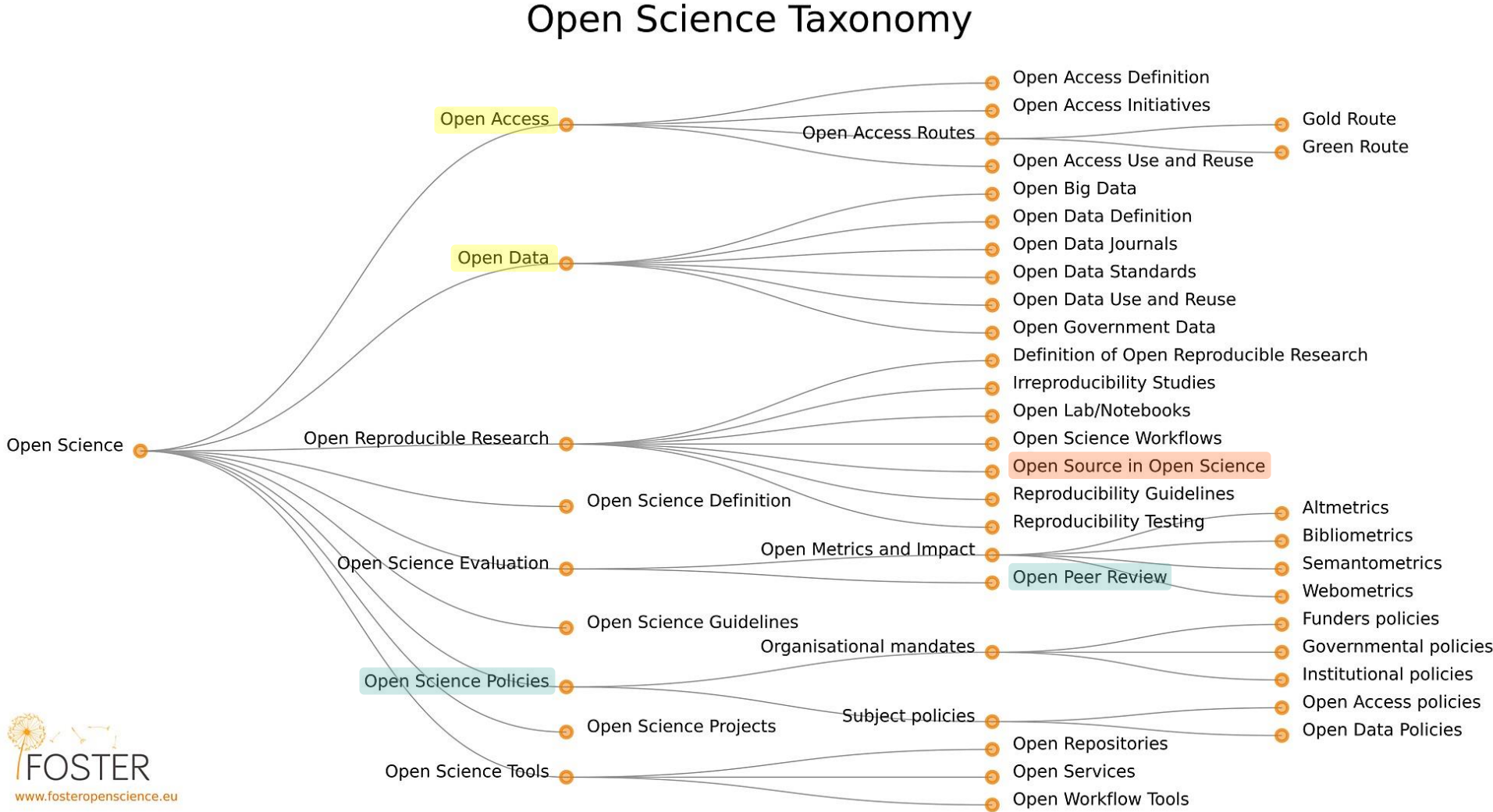
List & find academic jobs on PeerJ for free.

[Learn more >](#)

Get PeerJ content alerts



A somewhat different (more detailed) open science taxonomy has been collated by the FOSTER network



open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

In 2010, Elsevier's reported a 36% profit margin – higher than Apple, Google, or Amazon that year



The long read

Is the staggeringly profitable business of scientific publishing bad for science?

It is an industry like no other, with profit margins to rival Google - and it was created by one of Britain's most notorious tycoons: Robert Maxwell. By Stephen Buranyi

Three types of open access: gold, green, and delayed

Gold open access means you pay, as author, to have your paper open access. *Other papers in the same journal may not be, and so subscriptions remain necessary!* Journals love this as it means they get paid twice.

Green open access means that you deposit a preprint (without the 'added value' of peer review and typesetting) to a public repository such as arXiv, BioRxiv, MedRxiv, or ChemRxiv.

Delayed open access means that your paper only becomes open access after an embargo period (often a year)

The Directory of Open Access Journals helps you to find open access journals



SEARCH ▾ DOCUMENTATION ▾ ABOUT ▾

LOGIN →]

DIRECTORY OF OPEN ACCESS JOURNALS

Find open access journals & articles.

Journals Articles

In all fields ▾ [SEARCH](#)

80
LANGUAGES

135
COUNTRIES
REPRESENTED

13,635
JOURNALS
WITHOUT FEES

20,338
JOURNALS

9,792,647
ARTICLE RECORDS



Green open access as a means to fast and free open access



zenodo



figshare



bioRxiv
beta

THE PREPRINT SERVER FOR BIOLOGY



Innovations in peer review

JUNE 26, 2017 PROTOHEDGEHOG OPEN SCIENCE

Peer review is broken, but we hold the technological and social solutions to fix it. Right now.

We just submitted a monster paper on the history and present diversity of peer review practices to F1000 Research. It's [available in advance here](#), and soon will be open to public commenting from anyone as it undergoes formal peer review. We wrote it in a [similar manner](#) to another paper published on [Open Access](#) last year.

INSTITUTIONAL REPOSITORIES



CC BY-SA 4.0

Ghent University has built its own institutional repository



ACADEMIC BIBLIOGRAPHY

[Add publications and datasets](#) [Lists](#) [Sign in](#)

Search 200 years of publications by Ghent University researchers.

[Search publications and datasets](#)

[Advanced search](#)

Recently added research

New open access publications and datasets

Ghent University has a [policy](#) setting out its vision and guidelines for scholarly publishing.

- Blood-borne extracellular vesicles of bacteria and intestinal cells in patients with ps...
- Characterization of the genetic composition and establishment of a core collection for ...
- Decolonizing Rather than Decentering 'Europe'
- Data from: An Impact Assessment of Par-Baking and Storage on the Quality of Wheat, Whol...
- Auto-scaling dataset based on the gym-hpa framework
- jpedro1992/gym-hpa: v0.0.1-alpha

[View all open access publications and datasets.](#)



Photo by [Alessandro Bianchi](#) on Unsplash

Connected with ORCID

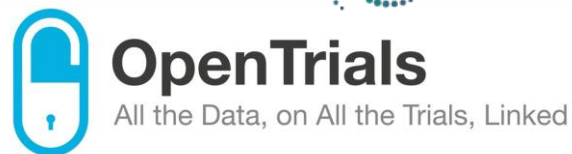
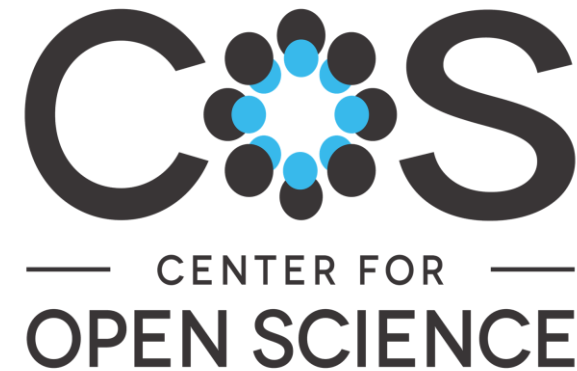
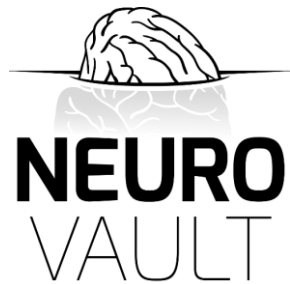
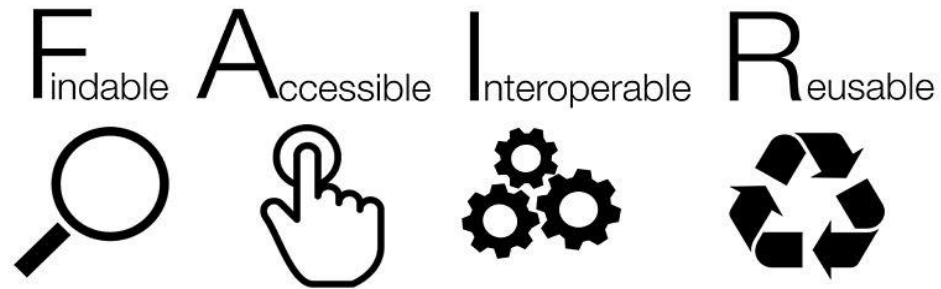
ORCID provides a persistent identifier that will distinguish you from other researchers throughout your scholarly career.

- Choosing one type of advocacy tactics over the other?
- The role of orphan crops in the transition to nutritional quality-oriented crop improve...
- Mucosa-associated lymphoid tissue lymphoma translocation protein 1 inhibition alleviate...

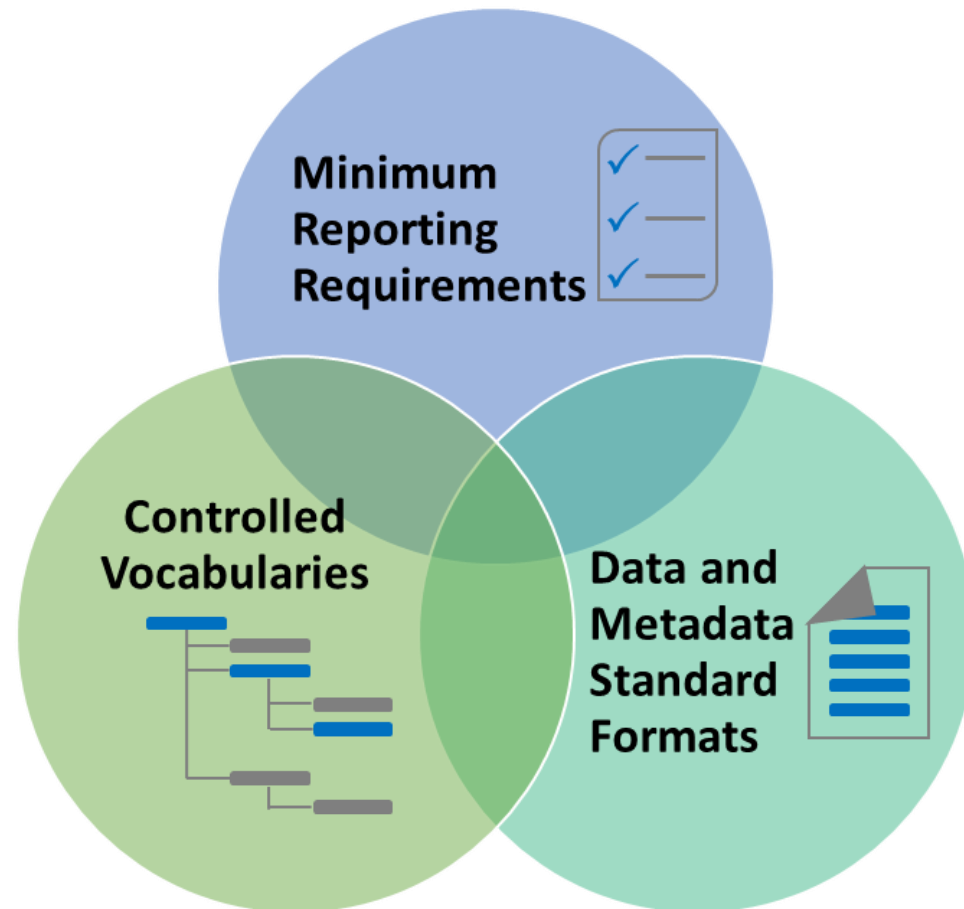
New PhD dissertations

- The parasite community of solitary bees and Anthropogenic filtering of the gut microbiome.
- 3D-explicit tree representation from terrestrial laser scanning to improve radiative t...
- Identification and characterization of protein interactions between the hepatitis E vir...
- Studying microglia and astrocytes in the liver-brain axis during chronic experimental h...
- Endectocide treatment of cattle as complementary tool for malaria control in Sub-Sahara...

Open data is widely organised around FAIR principles, and compliant (generic) data storage systems exist



Data sharing requires three building blocks: minimal requirements, CVs, and standard formats



Of course, a suitable data licence needs to be chosen as well, and here, Creative Commons licences are the most popular

Data without license may NOT be shared at all

Two Creative Commons licenses should be your top choices:

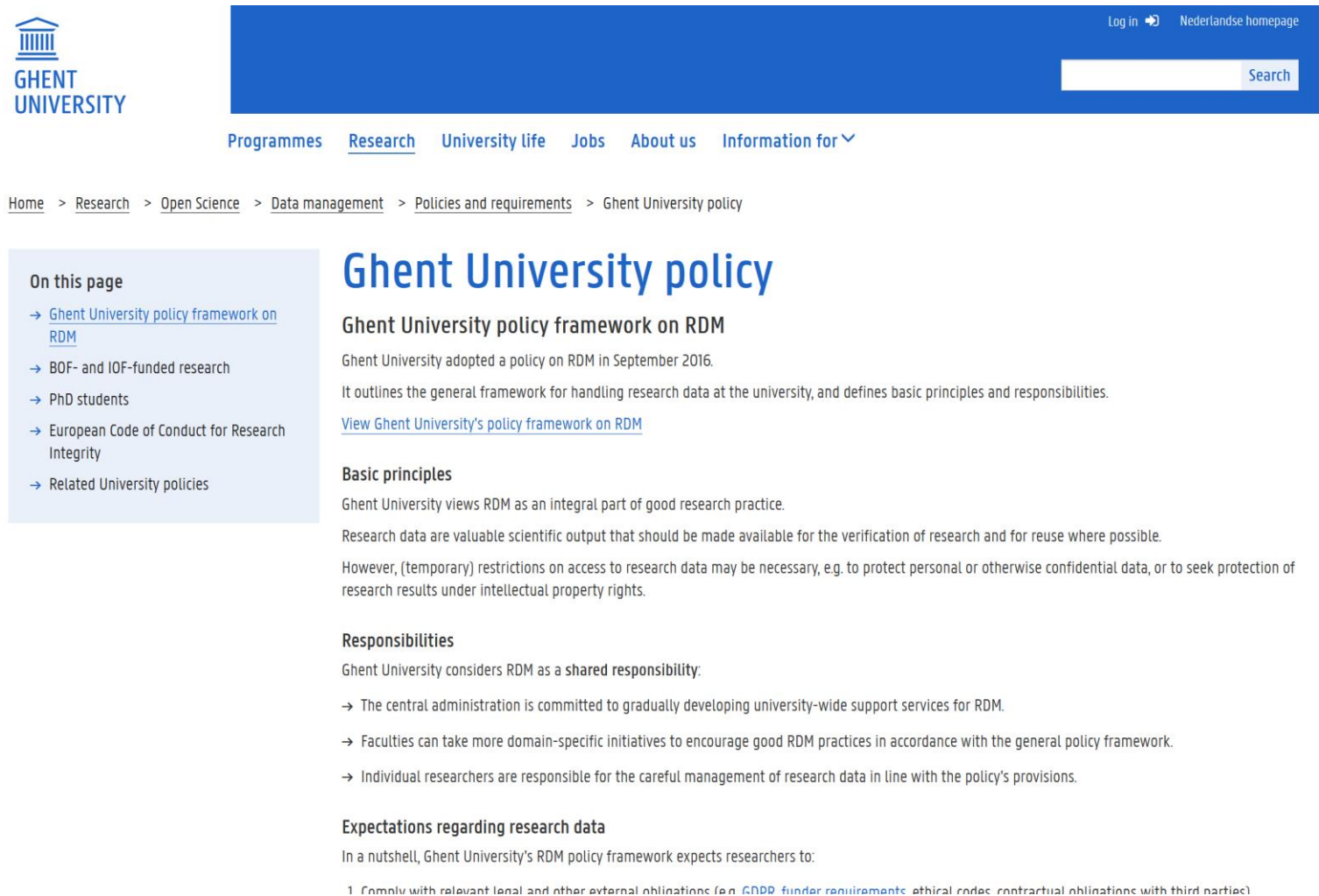
Attribution (CC BY), only mandates recognition of the author

Attribution-ShareAlike (CC BY-SA), as CC BY above, but all derived works need to be licensed CC-BY-SA as well (*infectious licence*)

Note that CC licenses are **not** meant for software code; see later for examples of suitable open source licenses!

Wikipedia and a lot of Flickr uses CC, amongst many others

Ghent University has defined a detailed policy framework on research data management (RDM)



The screenshot shows the Ghent University website. At the top left is the Ghent University logo. To the right is a blue navigation bar with a search box and links for "Log in" and "Nederlandse homepage". Below the navigation bar is a menu with "Programmes", "Research", "University life", "Jobs", "About us", and "Information for". A breadcrumb trail reads: "Home > Research > Open Science > Data management > Policies and requirements > Ghent University policy".

On this page

- [Ghent University policy framework on RDM](#)
- BOF- and IOF-funded research
- PhD students
- European Code of Conduct for Research Integrity
- Related University policies

Ghent University policy

Ghent University policy framework on RDM

Ghent University adopted a policy on RDM in September 2016. It outlines the general framework for handling research data at the university, and defines basic principles and responsibilities.

[View Ghent University's policy framework on RDM](#)

Basic principles

Ghent University views RDM as an integral part of good research practice. Research data are valuable scientific output that should be made available for the verification of research and for reuse where possible. However, (temporary) restrictions on access to research data may be necessary, e.g. to protect personal or otherwise confidential data, or to seek protection of research results under intellectual property rights.

Responsibilities

Ghent University considers RDM as a **shared responsibility**:

- The central administration is committed to gradually developing university-wide support services for RDM.
- Faculties can take more domain-specific initiatives to encourage good RDM practices in accordance with the general policy framework.
- Individual researchers are responsible for the careful management of research data in line with the policy's provisions.

Expectations regarding research data

In a nutshell, Ghent University's RDM policy framework expects researchers to:

- 1 Comply with relevant legal and other external obligations (e.g. [GDPR](#), [funder requirements](#), [ethical codes](#), [contractual obligations with third parties](#))

Data management plans are mandatory for all UGent funding as well as for most external funders

DMP ONLINE .BE My Dashboard Create plans Reference Help Lennart Martens

Ghent University (UGent - UZ Gent)

www.ugent.be Send mail to request DMP feedback (attach pdf!)

Ghent University RDM webpages rdm.support@ugent.be

My Dashboard

The table below lists all the plans associated with the **current user account**. This includes the plans you have created with this account and the plans that have been shared with you for this account.

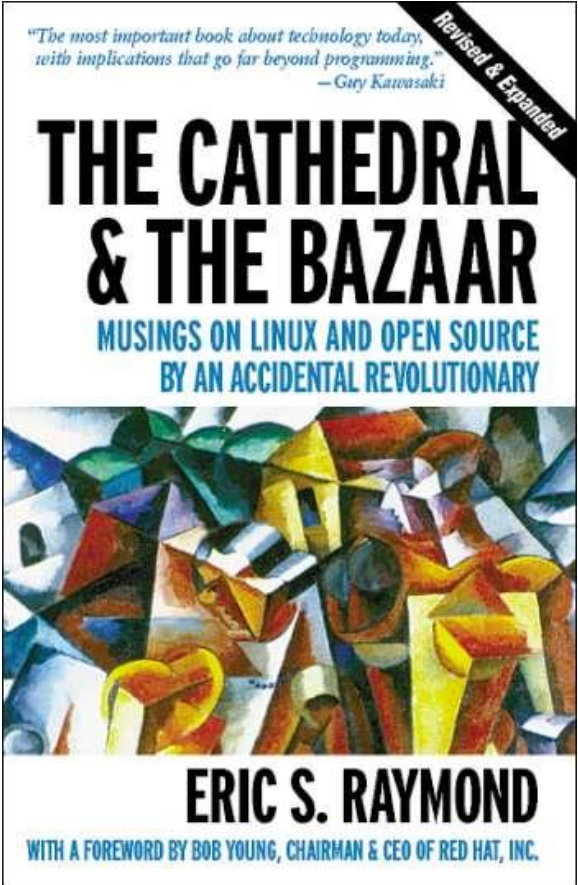
Project Title	Template	Edited	Role	Owner	Test	Visibility	Shared	
Immunopeptidomics-based Development of Next-Generation Bacte ...	Horizon Europe DMP +	09-01-2024	Co-owner	lyudmila.kovalchuke@ugent.be	<input type="checkbox"/>	Private	Yes	Actions
A computational pipeline for highly sensitive profiling of t ...	FWO DMP +	24-05-2022	Co-owner	tineclae.claeys@ugent.be	<input type="checkbox"/>	Private	Yes	Actions
A bioinformatics toolbox for forensic proteomics	Generic DMP +	26-04-2022	Co-owner	toon.callens@ugent.be	<input type="checkbox"/>	Private	Yes	Actions
Epitope GOA (BOF21/GOA/033)	Generic DMP +	29-06-2021	Owner	You	<input type="checkbox"/>	Private	Yes	Actions
ProteinContour	FWO DMP	14-06-2021	Co-owner	wim.vranken@vub.be	<input type="checkbox"/>	Private	Yes	Actions
MASS Spectrometry TRaining network for Protein Lipid adduct ...	Horizon 2020 DMP	27-05-2021	Owner	You	<input type="checkbox"/>	Private	No	Actions

Create plan

Ghent University (UGent - UZ Gent)'s Plans

The table below lists the plans that users at your organisation have created and shared within your organisation. This allows you to download a PDF and view their plans as samples or to discover new research data.

The open source paradigm is old and venerable, and certainly not only linked to science



There's some choice in open source or free software licenses, and making that choice is not entirely trivial

GNU GPL: copyleft, infective

GNU LGPL: copyleft, but linking is non-infective

Apache2: open source, permissive

MIT: open source, permissive

BSD: open source, permissive

Creative Commons: not meant for software!

Not all licenses can be altered afterwards, and most need explicit permission from all contributing authors!

<https://opensource.org/licenses/category>

https://en.wikipedia.org/wiki/Comparison_of_free_and_open-source_software_licenses

<http://choosealicense.com/licenses>

Open source code should be hosted on a third-party platform, like GitHub, BitBucket, or similar

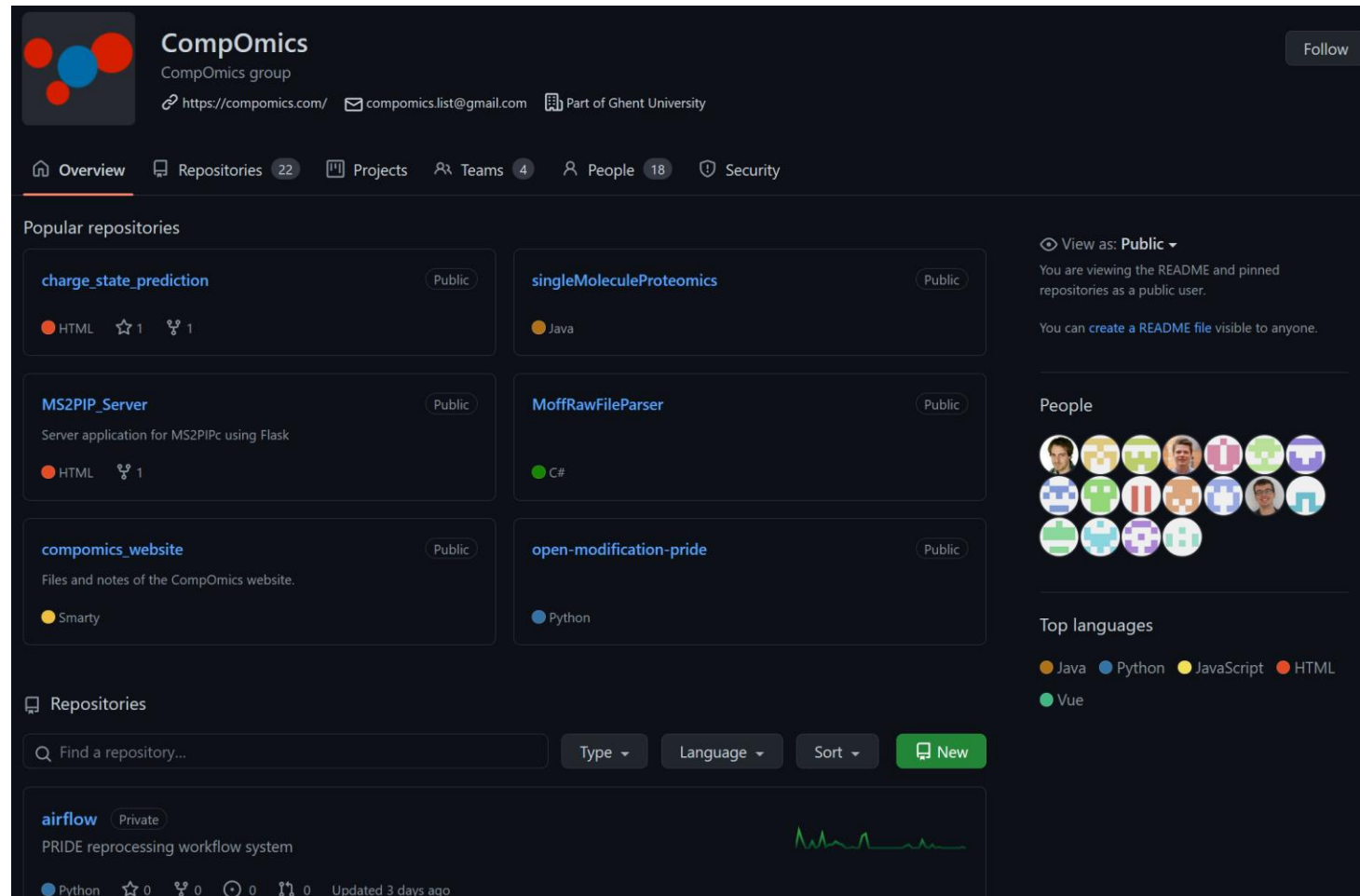
The screenshot shows the GitHub profile page for the 'compomics' organization. The profile name is 'Computational Omics and Systems Biology Group', with a bio stating they specialize in high-throughput Omics data management and analysis. The profile has 64 followers and lists contact information including a website, Twitter handle, and email. The 'Pinned' section displays six repositories: ThermoRawFileParser (C#, 150 stars), DeepLC (Python, 43 stars), ms2rescore (Python, 32 stars), compomics-utilities (Java, 26 stars), peptide-shaker (Java, 41 stars), and searchgui (Java, 35 stars). The 'Repositories' section shows a search bar and a list of repositories, including ms2rescore and DeepLC, with their respective languages, licenses, and update dates. The right sidebar contains options to view as public, a 'Discussions' section, a 'People' section with a grid of member avatars, and a 'Top languages' section showing Java, Python, HTML, Jupyter Notebook, and JavaScript.

<https://github.com/compomics>



CC BY-SA 4.0

Ghent University has its own local GitHub server, used during development or in case of potential IP



The screenshot shows the GitHub profile page for the 'CompOmics' group. The profile header includes the group name, a 'Follow' button, and contact information: 'https://compomics.com/', 'compomics.list@gmail.com', and 'Part of Ghent University'. The navigation bar shows 'Overview' (selected), 'Repositories' (22), 'Projects', 'Teams' (4), 'People' (18), and 'Security'. The main content area is divided into several sections:

- Popular repositories:** A grid of repository cards. Each card shows the repository name, a 'Public' label, a language icon, and star/fork counts. The visible repositories are:
 - charge_state_prediction** (HTML, 1 star, 1 fork)
 - MS2PIP_Server** (HTML, 1 fork)
 - compomics_website** (Smarty)
 - singleMoleculeProteomics** (Java)
 - MoffRawFileParser** (C#)
 - open-modification-pride** (Python)
- View as:** A dropdown menu set to 'Public'. Below it, text reads: 'You are viewing the README and pinned repositories as a public user. You can create a README file visible to anyone.'
- People:** A grid of profile pictures for group members.
- Top languages:** A horizontal bar chart showing the distribution of languages used in the repositories. The languages listed are Java, Python, JavaScript, HTML, and Vue.
- Repositories:** A section with a search bar 'Find a repository...', filters for 'Type', 'Language', and 'Sort', and a 'New' button. The first repository shown is **airflow** (Private), described as 'PRIDE reprocessing workflow system', with 0 stars, 0 forks, 0 clones, and 0 pulls, updated 3 days ago.

open science is here to stay

what does it all entail?

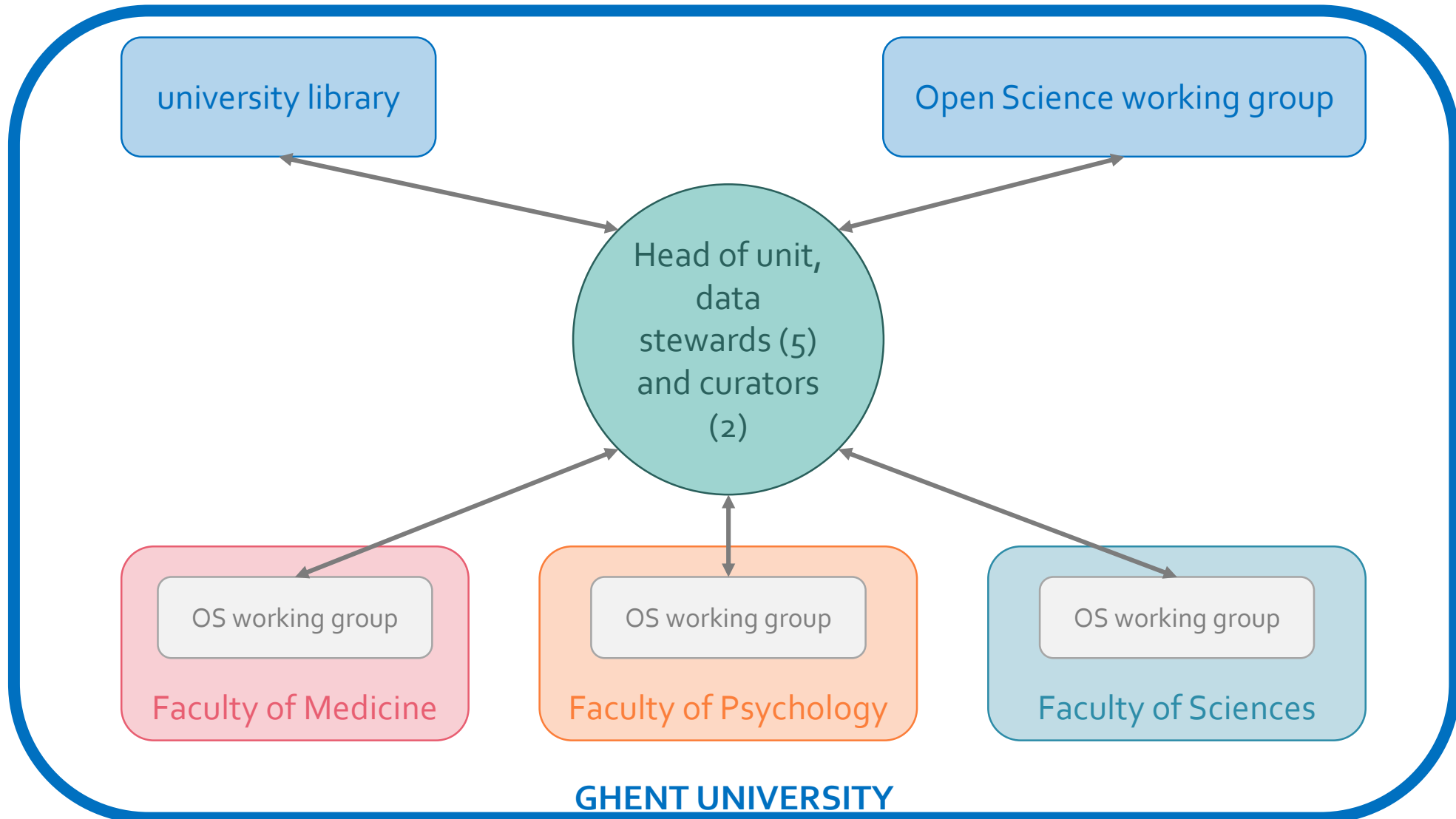
open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

Ghent University has working groups at different levels, with a central, connecting role reserved for data stewards



The roles of the data stewards is to broadly support open science practices, and help fulfill requirements

How can we help?

We offer a range of advisory, training, and curation services regarding (the management of) research data:

- [Data Management Plan](#) support (templates, guidance, feedback, examples)
- Online information and guidance
- [RDM helpdesk](#)
- Consultations with individual researchers & groups
- [Training, info sessions, & other events](#)
- User support for Library RDM tools (e.g. [DMPonline.be](#); [OSF for UGent](#))
- Curation of datasets and software registered in [Biblio](#)
- Curation of datasets and software submitted e.g. to the Ghent University [Zenodo Community](#), or on request

The six data stewards are each assigned to a topic cluster

Who are we?

As data stewards and data curators we are members of the **open science team** within [Ghent University Library](#) (Boekentoren), which is part of the University's Research Department.

We have experience with research and data in various disciplines, thus forming a team with complementary skills and areas of expertise.

Our group includes 1 coordinator, 5 data stewards, and 2 data curators:

Open science coordinator

→ [Myriam Mertens](#)

Data stewards

We act as a main RDM point of contact for each of the 5 faculty clusters.

We aim to be as visible and accessible as possible within the faculties, for example through activities such as consultations with researchers, training, and other events.

→ Cluster **Law, Arts & Humanities** (Faculties of Arts & Philosophy; Law & Criminology): [Thomas Van de Velde](#)

→ Cluster **Social and Behavioural Sciences** (Faculties of Economics & Business Administration; Psychology & Educational Sciences; Political & Social Sciences): [Ziad Choueiki](#)

→ Cluster **Natural Sciences** (Faculty of Sciences): [Paula Oset](#)

→ Cluster **(Bioscience) Engineering** (Faculties of Engineering & Architecture; Bioscience Engineering): [Stefanie De Bodt](#)

→ Cluster **Life Sciences & Medicine** (Faculties of Medicine & Health Sciences; Veterinary Medicinen; Pharmaceutical Sciences): [Laura Standaert](#)

Data curators

→ [Evelien Dhollander](#)

→ [Kevin Leonard](#)

open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

Focusdag



Woensdag 8 november 2017
9.30u - 17.00u
Paleis der Academiën, Brussel

lennart martens, Jonge Academie



Jonge Academie



DEPARTEMENT
ECONOMIE
WETENSCHAP &
INNOVATIE

Open Science is actually expanding the value of our research

Value is increasingly recognized to be found in (collections of) data from scientific research, and in shared code

Opportunities abound for those who wish to tap into the information made available through open science.

Many of the most valuable companies on the planet already know that large scale data analysis is very much worthwhile

We also need to introduce changes into how we train, and into how we evaluate

We have to provide people with the right skillset

- *Basic data management skills for everyone*
- *Basic data consumption skills for everyone*
- *Specialisation for Data Stewards*
- *Specialisation for Data Scientists*

We need to stimulate and reward original thinking on the (re-)use of open data, as it contains enormous potential!

In the Faculty of Medicine, we created adaptable educational guidance to bring open science skills into the curricula

		RDM-eindcompetenties	Volgorde van implementatie
1. Verantwoordelijke dataproductenten	1.1 Inbedding in doelstellingen voor verantwoord en correct onderzoek	1.1.1 Kennis van regels en normen inzake RDM 1.1.2 Inzicht in verantwoordelijkheid onderzoeker voor RDM 1.1.3 Kennis van RDM oplossingen 1.1.4 Toepassen van RDM oplossingen	
	1.2 Uitbreiding van technische kennis/vaardigheden	1.2.1 Kennis van beveiliging en encryptie 1.2.2 Toepassen van beveiliging en encryptie 1.2.3 Kennis van opslag/archiveringsstrategieën	
2. Gesofistikeerde dataconsumenten	2.1 Inbedding in doelstellingen voor (statistische) data analyse	2.1.1 Kennis van publieke databronnen 2.1.2 Inzicht in nut en beperkingen publieke data 2.1.3 Toepassen: het opvragen van publieke data 2.1.4 Kennis van grootschalige analyse van publieke data	
	2.2 Uitbreiding van technische kennis/vaardigheden	2.2.1 Kennis van formaten voor publieke data 2.2.2 Toepassen automatische dataverwerking 2.2.3 Toepassen van grootschalige (her)analyse van publieke data	

basic level, advanced level, specialisation level

We created detailed learning goals, matched to the final competencies in the previous table (i)

1. DATA & WET

1.1 Ethiek en confidentialiteit		
Leerinhoud	Specificiteit	Eindcomp.
Algemene Verordening Gegevensbescherming (AVG) of <i>General Data Protection Regulation (GDPR)</i>		
– basisprincipes	Generiek	1.1.1
– <i>GDPR Code of Conduct</i> UGent / UZ Gent	Generiek/FGE	1.1.1
– <i>GDPR</i> versus <i>RDM</i>	Generiek	1.1.1
Privacy en persoonsgegevens		
– Classificatie		
– confidentiële vs niet-confidentiële data	Generiek	1.1.1
– anonieme data vs persoonsgegevens	Generiek	1.1.1
– sensitieve vs niet-sensitieve persoonsgegevens	FGE	1.1.1
– geanonimiseerde vs gepseudonimiseerde persoonsgegevens	FGE	1.1.1
– <i>Informed Consent</i> (Wat?, Waarom?)	FGE	1.1.1
– <i>Small Cell Risk Analysis (SRCA)</i>	FGE	1.1.1
<i>Good Clinical Practice (GCP)</i> (Wat?, Waarom?)	FGE	1.1.1

basic level
advanced level
specialisation level

2. SITUERING RDM

2.1 Belang RDM		
Leerinhoud	Specificiteit	Eindcomp.
Definitie (Wat is RDM?)	Generiek	1.1.2
<i>FAIR Guiding Principles - Findable, Accessible, Interoperable and Reusable / Open Science (Open Research Data)</i>	Generiek	1.1.2
Voordelen/uitdagingen (Waarom RDM?)	Generiek	1.1.2

2.2 Research lifecycle		
Leerinhoud	Specificiteit	Eindcomp.
Definitie verschillende fasen levenscyclus onderzoeksdata	Generiek	1.1.2

We created detailed learning goals, matched to the final competencies in the previous table (ii)

2.3 RDM & wetenschappelijke integriteit		
Leerinhoud	Specificiteit	Eindcomp.
Definitie (Wat wordt er verstaan onder wetenschappelijke integriteit?, schendingen)	Generiek	1.1.2
RDM binnen Europese gedragscode voor wetenschappelijke integriteit (<i>EU Code of Conduct</i>)	Generiek	1.1.2

3. PLANFASE

3.1 Datamanagementplan		
Leerinhoud	Specificiteit	Eindcomp.
Definitie DMP	Generiek	1.1.1 1.1.3
Belang/Voordelen/Beperkingen	Generiek	1.1.2
DMPonline.be tool	Generiek	1.1.3 1.1.4

4. ONDERZOEKSFASE

4.1 Data verzamelen		
Leerinhoud	Specificiteit	Eindcomp.
Datatypes	Generiek	1.1.3
Bestandsformaten		
– soort: open formaat vs. gesloten, ...	Generiek	1.1.3
– voorkeursformaten (<i>Recommended Standard Data Formats</i>)	Opleiding	1.1.3
Datavolumes en schaalbaarheid	Generiek	1.1.3
<i>Data Capture Methods</i>	Generiek	1.1.3
– typevoorbeeld REDCap	FGE	1.1.3 1.1.4
Datakwaliteit	Generiek	1.1.3 1.1.4
Dataverzameling in samenwerkingsverband (<i>Collaborative Data Collection</i>)	Generiek	1.1.3

4.2 Data documenteren		
Leerinhoud	Specificiteit	Eindcomp.
Belang	Generiek	1.1.2
Niveau: projectniveau vs dataniveau	Generiek	1.1.3 1.1.4
Metadata	Generiek	1.1.3 1.1.4
Consistente data: <i>Codebook / Data Dictionary</i>	Generiek	1.1.3 1.1.4
Activiteitenverslag: <i>Audit Trail</i>	Generiek	1.1.3 1.1.4
Standaarden		
– <i>Minimal Reporting Standards</i>	Opleiding	1.1.3
– Relevante ontologieën (bvb. biomedische)	Opleiding	1.1.3
– <i>Standard Operating Procedures (SOPs)</i>	Generiek	1.1.3 1.1.4

4.3 Data organiseren		
Leerinhoud	Specificiteit	Eindcomp.
Betekenisvolle bestandsnamen (<i>File Naming Conventions</i>)	Generiek	1.1.3 1.1.4
Gestructureerde en consequente folderstructuur	Generiek	1.1.3 1.1.4

basic level
advanced level
specialisation level

We created detailed learning goals, matched to the final competencies in the previous table (iii)

5. GEBRUIKERSFASE

5.1 Data zoeken/opvragen		
Leerinhoud	Specificiteit	Eindcomp.
Publieke databronnen in de levenswetenschappen <ul style="list-style-type: none"> – publiek beschikbare statistische info (vb. Statbel) – typevoorbeelden: eHealth, healthdata.be, NIH – biomedische databronnen – <i>World Health Organization (WHO)</i> platform (incl. clinicaltrials.gov, EU Clinical Trials Register, International Clinical Trials Registry Platform) 	Generiek FGE FGE FGE	2.1.1 2.1.1 2.1.1 2.1.1
Publiek beschikbare data <ul style="list-style-type: none"> – belang citeren (vindbaarheid, geven van credit (impact, Altmetrics)) – nut en beperkingen <ul style="list-style-type: none"> – heterogeniteit van publiek beschikbare data – kwaliteit van publiek beschikbare data – volledigheid van publiek beschikbare data – GDPR, nl. persoonsgegevens als belangrijke limitatie – opvragen d.m.v. databronnen (vb. PubMed, NCBI) – (meta)verwerking 	Generiek Generiek Generiek Generiek FGE Generiek Opleiding	1.1.2 2.1.2 2.1.2 2.1.2 2.1.2 2.1.3 2.1.3

basic level
advanced level
specialisation level

5.2 Data (her)gebruiken		
Leerinhoud	Specificiteit	Eindcomp.
Belang (heranalyse, serendipiteit, individueel hergebruik)	Generiek	1.1.2
Methoden voor grootschalige, heterogene (her)analyse van publieke data <ul style="list-style-type: none"> – omgaan met grote hoeveelheid data – omgaan met heterogene data – omgaan met onvolledige data – kwaliteit van data bepalen 	Generiek / Opleiding Generiek / Opleiding Generiek / Opleiding Generiek / Opleiding	2.1.4 2.2.3 2.1.4 2.2.3 2.1.4 2.2.3 2.1.4 2.2.3
Bestandsformaten voor de meest courante publieke databronnen	Opleiding	2.2.1
Geautomatiseerde dataverwerking in populaire omgevingen of programmeertalen <ul style="list-style-type: none"> – gebruik maken van bestandsformaten (vb. csv, json, xml) – van automatisatie in bestaande omgeving naar eigen scripts in Python of R – notie van machine leren (ML) en artificiële intelligentie (AI) (vb. classificatie van radiologische beeldvorming d.m.v. AI) 	Generiek Opleiding Opleiding	2.2.2 2.2.2 2.2.2

An example implementation from the Biomedical Sciences was also provided, here for the Bachelor

		Bachelor BMW									
		1 ^e	2 ^e				3 ^e				
		Biom. informatie en informatieverwerking	Inleiding tot de biostatistiek	Chemische en biomedische analyse	Biomedische basistechnieken	Literatuur review biom. onderzoek I	Humane moleculaire genetica	Bio-informatica	Toegepaste biomedische praktijk	Literatuur review biom. onderzoek II	Epidemiologie
1. Verantwoordelijke dataproductenten	1.1.1 Kennis van regels en normen inzake RDM										
	1.1.2 Inzicht in verantwoordelijkheid onderzoeker voor RDM										
	1.1.3 Kennis van RDM oplossingen										
	1.1.4 Toepassen van RDM oplossingen										
2. Gesofistikeerde dataconsumenten	2.1.1 Kennis van publieke databronnen										
	2.1.2 Inzicht in nut en beperkingen publieke data										
	2.1.3 Toepassen: het opvragen van publieke data										
	2.1.4 Kennis van grootschalige analyse van publieke data										
	2.2.1 Kennis van formaten voor publieke data										
	2.2.2 Toepassen automatische dataverwerking										
	2.2.3 Toepassen van grootschalige (her)analyse van publieke data										

An example implementation from the Biomedical Sciences was also provided, and here for the Master

		Master BMW												
		1 ^e								2 ^e				
		Goede laboratoriumpraktijk (GLP)	Bio-ethiek in experimentele geneeskunde	Geavanceerde bio-informatica	Good clinical practice (GCP) - Klinische studies	Majorstage	Onderzoeksstage	Massaal parallel sequencing data-analyse	Advanced Imaging and Image Analysis	Machine Leren methoden voor biom. gegevens	Klinische aspecten van kanker	Systeembioïogie	Genetische diagnostiek	Masterproef
1. Verantwoordelijke dataproducenten	1.1.1 Kennis van regels en normen inzake RDM													
	1.1.2 Inzicht in verantwoordelijkheid onderzoeker voor RDM													
	1.1.3 Kennis van RDM oplossingen													
	1.1.4 Toepassen van RDM oplossingen													
2. Gesofistikeerde dataconsumenten	2.1.1 Kennis van publieke databronnen													
	2.1.2 Inzicht in nut en beperkingen publieke data													
	2.1.3 Toepassen opvragen publieke data													
	2.1.4 Kennis van grootschalige analyse van publieke data													
	2.2.1 Kennis van formaten voor publieke data													
	2.2.2 Toepassen automatische dataverwerking													
	2.2.3 Toepassen van grootschalige (her)analyse van publieke data													

open science is here to stay

what does it all entail?

open access, open data, open code

the administration of open science

the impact of open science on education

a summary in three parts

Open science brings many good things, and we should be ready for these

Open science makes the work accessible to anyone

Open science allows people to build much more efficiently on previous work

Open science helps maximize the usefulness of each individual research effort

Data tend to have a (much!) longer shelf life than our (limited) interpretations

Open science fosters creativity, and stimulates revolutionary research

Summary – policy aspects

Focus on promise

- Share successful ideas/approaches/grant proposals

- Organise communities for exchange

- Inspire uptake

Positive reinforcement

- Highlight and reward open science success stories

- Adapt CVs to value meritorious open science effort

- Financially support firmly open science-related efforts (creation and maintenance of data, code, tools – can be hardware, software, skilled support staff)

Passive enforcement

- Count only open artefacts in bean-counting exercises

- Force non-paper contributions onto CVs

- Introduce valuation of actual re-use/impact metrics

Summary – implementation steps

Facilitate adherence to requirements

- Solid support for administrative tasks (DMPs)

- Support for open science practices (institutional, or third-party repositories)

- Support for open science-related text in proposals

Invest in education

- Train responsible data producers

- Train sophisticated data consumers

- Train future large-scale data scientists across domains!

Connect to industry

- Stay in touch with where they are moving

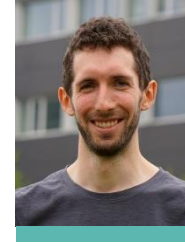
- What do they need in terms of:

 - Skilled staff

 - Resources and tools

 - Networking





Comp
omics



www.compomics.com
compomics.github.io