

Breaking down barriers in the scientific use of EO data



ifgi

Institute for Geoinformatics
University of Münster

Edzer Pebesma

EODC workshop, TU Vienna, May 30-31, 2016



Open Innovation Open Science Open to the World

– a vision for Europe





EUROPEAN COMMISSION

DIRECTORATES-GENERAL FOR RESEARCH AND INNOVATION (RTD) AND
COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY (CONNECT)

BACKGROUND DOCUMENT

PUBLIC CONSULTATION

‘SCIENCE 2.0’: SCIENCE IN TRANSITION

0. OVERVIEW OF THE CONSULTATION PROCESS

This background paper to the online public consultation gives a short description of the term 'Science 2.0' as used in this consultation. The goal of the consultation is to better understand the full potential of 'Science 2.0' as well as the desirability of any possible policy action.

European Commission - Press release

European Cloud Initiative to give Europe a global lead in the data-driven economy

Brussels, 19 April 2016

The Commission today presented its blueprint for cloud-based services and world-class data infrastructure to ensure science, business and public services reap benefits of big data revolution.

Europe is the largest producer of scientific data in the world, but insufficient and fragmented infrastructure means this 'big data' is not being exploited to its full potential. By bolstering and interconnecting existing research infrastructure, the Commission plans to create a new **European Open Science Cloud** that will offer Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to store, share and re-use their data across disciplines and borders. This will be underpinned by the **European Data Infrastructure**, deploying the high-bandwidth networks, large scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud. This world-class infrastructure will ensure Europe participates in the global race for high performance computing in line with its economic and knowledge potential.

Focusing initially on the scientific community - in Europe and among its global partners -, the user base will over time be enlarged to the public sector and to industry. This initiative is part of a package of measures to strengthen Europe's position in data-driven innovation, to improve competitiveness and cohesion and to help create a [Digital Single Market](#) in Europe ([press release](#)).

Carlos **Moedas**, Commissioner for Research, Science and Innovation, said: *"Our goal is to create a European Open Science Cloud to make science more efficient and productive and let millions of researchers share and analyse research data in a trusted environment across technologies, disciplines and borders. We listened to the scientific community's plea for an infrastructure for Open Science and with this comprehensive plan we can get down to work. The benefits of open data for Europe's science, economy and society will be enormous."*

Günther H. **Oettinger**, Commissioner for the Digital Economy and Society, said: *"The European Cloud Initiative will unlock the value of big data by providing world-class supercomputing capability, high-speed connectivity and leading-edge data and software services for science, industry and the public sector. With this initiative, our ambition is to be in the global top-three in high performance computing by 2020. We will also be looking into the potential of quantum technologies which hold the promise to solve computational problems beyond current supercomputers."*

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016



PDF



Rights & Permissions

Is there a reproducibility crisis in science?



00:00

02:03



More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature's* survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

Crisis talks



1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.



Like

Recent

Read

Commented

1. **US advised to stick with troubled fusion reactor ITER**
Nature | 26 May 2016
2. **Two-hundred-terabyte maths proof is largest ever**
Nature | 26 May 2016
3. **Sights and sounds of seabed cities**
Nature | 26 May 2016

Newsletter

The best science news from *Nature* and beyond, direct to your inbox every day.

Email*

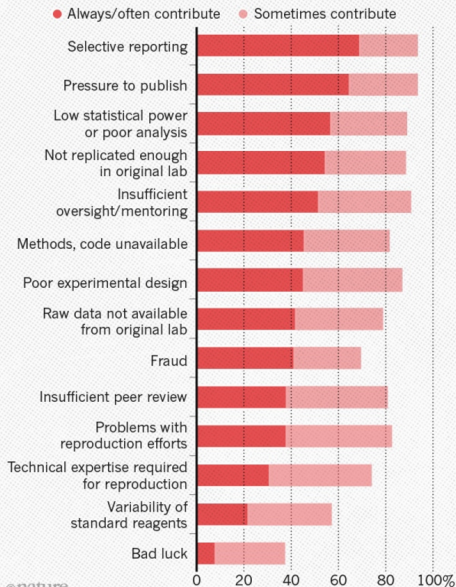
Submit

IS THERE A REPRODUCIBILITY CRISIS?



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

Can we do that by opening the EODC?

What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

Can we do that by opening the EODC?

What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

Can we do that by opening the EODC?

What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...,
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

Can we do that by opening the EODC?

What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...,
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

Can we do that by opening the EODC?

What is “Open Science”?

- ▶ ... and excuse for the EC to spend tax money to push innovation and the economy
- ▶ an ethics/morality thing pedantic scientists like to brag about
- ▶ ... but then what is “*Closed Science*”?
- ▶ shouldn't *Open Science* be the tautology...,
- ▶ ... and *Closed Research* be left to the industry?

How can we open Earth Observation Research?

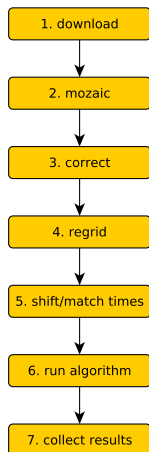
Can we do that by opening the EODC?

Questions:

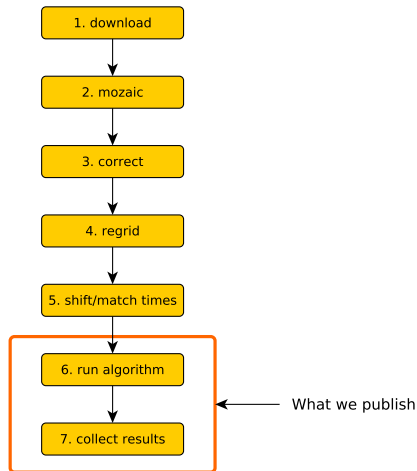
EO Workflow (WF) = [from scene download to final figure in paper]

- ▶ who has executed an EO WF that another research group had shared?
- ▶ who has shared their EO WF with (an)other research group(s)?
- ▶ who has published a complete EO WF?

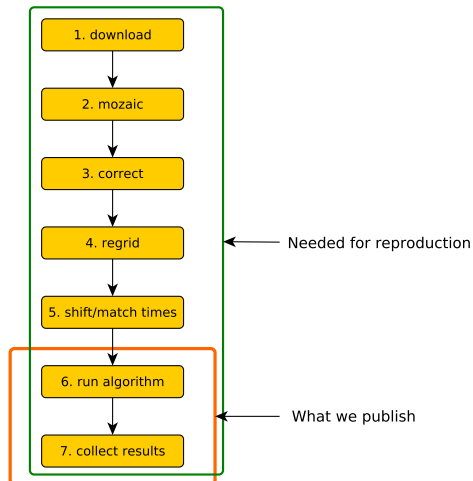
Current Earth Observation Research:



Current Earth Observation Research:



Current Earth Observation Research:



Science domains that do share workflows

- ▶ statistics (R),
- ▶ bioinformatics (R, taverna, kepler)
- ▶ astrophysics
- ▶ high energy physics (WLCG: shared usage of 170 data centers)
- ▶ EO: where agencies convert raw signal into level 0/1/2/3

Science domains that do share workflows

- ▶ statistics (R),
- ▶ bioinformatics (R, taverna, kepler)
- ▶ astrophysics
- ▶ high energy physics (WLCG: shared usage of 170 data centers)
- ▶ EO: where agencies convert raw signal into level 0/1/2/3

Opportunities for Open EODC's - 1/2

If Earth observation researchers would also share steps 1-5, this would

- ▶ make EO research open (reproducible)
- ▶ clear up processing steps that are now obscure,
- ▶ improve their understanding, speed up their improvement,
- ▶ speed up innovation,
- ▶ increase re-use, and hence citations and recognition

Opportunities for Open EODC's - 2/2

- ▶ provide the EO data archived read-only, in a simple, standardized way:
 - ▶ make the download step obsolete
 - ▶ stay close to how they're offered by the agencies
- ▶ provide flexible software setups:
 - ▶ based on docker containers,
 - ▶ provide docker images with common open source software/toolboxes
 - ▶ teach how to use it
- ▶ don't constrain access to the infrastructure
- ▶ provide transparent pricing, friendly for PhD students
- ▶ free for those who open scripts, prior to running them

Opportunities for Open EODC's - 2/2

- ▶ provide the EO data archived read-only, in a simple, standardized way:
 - ▶ make the download step obsolete
 - ▶ stay close to how they're offered by the agencies
- ▶ provide flexible software setups:
 - ▶ based on docker containers,
 - ▶ provide docker images with common open source software/toolboxes
 - ▶ teach how to use it
- ▶ don't constrain access to the infrastructure
- ▶ provide transparent pricing, friendly for PhD students
- ▶ free for those who open scripts, prior to running them

Opportunities for Open EODC's - 2/2

- ▶ provide the EO data archived read-only, in a simple, standardized way:
 - ▶ make the download step obsolete
 - ▶ stay close to how they're offered by the agencies
- ▶ provide flexible software setups:
 - ▶ based on docker containers,
 - ▶ provide docker images with common open source software/toolboxes
 - ▶ teach how to use it
- ▶ don't constrain access to the infrastructure
- ▶ provide transparent pricing, friendly for PhD students
- ▶ free for those who open scripts, prior to running them

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Requirements to Open EODC's

- ▶ set them up transparently, open source all config scripts
- ▶ *keep simple things simple*
- ▶ allow to replicate them easily, and flexibly:
 - ▶ e.g. for other, or smaller data archives
 - ▶ allow them to run on small computers too
- ▶ be transparent in prioritizing large jobs; provide SLA's
- ▶ think about decentralized governance, build a community
- ▶ think evolution, build bridges, compete fair (study the R community)

After that:

- ▶ document and share experiences, publish workflows, write books, teach
- ▶ change publication ethics and demand reproducibility

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the "exclusive for partners" game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the "exclusive for partners" game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the "exclusive for partners" game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the "exclusive for partners" game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the "exclusive for partners" game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the “exclusive for partners” game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the “exclusive for partners” game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the “exclusive for partners” game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow

Can an Open EODC break down barriers in Earth Observation Research?

YES! But how?

- ▶ reserve 25% of computation capacity for free use, provided workflows are shared
- ▶ open up your doors, and tell everyone to do the same
- ▶ don't play the “exclusive for partners” game
- ▶ forget fear, drop inertia
- ▶ build trust in the larger EO community
- ▶ show the way, and everyone will follow