

Water science Open Science must be

Supported by FNR ATTRACT
(A16/SR/11254288 and A18/BM/12341006)

Celebrating the launch of Nature
Water - Part 3: Focus on social
sciences and open science
2 Feb. 2023

Emma L. Schymanski¹, Stanislaus J. Schymanski^{2,3}

¹Luxembourg Centre for Systems Biomedicine (LCSB), University of
Luxembourg, Belvaux, Luxembourg,

²Environmental Research and Innovation (ERIN), Luxembourg Institute of Science and Technology (LIST),

³Swiss Data Science Center (SDSC)



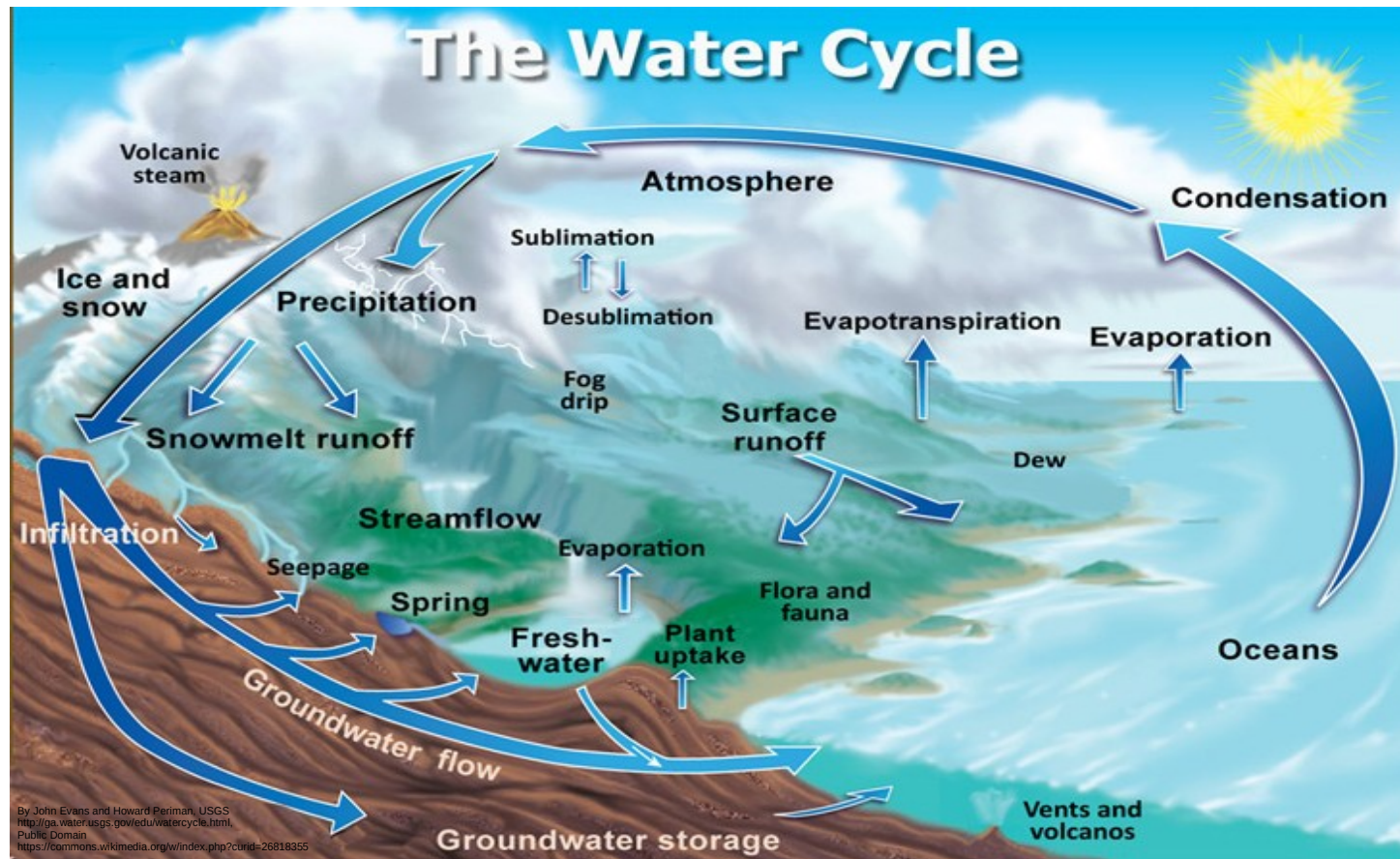
WATER CROSSES BORDERS, LINKS COMMUNITIES & GENERATIONS

Why OS?

What is OS?

How to do OS?

Conclusions



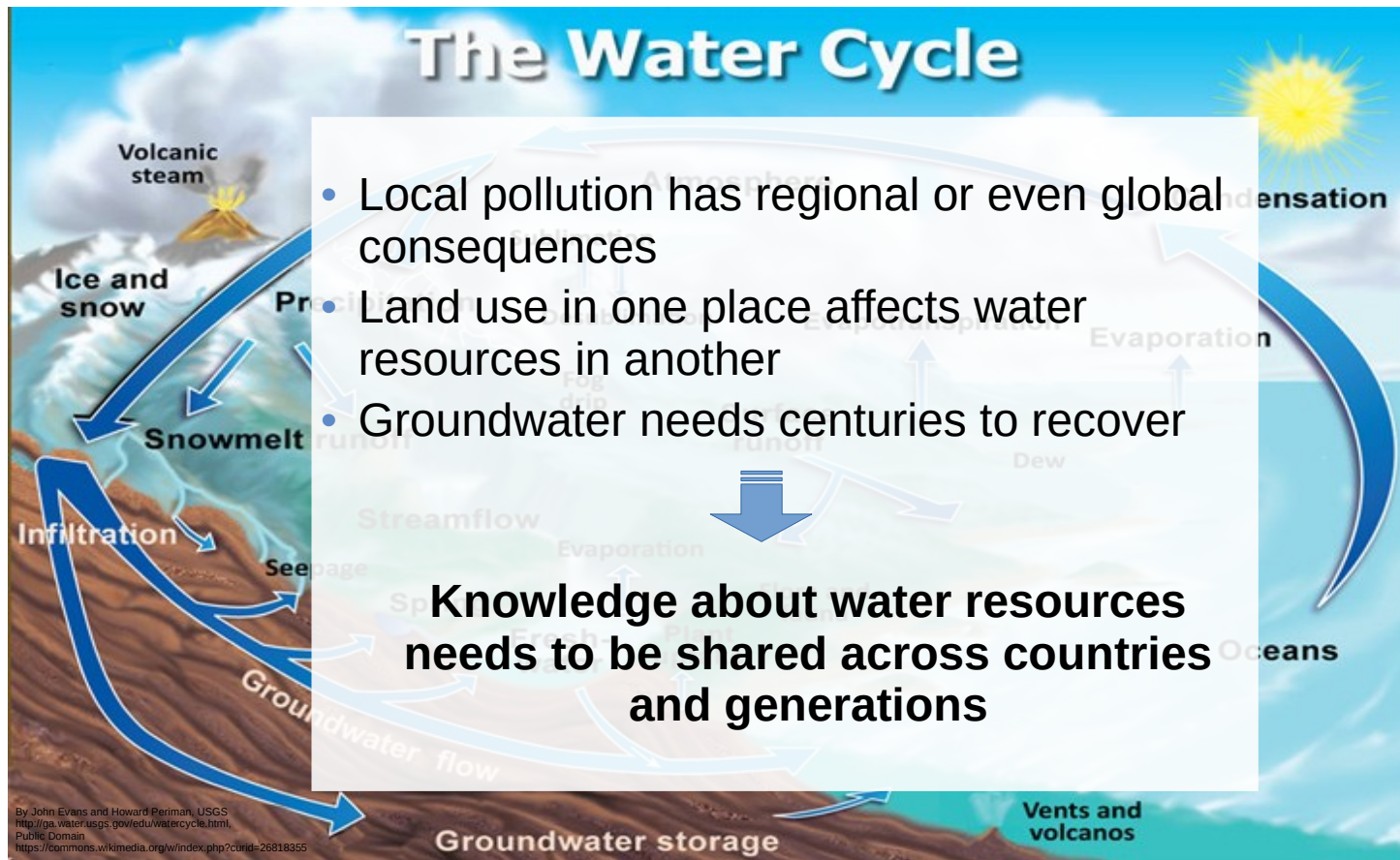
WATER CROSSES BORDERS, LINKS COMMUNITIES & GENERATIONS

Why OS?

What is OS?

How to do OS?

Conclusions



EXAMPLE: FISH KILL IN ODER RIVER (GERMANY)

Due to spill in Poland?

- High salinity
- Herbicides
- >1200 known chemicals analysed

Why OS?

What is OS?

How to do OS?

Conclusions



<https://www.tagesschau.de/ausland/europa/oder-fischsterben-polen-algen-103.html>

BROAD COLLABORATION NEEDED

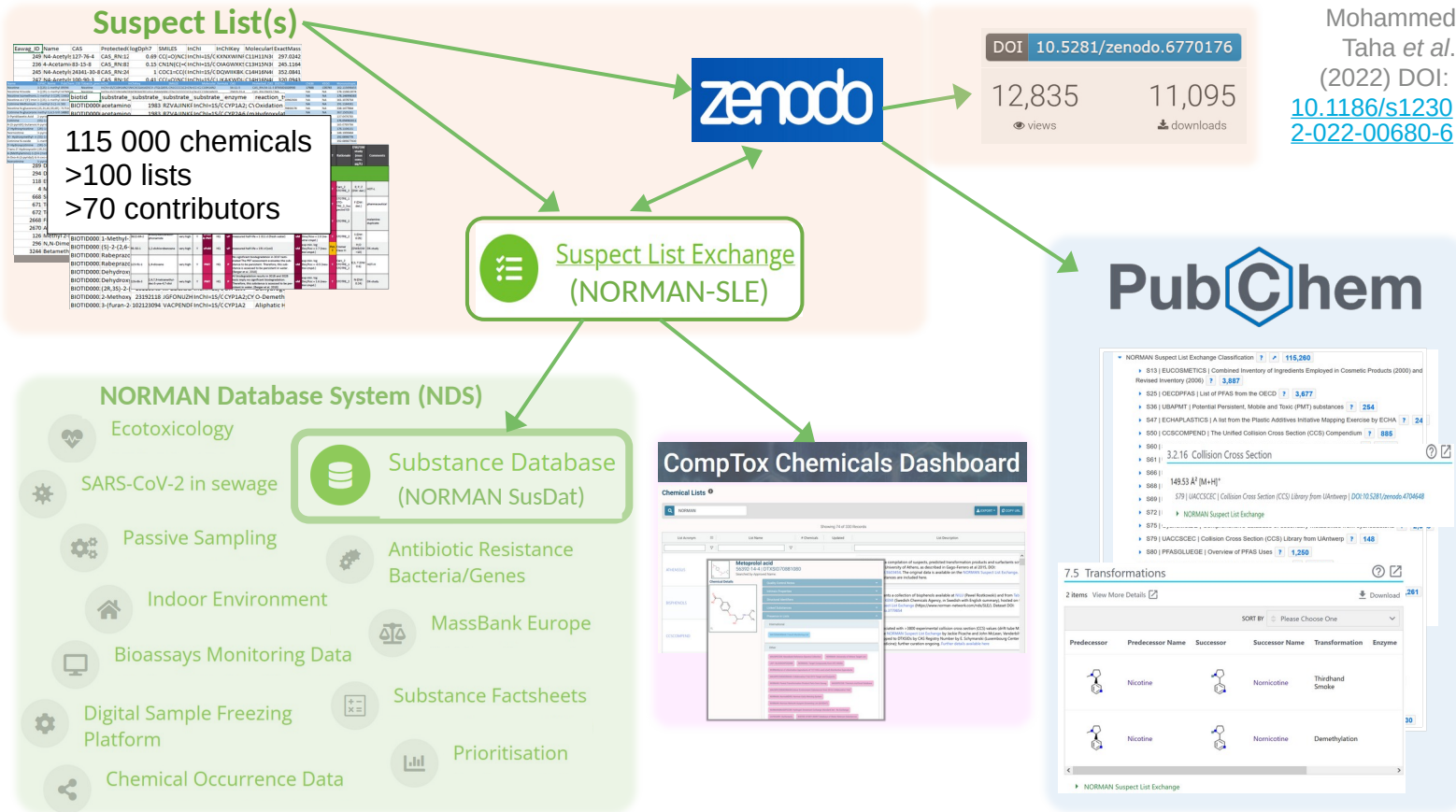
Mohammed
Taha *et al.*
(2022) DOI:
[10.1186/s12301-022-00680-6](https://doi.org/10.1186/s12301-022-00680-6)

Why OS?

What is OS?

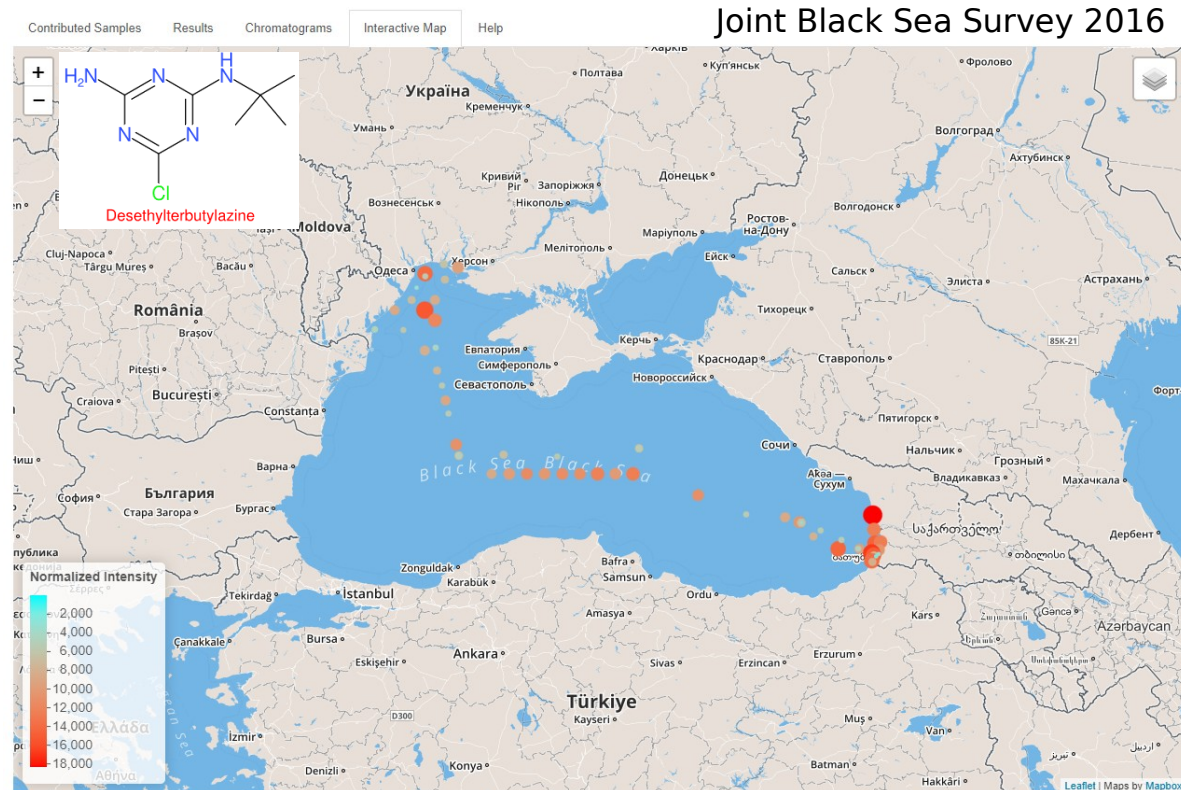
How to do OS?

Conclusions



NORMAN DIGITAL SAMPLE FREEZING PLATFORM

Why OS?
What is OS?
How to do OS?
Conclusions



Find out
where the
suspect is
hiding ...

Image provided by Nikiforos Alygizakis.
Alygizakis *et al.*, 2019, TrAC,
DOI: [10.1016/j.trac.2019.04.008](https://doi.org/10.1016/j.trac.2019.04.008)

Why OS?

What is OS?

How to do OS?

Conclusions



REPORT

A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon

ZHENYU TIAN , HAOQI ZHAO , KATHERINE T. PETER , MELISSA GONZALEZ , JILL WETZEL , CHRISTOPHER WU , XIMIN HU , JASMINE PRAT ,
EMMA MUDROCK, [...], AND EDWARD P. KOLODZIEJ [+17 authors](#) [Authors Info & Affiliations](#)

SCIENCE • 3 Dec 2020 • Vol 371, Issue 6525 • pp. 185-189 • DOI: 10.1126/science.abd6951

PubChem N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine

7.1 Transformations

1 item View More Details

Download

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	6PPD		6PPD-quinone	Ozone	

NORMAN Suspect List Exchange



WATER QUANTITY: DATA, MODELS AND PREDICTION

Forecasting of floods and water shortage relies on multidisciplinary **data**:

- climate
- topography
- soil & geo(morph)ology
- vegetation
- land use
- water abstraction...

... and reliable simulation **models** !

Why OS?

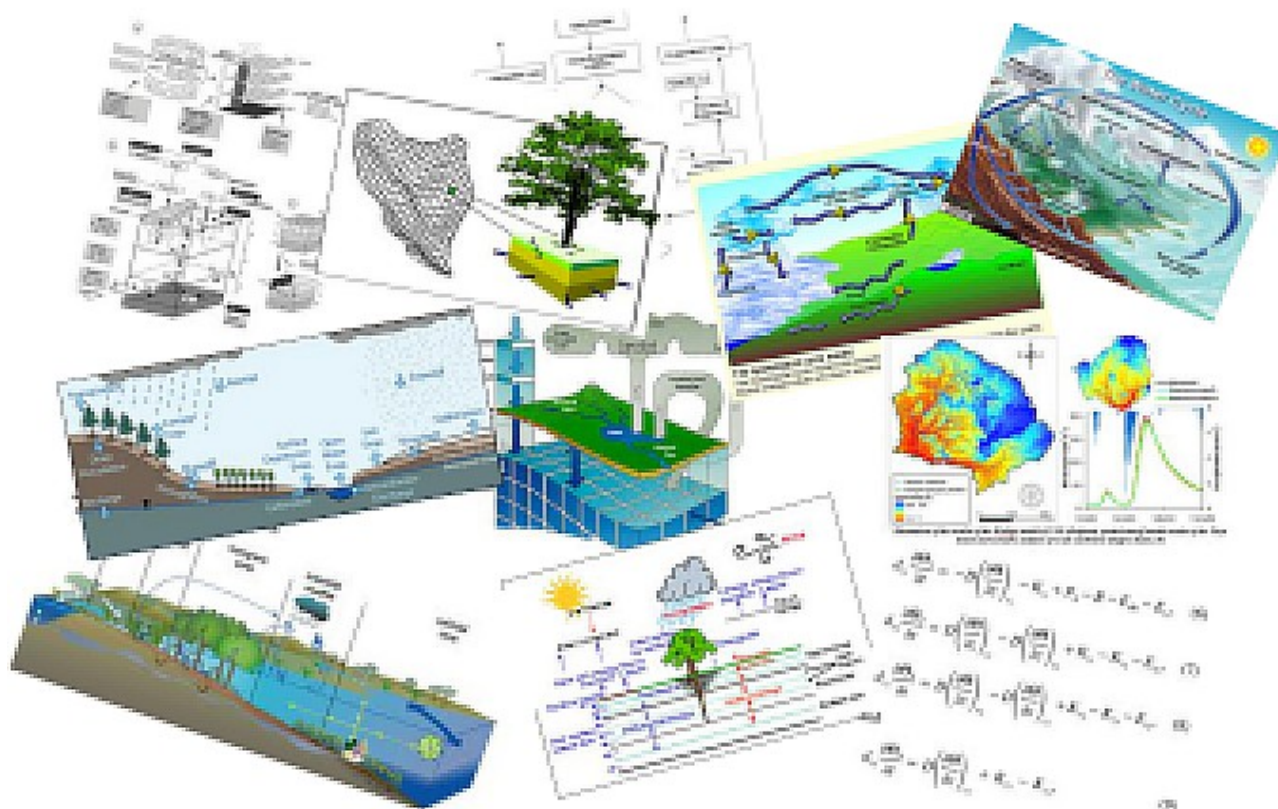
What is OS?

How to do OS?

Conclusions



Conclusions



SAD REALITY IN HYDROLOGY: 1% OF RESULTS REPRODUCIBLE

Stagge et al. (2019): **Assessing data availability and research reproducibility in hydrology and water resources.**

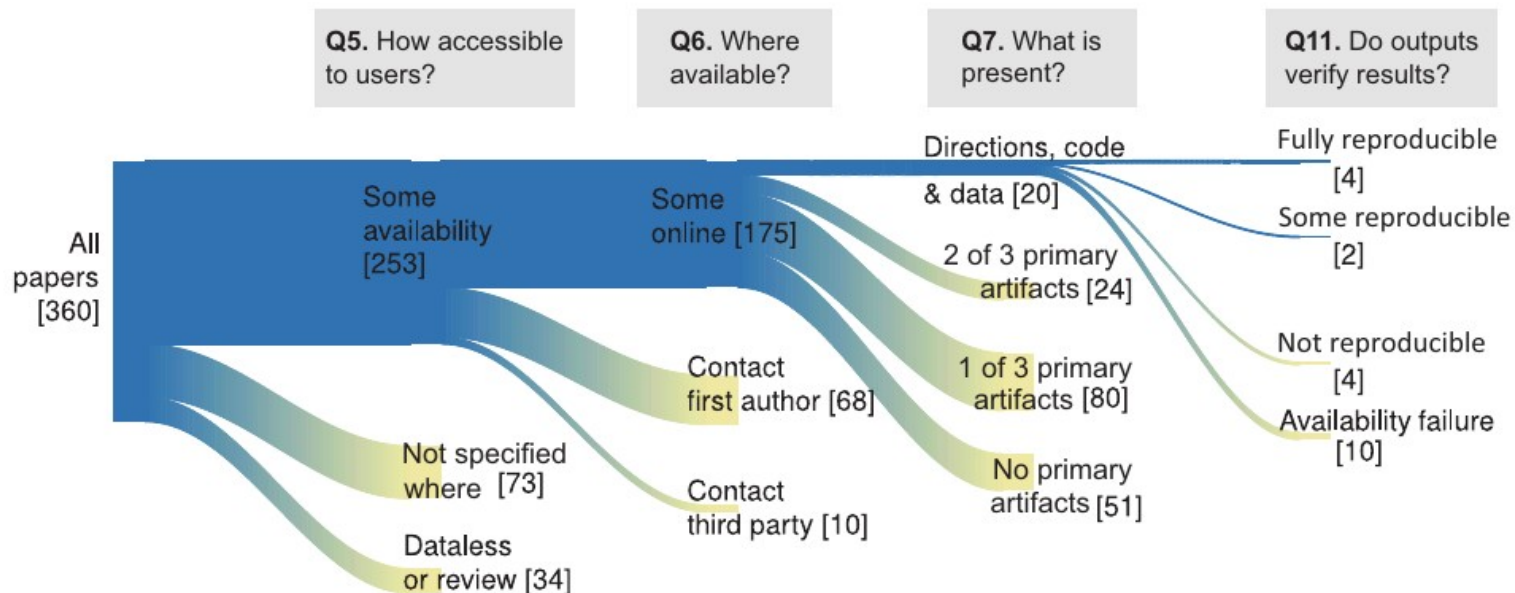
SCIENTIFIC DATA, 6:190030, <https://doi.org/10.1038/sdata.2019.30>

Why OS?

What is OS?

How to do OS?

Conclusions



SAD REALITY IN HYDROLOGY: 1% OF RESULTS REPRODUCIBLE

Stagge et al. (2019): **Assessing data availability and research reproducibility in hydrology and water resources.**

SCIENTIFIC DATA, 6:190030. <https://doi.org/10.1038/sdata.2019.30>

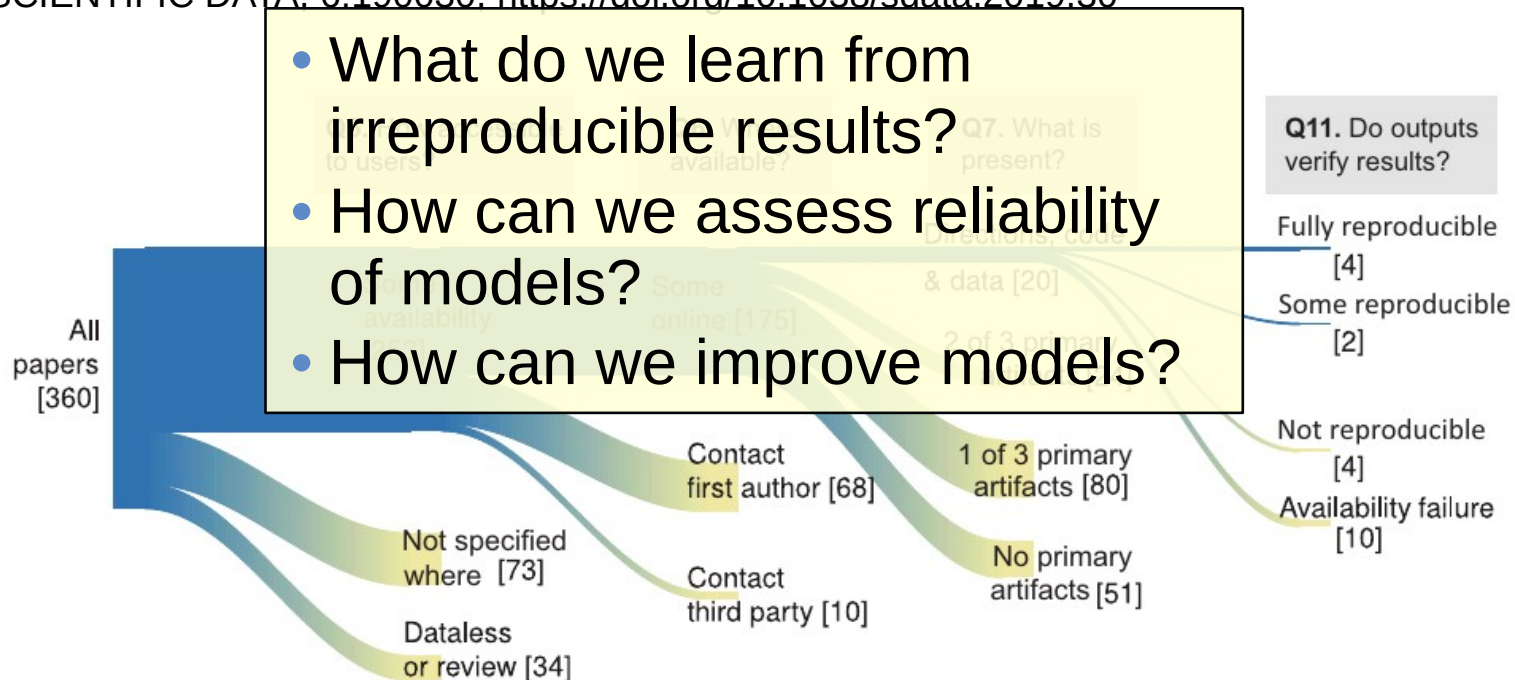
Why OS?

What is OS?

How to do OS?

Conclusions

- What do we learn from irreproducible results?
- How can we assess reliability of models?
- How can we improve models?



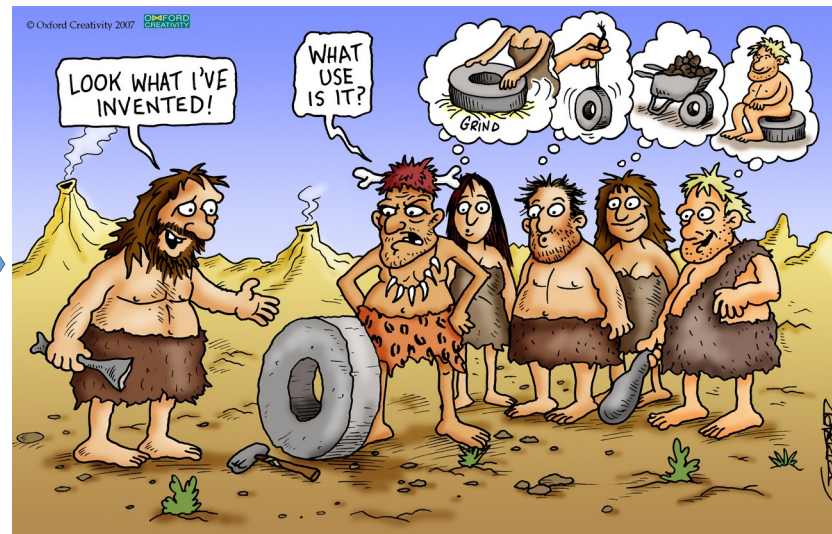
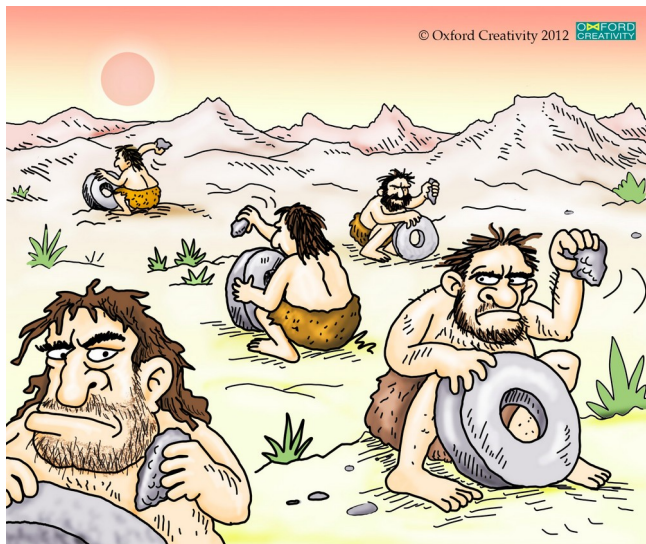
OPEN SCIENCE: SHARING → COMMUNITY → PROGRESS

Why OS?

What is OS?

How to do OS?

Conclusions



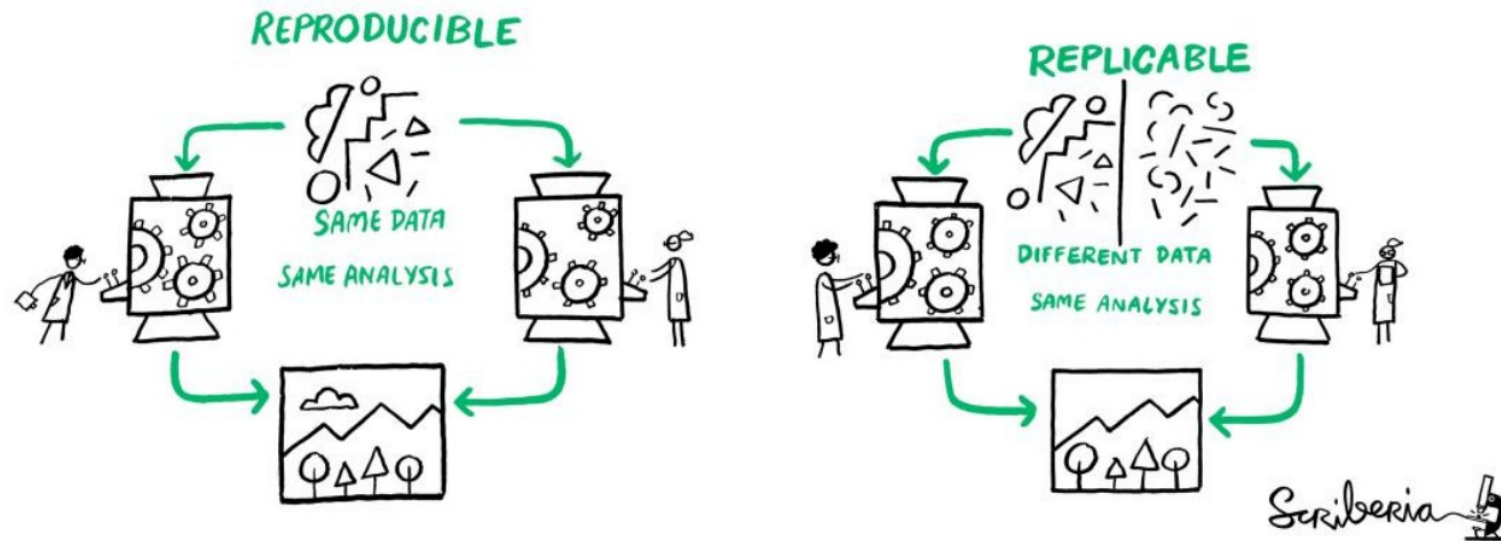
REPRODUCIBLE VS. REPLICABLE

Why OS?

What is OS?

How to do OS?

Conclusions



A HYDROLOGIST'S GUIDE TO OPEN SCIENCE

Why OS?
What is OS?
How to do OS?
Conclusions

For and by hydrology
community members.

Join the conversation at

[https://open-hydrology.](https://open-hydrology.github.io/)

[github.io/](https://open-hydrology.github.io/)

@Open_Hydrology

by:
Caitlyn A. Hall,
Sheila Saia,
Andrea Popp,
Nilay Dogulu,
Stan Schymanski,
Niels Drost,
Tim van Emmerik,
Lieke Melsen, and
Rolf Hut

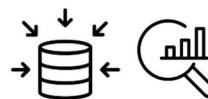
1
OPEN
RESEARCH
PROCESS AND
APPROACH



Principle 1

Open hydrologists intentionally plan for, describe and share the entire research approach and process from motivation to the final product.

2
OPEN DATA
COLLECTION
AND ANALYSIS



Principle 2

Open hydrologists document all components of their data collection and processing pipeline, favoring open and non-proprietary technologies.

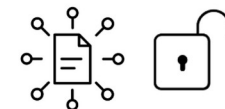
3
OPEN
SOFTWARE USE
AND
DEVELOPMENT



Principle 3

Open hydrologists test, archive, document, and version control their research software using standard open source software protocols and accessible documentation language.

4
OPEN
SCIENCE
PUBLISHING



Principle 4

Open hydrologists publish all components of their research on citable platforms and in journals that are accessible to both the research community and the general public following ethical standards.

<https://hess.copernicus.org/articles/26/647/2022/>

Why OS?

What is OS?

How to do OS?

Conclusions



ELEGANT, BUT WHERE EXACTLY
DOES EQ. 3 COME FROM?

OH, AND WHAT ARE THE
UNITS OF P_1 AND V_1 ?

ESSM

Python package for dealing with physical variables and units.

essm.readthedocs.org



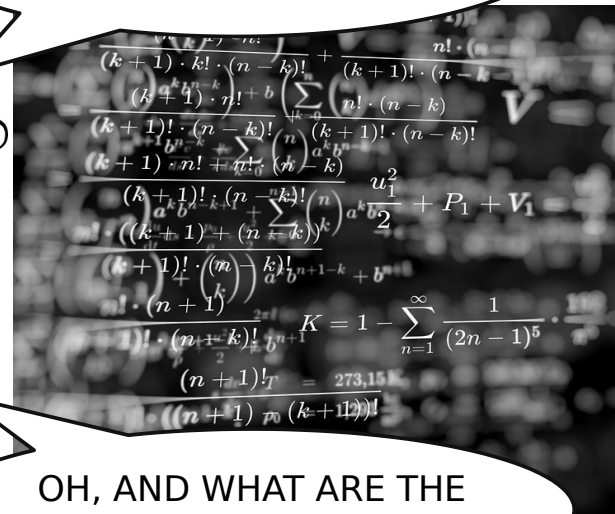
Environmental Science using Symbolic Math

DOI [10.5281/zenodo.4380979](https://doi.org/10.5281/zenodo.4380979) build failing coverage 97% tag v1.1 downloads 102/month
license GPL-2.0

This package contains help for dealing with physical variables and units.



ELEGANT, BUT WHERE EXACTLY DOES EQ. 3 COME FROM?



OH, AND WHAT ARE THE UNITS OF P_1 AND V_1 ?

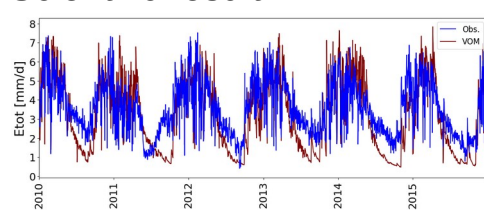
Why OS?

What is OS?

How to do OS?

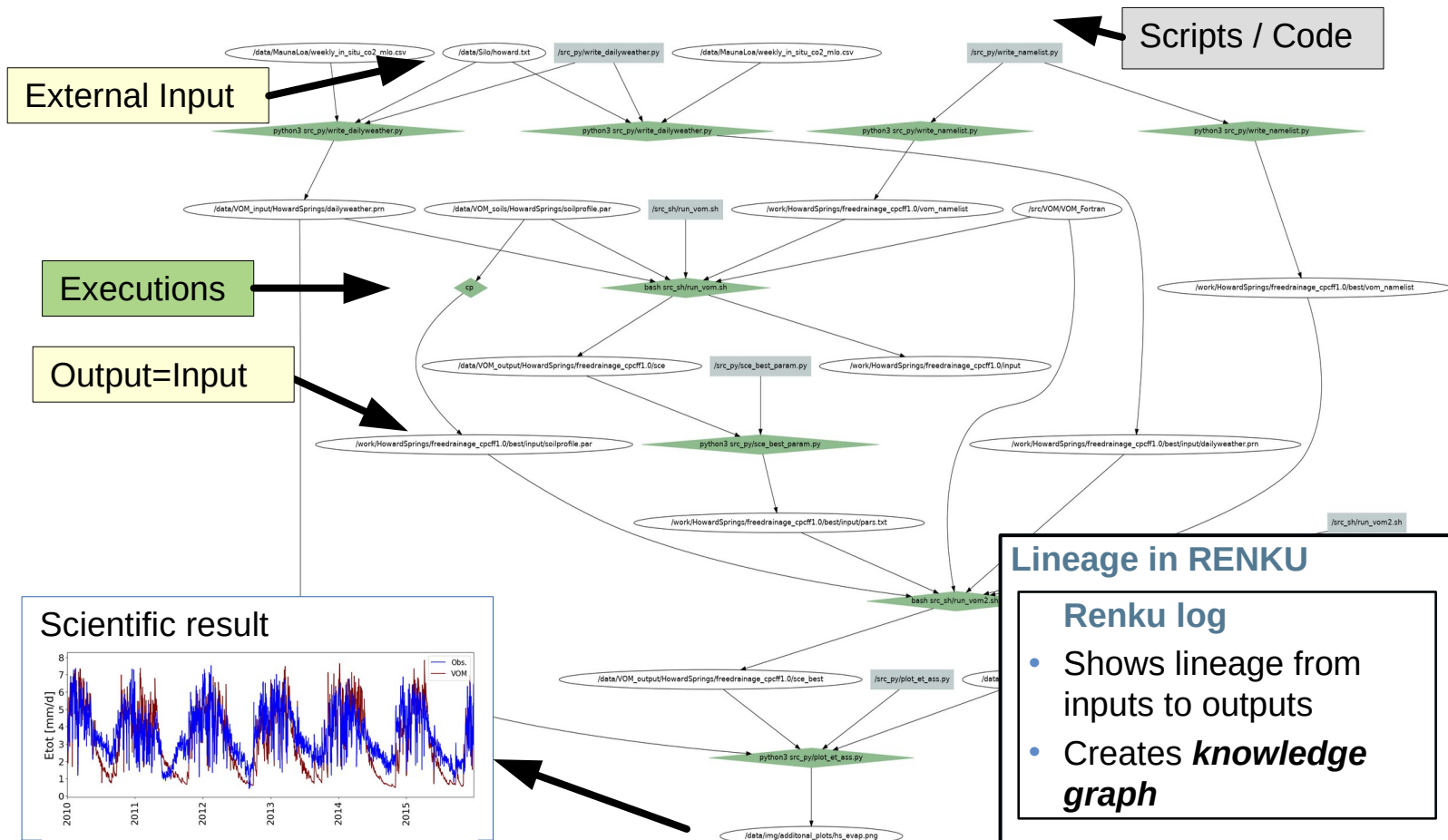
Conclusions

Scientific result



COOL, BUT WHAT WENT INTO THIS PLOT?

Why OS?
What is OS?
How to do OS?
Conclusions



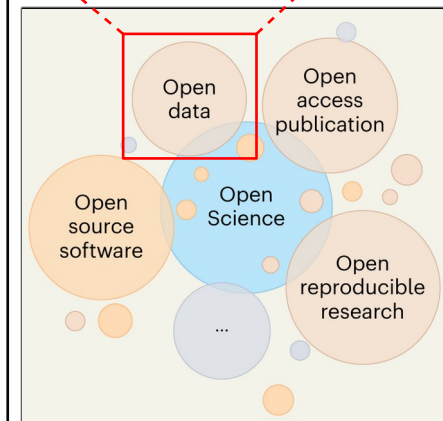
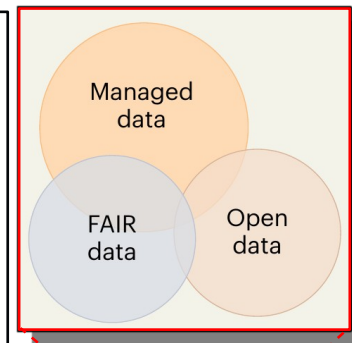
Why OS?

What is OS?

How to do OS?

Conclusions

- **Water science** has to be **Open Science** because:
 - Large scale problems require large **collaborative efforts** (e.g. **NORMAN-SLE**)
 - Collaboration builds on Findable, Accessible, Interoperable and Reusable (**FAIR**), **open** and well **documented** data & code
- **Open Science** builds on:
 - Open & FAIR **data** (e.g. HydroShare, zenodo.org, figshare)
 - Reproducible **mathematical derivations** (e.g. **essm.readthedocs.io**)
 - **Provenance** of data, code and results (e.g. **renkulab.io**)
 - Adequate **documentation** (e.g. **jupyter.org**, **RMarkdown**)
 - **Open Access** publication
 - **Open Source** software



Schymanski & Schymanski (2023):
Nat Water, 1, 4–6,
<https://doi.org/10.1038/s44221-022-00014-z>



Presented: Tue, 2 Feb 2023, 2-3pm, “Celebrating the launch of Nature Water - Part 3: Focus on social sciences and open science”
<https://cassyni.com/events/8ytB23CP1KG6yFL3TVRB5o>

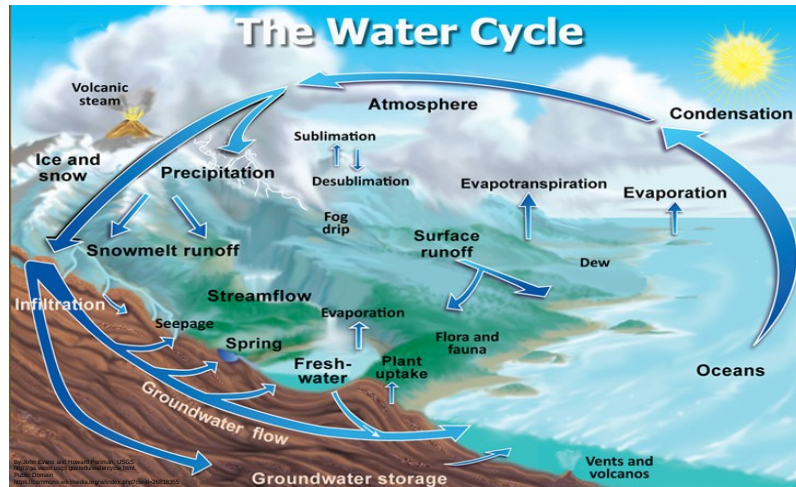
I am starting the presentation with a provocative statement, directly lifted from an opinion paper in the first issue of nature water by my wife Emma, who unfortunately cannot attend, and myself: Water science must be Open Science. In the next 10+ minutes I will try to explain why we are saying this, what we mean by open science and how Open Science can be achieved.

Why OS?

What is OS?

How to do OS?

Conclusions



2

What is special about water science?

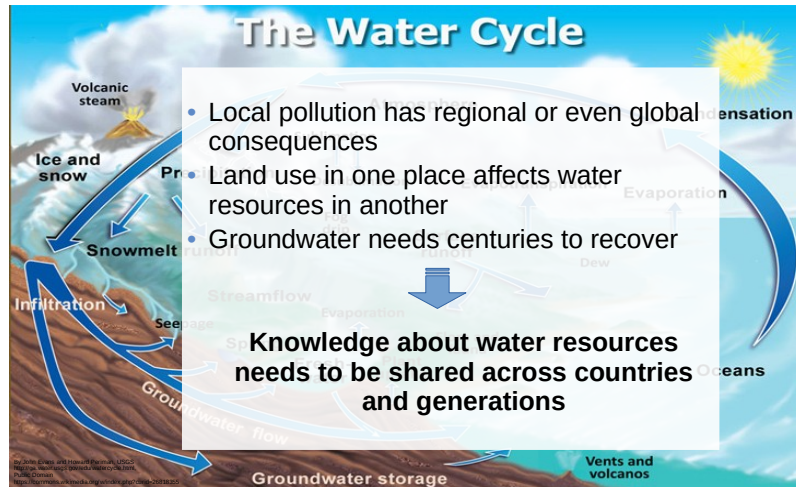
Reminding ourselves of the global water cycle, we have to recognise that water is essential for life at the local scale, but at the same time crosses borders and links communities and generations. For example, the atmospheric side of the water cycle is very sensitive to global CO₂ emissions, surface waters flow over 1000s of kilometers, whereas groundwater essentially stays in place for much more than our life times.

Why OS?

What is OS?

How to do OS?

Conclusions



This means that for example local pollution has regional or even global consequences, that land use in one place affects water resources in another, and groundwater resources have a “memory” of 100s to 1000s of years, as this is how long it takes for them to be replenished, especially in arid environments, where we most strong rely on them. Therefore, knowledge about water resources needs to be shared across countries and generations, i.e. in space and time.

EXAMPLE: FISH KILL IN ODER RIVER (GERMANY)

Why OS?

What is OS?

How to do OS?

Conclusions

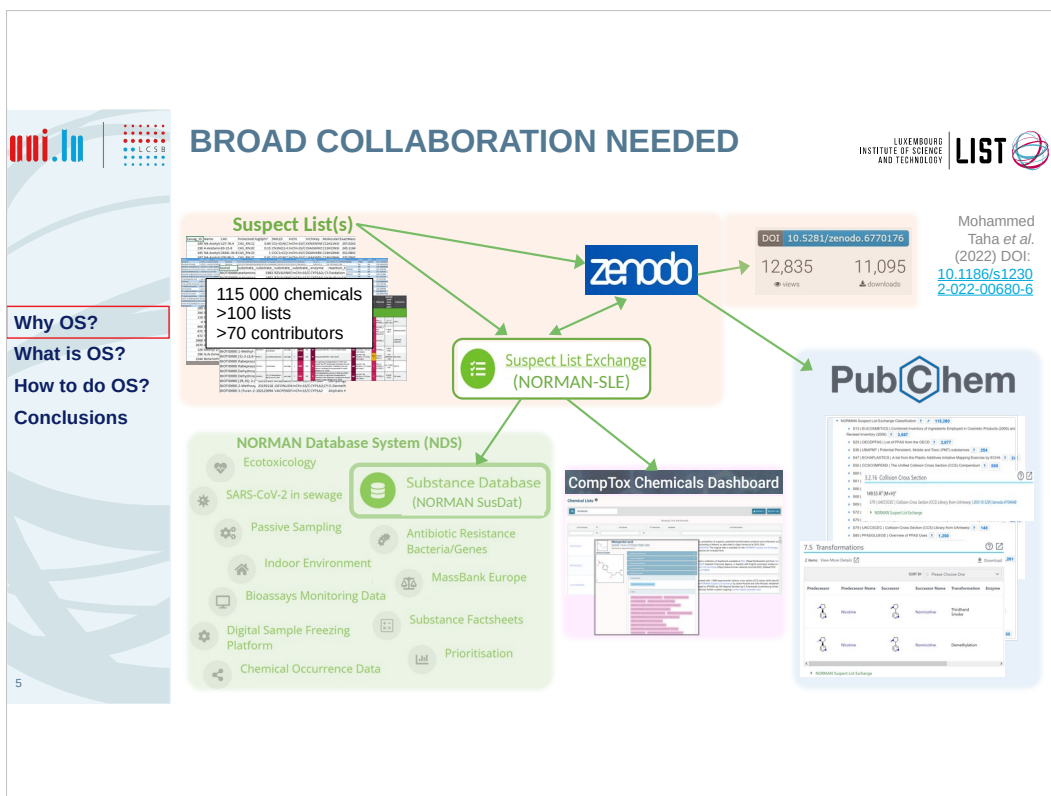
Due to spill in Poland?

- High salinity
- Herbicides
- >1200 known chemicals analysed



<https://www.tagesschau.de/ausland/europa/oder-fischsterben-polen-algen-103.html>

A recent example for the regional consequences of local pollution is the epic fish kill in the river Oder in Germany last year. A frantic search for possible reasons pointed to some sort of a spill in Poland, as it coincided with an extreme spike in salinity, but researchers also found herbicides and despite screening for over 1200 known chemicals, the exact cause of the fish death is still not clear. Currently the understanding is that the spill along with high temperatures caused massive growth of a toxic algae, which likely killed more than 50% of the fish.



In fact, to identify sources of catastrophic or, perhaps worse, slow and persistent pollution, requires a huge community effort, such as the NORMAN Suspect List Exchange (NORMAN-SLE). This community effort now includes information on >115 thousand chemicals of concern from over 100 different suspect lists submitted by more than 70 contributors. This data is shared openly on zenodo and is integrated into other major chemical resources (CompTox, PubChem) to increase dissemination of this information to help understand and track pollutants in the environment.

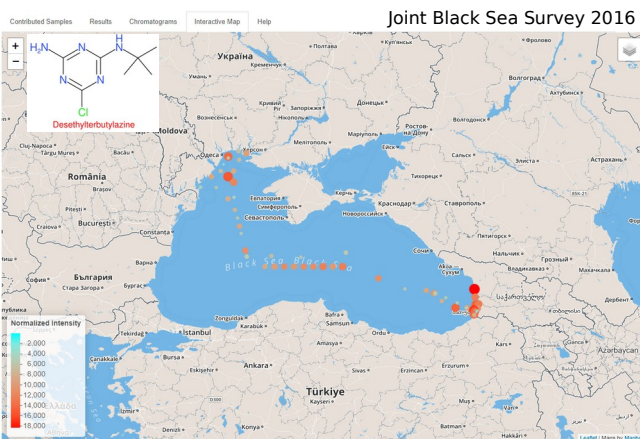
NORMAN DIGITAL SAMPLE FREEZING PLATFORM

Why OS?

What is OS?

How to do OS?

Conclusions



Find out
where the
suspect is
hiding ...

For example, in the NORMAN digital sample freezing platform, it is possible to select a chemical and generate a map of where this chemical has already been detected in the environment, such as shown here for the Black Sea.




SUCCESS STORY



Why OS?

What is OS?

How to do OS?

Conclusions




A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon

ZHENYU TIAN¹, HAOQI ZHAO¹, KATHERINE T. PETER¹, MELISSA GONZALEZ¹, JILL WETZEL¹, CHRISTOPHER WU¹, XIMIN HU¹, JASMINE PRAT¹, EMMA MUDROCK¹, L. J. AND EDWARD P. KOLODZIEJ¹

[+17 authors](#) [Authors Info & Affiliations](#)

SCIENCE • 3 Dec 2020 • Vol 371, Issue 6525 • pp. 185-189 • DOI: 10.1126/science.abd6951

PubChem N-(1,3-Dimethylbutyl)-N'-phenyl-p-phenylenediamine

7.1 Transformations

1 item View More Details

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	6PPD		6PPD-quinone	Ozone	

[NORMAN Suspect List Exchange](#)

In another fish kill case, this community-powered open science approach has helped identify the cause, which was a transformation product of a ubiquitous chemical contained in car tyres. Imagine the millions of chemical compounds produced in large quantities and eventually exposed to the environment, and their potential transformation products to grasp the effort needed to link the dots.

WATER QUANTITY: DATA, MODELS AND PREDICTION

Why OS?

What is OS?

How to do OS?

Conclusions

Forecasting of floods and water shortage relies on multidisciplinary **data**:

- climate
- topography
- soil & geo(morph)ology
- vegetation
- land use
- water abstraction...

... and reliable simulation **models** !



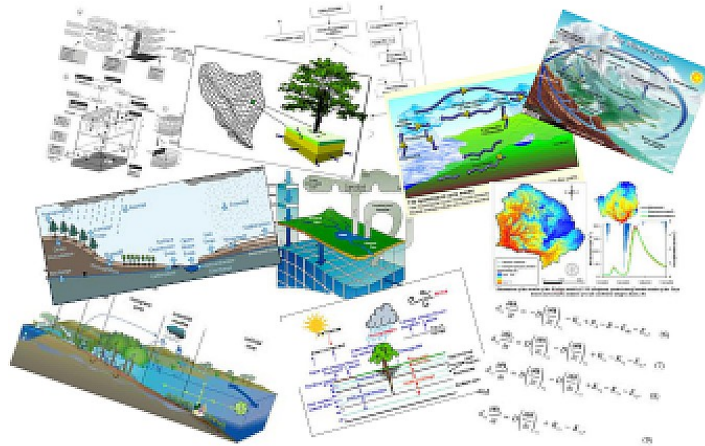
Now from water quality to quantity. Here we need to combine many different, multidisciplinary data sources, such as climate, topography, soil and geology, vegetation, land use, water abstraction etc. with simulation models to generate accurate flood warnings and generate flood and drought risk maps.

Why OS?

What is OS?

How to do OS?

Conclusions



However, we have the big problem that there are tens if not hundreds of hydrological models around. Essentially we could probably pick a different model for each data set, until it all becomes a blur.

SAD REALITY IN HYDROLOGY: 1% OF RESULTS REPRODUCIBLE

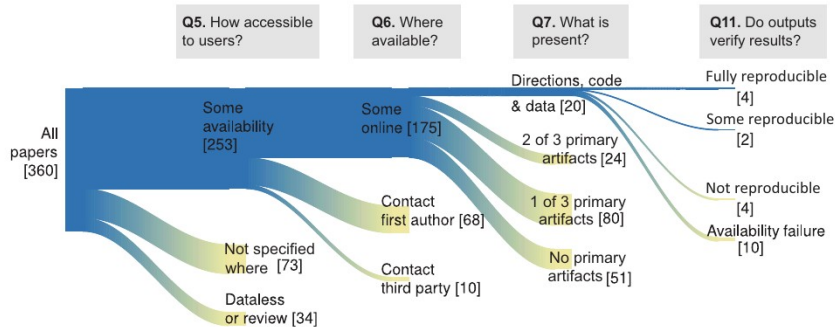
Stagge et al. (2019): **Assessing data availability and research reproducibility in hydrology and water resources.**
SCIENTIFIC DATA, 6:190030, <https://doi.org/10.1038/sdata.2019.30>

Why OS?

What is OS?

How to do OS?

Conclusions



To make matters worse, hydrologists do not often share their models and data. A study in 2019 revealed that out of 360 randomly sampled hydrology papers published in 2017, only 4 (i.e. 1%) were fully reproducible. Half of them already failed at the general data availability check, most of the others had incomplete supporting information to enable reproduction of results.

SAD REALITY IN HYDROLOGY: 1% OF RESULTS REPRODUCIBLE

Stagge et al. (2019): **Assessing data availability and research reproducibility in hydrology and water resources.**
SCIENTIFIC DATA, 6:190030. <https://doi.org/10.1038/sdata.2019.30>

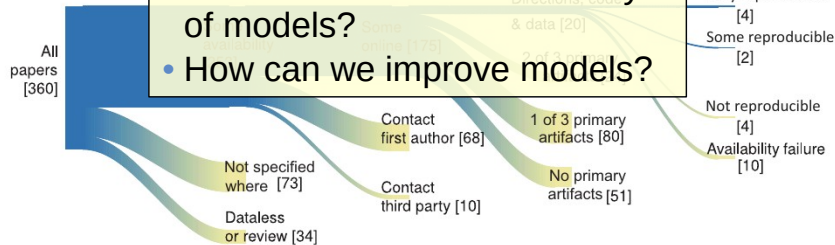
Why OS?

What is OS?

How to do OS?

Conclusions

- What do we learn from irreproducible results?
- How can we assess reliability of models?
- How can we improve models?



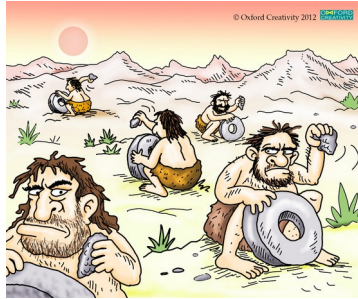
But what do we learn from irreproducible results? How can we...

Why OS?

What is OS?

How to do OS?

Conclusions



Artwork: Oxford Creativity, www.triz.co.uk

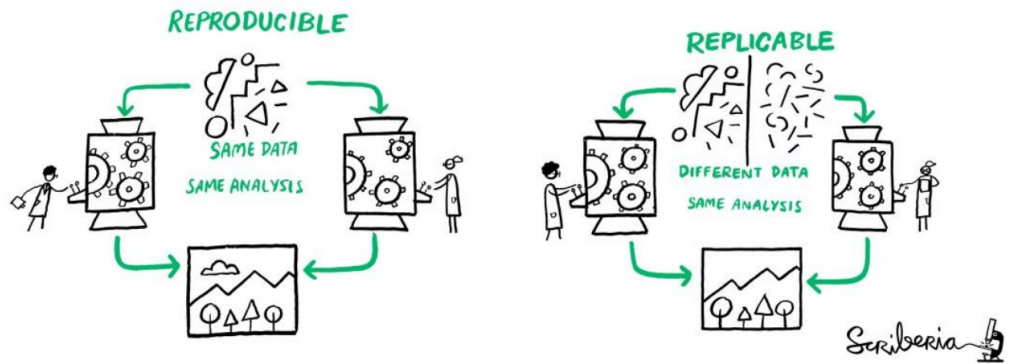
Essentially, everyone is working on their pet model and re-inventing the wheel. And this is why we need open science, where individual inventions can be not only scrutinized by many, but also re-used and adapted by many to solve the problem they are working on, which will benefit the global problems related to water science.

Why OS?

What is OS?

How to do OS?

Conclusions



European Commission. Directorate General for Research and Innovation. (2020). Reproducibility of scientific results in the EU: Scoping report. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/341654>

The problem is of course not limited to Hydrology and water science, which is illustrated by the fact that the European Commission has recently launched a general Open Science initiative, where they go beyond reproducibility of results and ask for results to be replicable, i.e. applicable to new data, which requires that the methods and tools are actually generalisable and re-usable.

Why OS?

What is OS?

How to do OS?

Conclusions

For and by hydrology
community members.

Join the conversation at

<https://open-hydrology.github.io/>

[@Open_Hydrology](https://twitter.com/Open_Hydrology)

by:
Caitlyn A. Hall,
Sheila Saia,
Andrea Popp,
Nilay Dogulu,
Stan Schymanski,
Niels Drost,
Tim van Emmerik,
Lieke Melsen, and
Rolf Hut

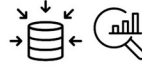
1
OPEN
RESEARCH
PROCESS AND
APPROACH



Principle 1

Open hydrologists intentionally plan for, describe and share the entire research approach and process from motivation to the final product.

2
OPEN DATA
COLLECTION
AND ANALYSIS



Principle 2

Open hydrologists document all components of their data collection and processing pipeline, favoring open and non-proprietary technologies.

3
OPEN
SOFTWARE USE
AND
DEVELOPMENT



Principle 3

Open hydrologists test, archive, document, and version control their research software using standard open source software protocols and accessible documentation language.

4
OPEN
SCIENCE
PUBLISHING



Principle 4

Open hydrologists publish all components of their research on citable platforms and in journals that are accessible to both the research community and the general public following ethical standards.

<https://hess.copernicus.org/articles/26/647/2022/>

To help hydrologists navigate the jungle of open science, in a group of early and mid-career hydrologists we wrote a Hydrologists Guide to Open Science, which is now published in HESS, and we also created a web site where the ideas, tools and approaches can be discussed, and where we can officially self-declare as Open Hydrologists.

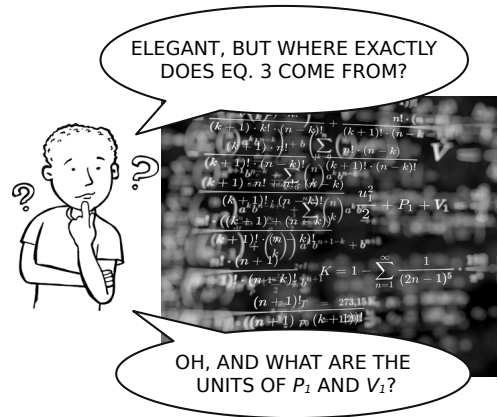
Why OS?

What is OS?

How to do OS?

Conclusions

15



There are tools to assist open science right from the first step, where reproducibility often breaks already when we try to follow a beautiful mathematical derivation but get stuck at a particular derivation step. Anyone who has tried to do this, probably has had this experience. And then after we decide to just trust the derivation, we find out that the units of the variables were not defined, so we have to start guessing before we can use the equation for any computations.



Why OS?
What is OS?
How to do OS?
Conclusions

ESSM

Python package for dealing with physical variables and units.

essm.readthedocs.org

 **jupyter** jupyter.org

 **python**
python.org

 **SymPy**
sympy.org

Environmental Science using Symbolic Math

DOI [10.5281/zenodo.4380979](https://doi.org/10.5281/zenodo.4380979)
build failing
coverage 97%
tag v1.1
downloads 102/month

license GPL-2.0

This package contains help...



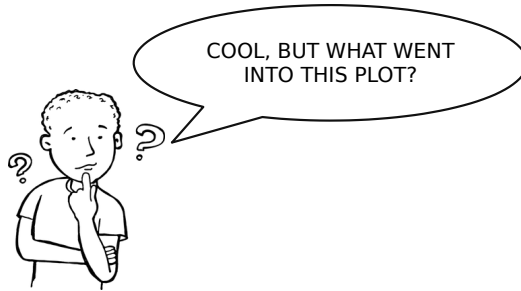
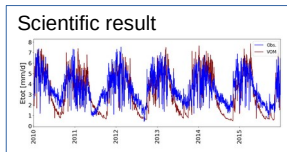


ELEGANT, BUT WHERE EXACTLY DOES EQ. 3 COME FROM?

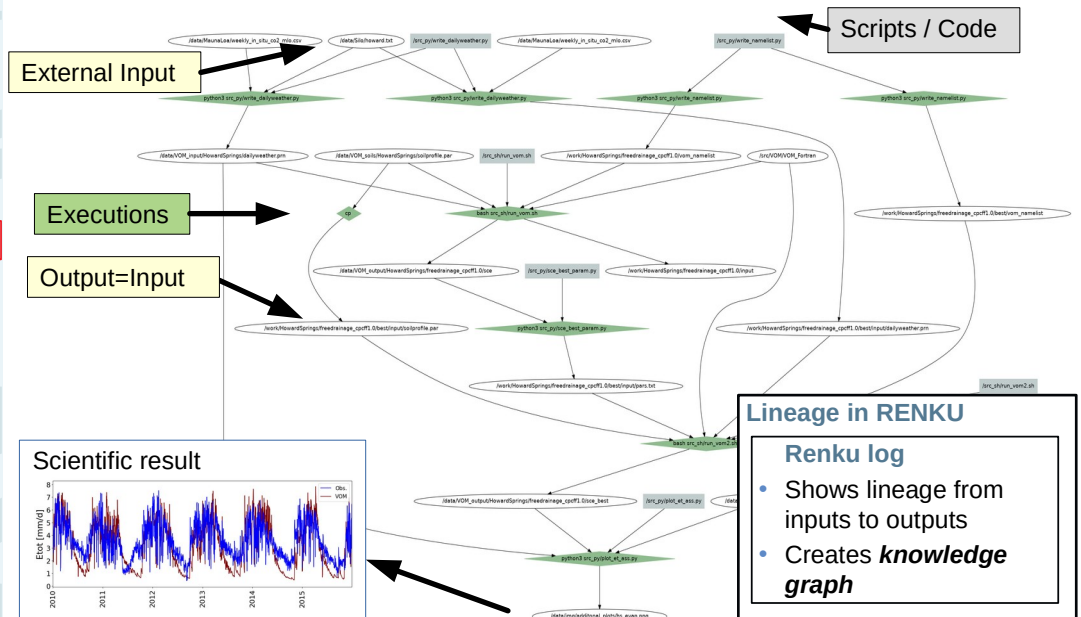


OH, AND WHAT ARE THE UNITS OF P_i AND V_i ?

Therefore, together with a former colleague, I have developed a Python package called ESSM, which we use in jupyter notebooks to transparently perform mathematical derivations, with automatic testing for dimensional consistency and a structured way of defining variables, their meaning and units, and inter-connected equations with their underlying assumptions.



The next step is how we get from data and code to results. How often do we see some cool results in a paper and ask ourselves what actually went into these plots?



Why OS?

What is OS?

How to do OS?

Conclusions

Here the platform renku is of great help, as it stores all the lineage information in a knowledge graph, as shown here. You can clearly see the external input, distinguished from scripts and code, the green diamonds are code executions, which then result in new data or plots. If you change one of the files, renku will tell you which part of the analysis has to be re-run to update all dependencies and can do it for you with just one command. This not only helps the future you in understanding what you did, but also others to reproduce and reuse your analysis on their data or in their workflows.

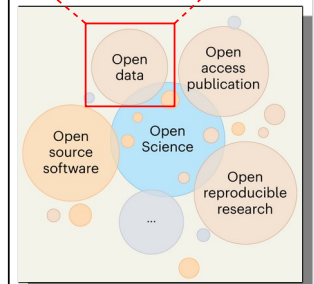
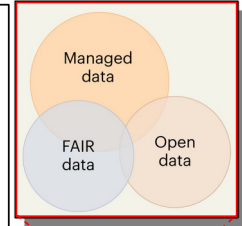
Why OS?

What is OS?

How to do OS?

Conclusions

- **Water science** has to be **Open Science** because:
 - Large scale problems require large **collaborative efforts** (e.g. **NORMAN-SLE**)
 - Collaboration builds on Findable, Accessible, Interoperable and Reusable (**FAIR**), **open** and well **documented** data & code
- **Open Science** builds on:
 - Open & FAIR **data** (e.g. **HydroShare**, **zenodo.org**, **figshare**)
 - Reproducible **mathematical derivations** (e.g. **essm.readthedocs.io**)
 - **Provenance** of data, code and results (e.g. **renkulab.io**)
 - Adequate **documentation** (e.g. **jupyter.org**, **RMarkdown**)
 - **Open Access** publication
 - **Open Source** software



Schymanski & Schymanski (2023):
Nat Water, 1, 4–6,
<https://doi.org/10.1038/s44221-022-00014-z>

So to conclude, we believe that water science has to be Open Science because large scale problems require large collaborative efforts, and because collaboration builds on FAIR, open and well documented data and code. This is also what Open Science builds on, where Open and FAIR data can be shared on e.g. HydroShare, zenodo or figshare, mathematical derivations can be made reproducible and traceable using ESSM, the provenance of data, code and results can be tracked in e.g. renkulab.io, and adequate documentation can be written conveniently and close to the code when using Jupyter notebooks or RMarkdown. Of course Open Access publications and Open Source software are also essential ingredients for Open Science.