



THE 5 W'S (AND THE H)

Open Data in Climate Research

Rocío Castaño-Primo, Benjamin Pfeil
Geophysical Institute
Bjerknes Climate Data Center
rocio.primo@uib.no

Open data. WHY?

Eat your veggies: legal requirements from funding agencies

EU

Horizon 2020 already mandates open access to all scientific publications



From 2017, research data is **open by default**, with possibilities to **opt out**

HORIZON 2020 GRANTEES ARE REQUIRED

take measures to ensure open access to the data underlying their scientific publications

provide open access to any other research data of their choice

Horizon 2020 grantees are encouraged to also share datasets beyond publication



Main principle

Open access

The Research Council's policy follows an open-by-default principle regarding access to research data.

[ISBN: 978-82-12-03660-4](#)

RCN

https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf



Open data. WHY?

More money, more data.

RCN NORKLIMA (2007-2011)

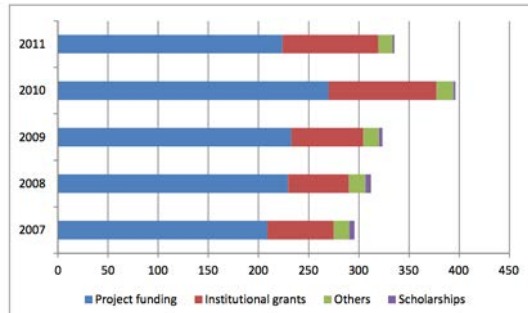
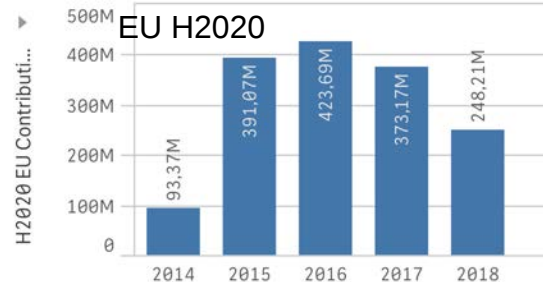
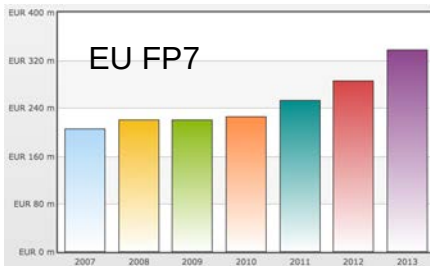
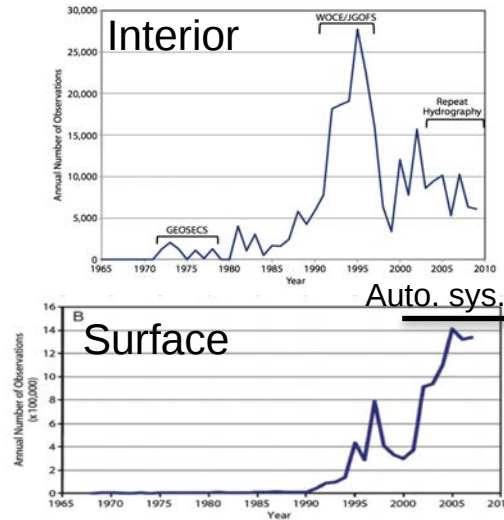


Figure 2.3.5. RCN funding for climate research by types of funding, 2007-2010. MNOK.
RCN KLIMAFORSK 2017: **153 mill**

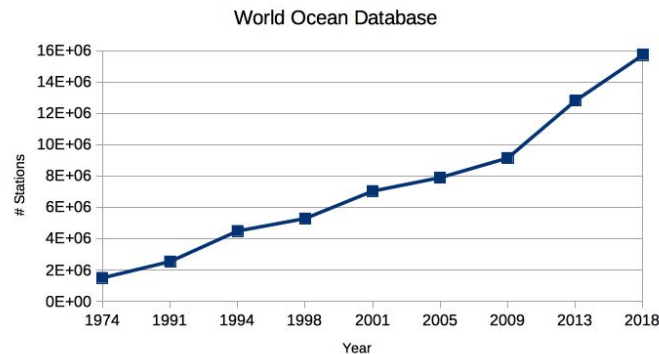


CO₂ # observations



Sabine et al. (2010)

Oceanographic # stations



Coupled Model
Intercomparison Program

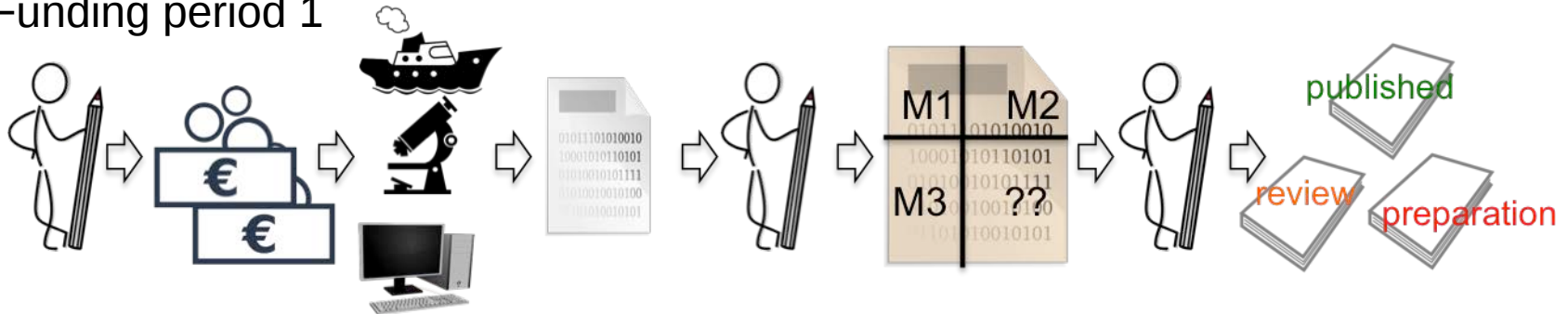


	#mod	#exp	size
CMIP	19	1	1 Gb
CMIP2	24	2	500 Gb
CMIP3	21	12	40 Tb
CMIP5	45	110	2 Pb
CMIP6	>32	?	???

Open data. WHY?

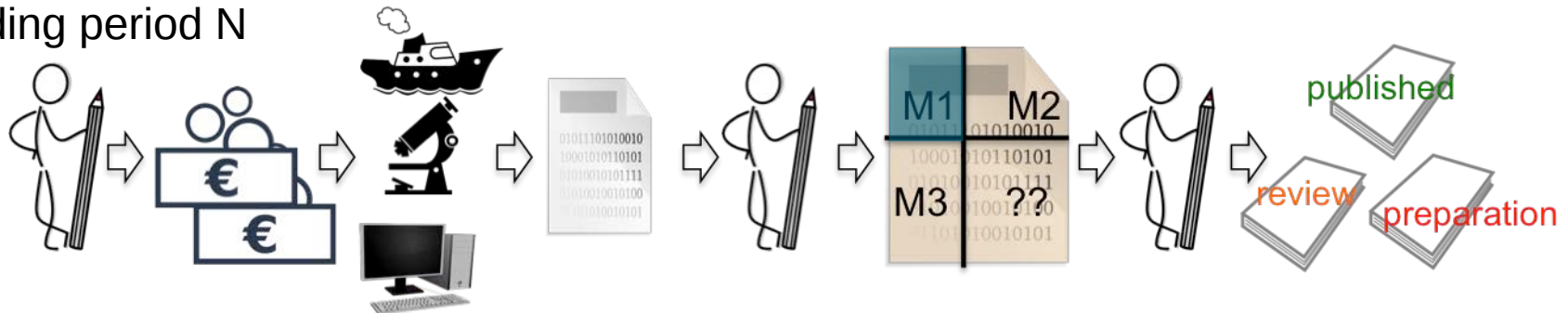
What is done with the data?

Funding period 1



...

Funding period N



Public / institutional repository?

Stored privately?

Lost?

~(ツ)~

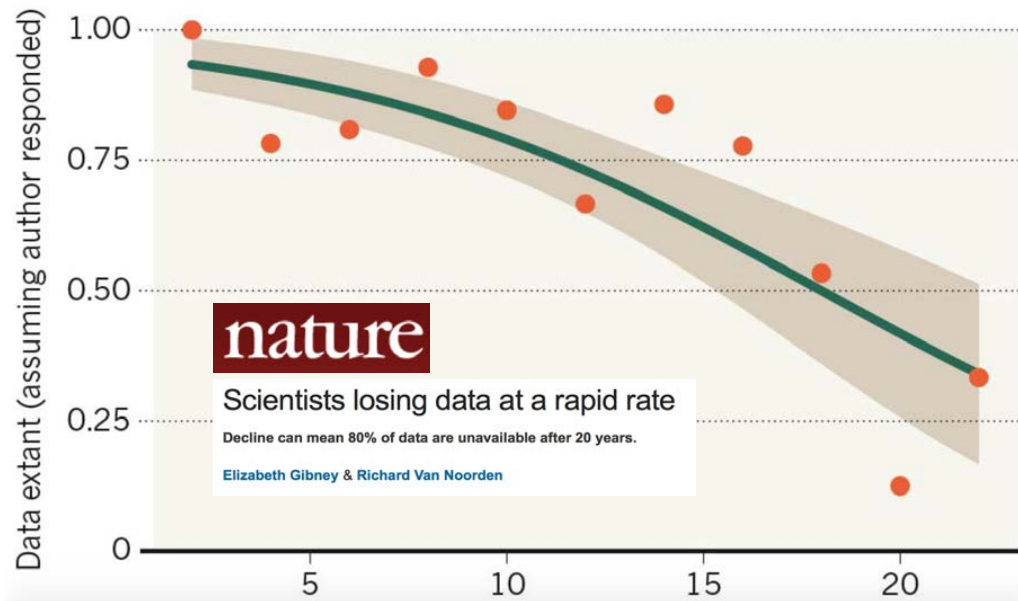


Open data. WHY?

Data gets lost

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



[Gibney and Noorden \(2013\). doi: 10.1038/nature.2013.14416](https://doi.org/10.1038/nature.2013.14416)

[Vines et al. \(2013\). doi: 10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014)



- Email not working
- No reply
- Do not know where
- Cannot access:
 - someone's attic
 - floppy disk
- Unwilling to share



- Unintelligible
- Non-standard format
- Physically destroyed

Old data lost → time series truncated

Open data. WHY?

Reproducibility, replicability, accountability



Most scientists 'can't replicate studies by their peers'

By Tom Feilden
Science correspondent, Today programme

© 22 February 2017



Opinion

GRAY MATTER

The New York Times

Why Do So Many Studies Fail to Replicate?

By Jay Van Bavel

May 27, 2016



MICHAEL SCHULSON SCIENCE 04.20.18 09:00 AM

WIRED

SCIENCE'S "REPRODUCIBILITY CRISIS" IS BEING USED AS POLITICAL AMMUNITION

Why we can't trust academic journals to tell the scientific truth

The Guardian
International edition

Academic journals don't select the research they publish on scientific rigour alone. So why aren't academics taking to the streets about this?

BJERKNES CENTRE
For Climate Research



nature
geoscience

Editorial | Published: 01 August 2010

Climategate closed

Nature Geoscience 3, 509 (2010) | Download Citation ↓

Climate science at the University of East Anglia is sound but lacking in transparency, according to the three official reports. But making data accessible will not be sufficient to guard against future attacks.

nature
climate change

Feature | Published: 29 March 2011

Data on demand

Kurt Kleiner ✉

Nature Climate Change 1, 10–12 (2011) | Download Citation ↓

✎ A Correction to this article was published on 06 May 2011

✎ This article has been updated

Climate scientists are under pressure to make their data — and their methods — more openly available, both to fellow scientists and the public. Now, open-access climate science is becoming easier than ever.



Nick Anderson (2009). Houston Chronicle

Open data. WHY?

Standing on the shoulders of giants



+



+

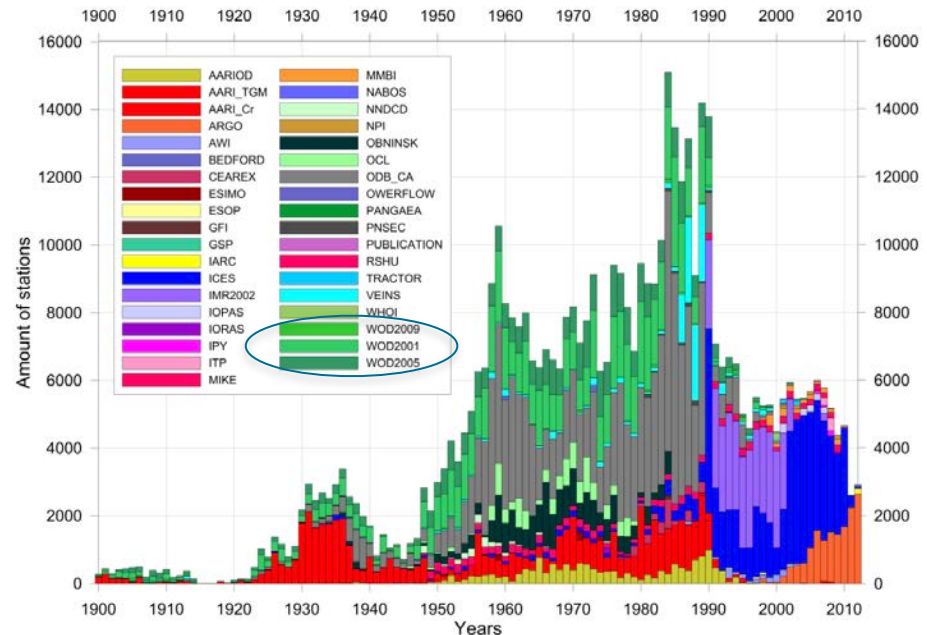
PACIFIC ocean Interior CARbon
(PACIFICA)

+ more
data



glodap_{v2}

Climatological atlas of the
Nordic Seas and northern
North Atlantic



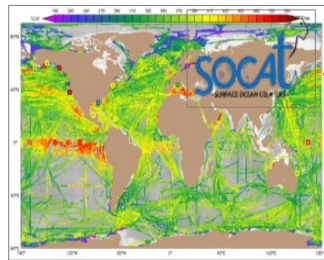
[Korablev et al. \(2014\) doi: 10.7289/V54B2Z78](#)



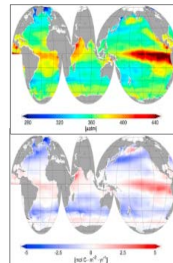
Open data. WHY?

Big questions, big data

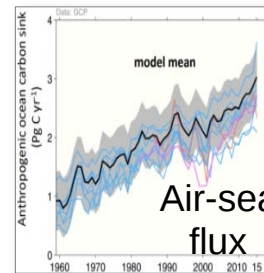
Global Carbon Budget



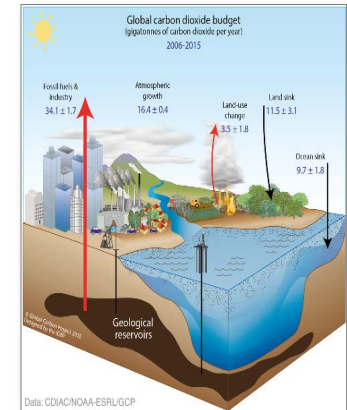
Assessment



Estimation



+ much more



Le Queré et al. (2018). doi: 10.5194/essd-2017-123

MODELS



NorESM

cam : atmospheric package (AMWG)
clm : land package (LMWG)
cice : sea-ice package
micom : ocean package
hamocc : biogeochemistry package

Tracer	Abbreviation	Unit
Carbonate system		
Dissolved inorganic carbon	DIC	kmol C m^{-3}
Total alkalinity	TA	keq m^{-3}
Nutrients		
Phosphate	PO_4	kmol m^{-3}
Nitrate	NO_3	kmol m^{-3}
Silicate	Si	kmol m^{-3}
Dissolved iron	dFe	kmol m^{-3}
Ecosystem state		
Phosphorus in phytoplankton	PHY	kmol P m^{-3}
Phosphorus in zooplankton	ZOO	kmol P m^{-3}
Dissolved organic carbon	DOC	kmol P m^{-3}
Particulate organic carbon	POC	kmol P m^{-3}
Particulate inorganic carbon	PIC (CaCO_3)	kmol C m^{-3}
Biogenic silica (opal)	BSi	kmol S m^{-3}
Number of aggregates ^{a,b}	NOS	cm^{-3}
Mass of dust in aggregates ^{a,b}	aDUST	kg m^{-3}
Mass of non-aggregated dust	IDUST	kg m^{-3}
Dissolved gases		
Oxygen	O_2	kmol m^{-3}
Nitrogen	N_2	kmol m^{-3}
Nitrous oxide	N_2O	kmol m^{-3}
Trichlorofluoromethane ^b	CFC-11	kmol m^{-3}
Dichlorodifluoromethane ^b	CFC-12	kmol m^{-3}
Sulfur hexafluoride ^b	SF_6	kmol m^{-3}
"Preformed" tracers^b (model diagnostics)		
Preformed oxygen	prO ₂	kmol m^{-3}
Preformed phosphate	prPO ₄	kmol m^{-3}
Preformed alkalinity	prALK	kmol m^{-3}

OBSERVATIONS

Forcing
Initialization
Model validation
Data assimilation



BJERKNES CENTRE
For Climate Research



Schwinger et al. (2016). doi: 10.5194/gmd-9-2589-2016

Open data. WHY?

Fine, but what's in it for me?

Data **IS** a research product!

Data can and **MUST** be cited → citations can be tracked

Data citation is becoming real with FORCE¹¹ and Elsevier

At Boston workshop, Elsevier detailed its efforts to credit researchers who make their data available by supporting research data citation

Elsevier.com

Data literature review

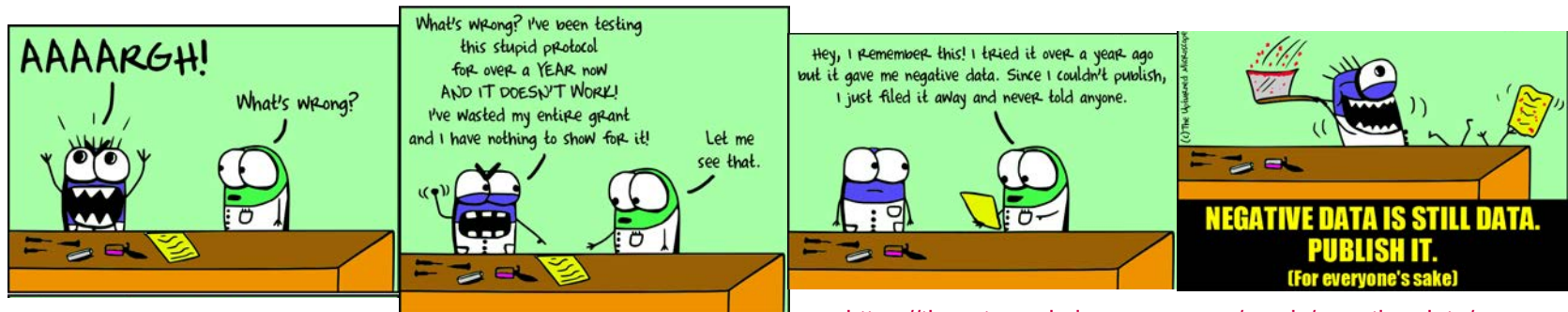
Explore potential hypothesis

Inter-disciplinary research

Avoid unnecessary replication



Web of Science

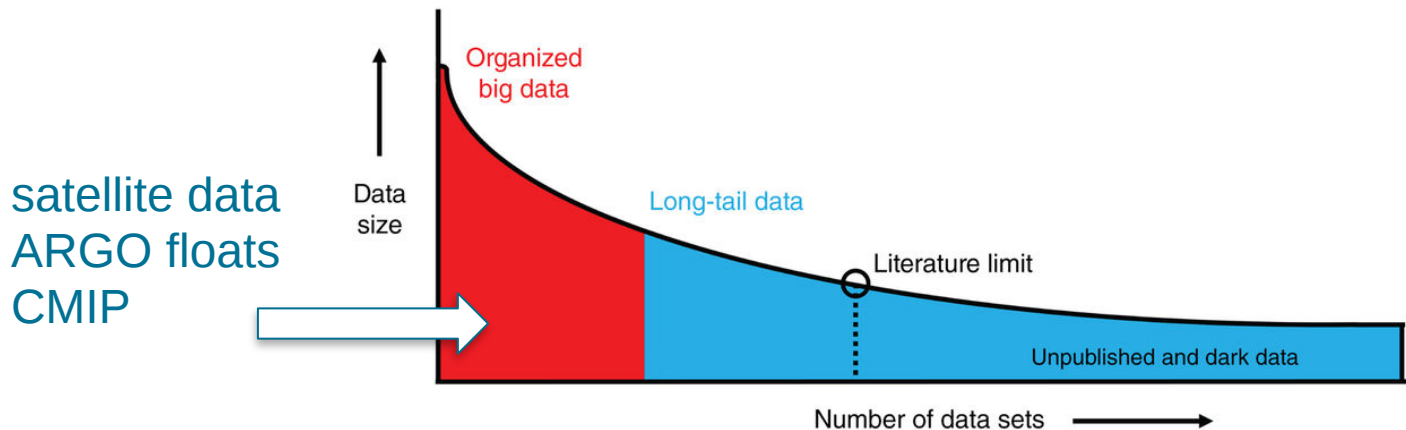


<https://theupturnedmicroscope.com/comic/negative-data/>



Open data. WHAT?

The more the merrier! Not all data is used/published



[Ferguson et al. \(2014\) doi: 10.1038/nm.3838](https://doi.org/10.1038/nm.3838)

HORIZON 2020 GRANTEES ARE **REQUIRED**

take measures to ensure open access to the data underlying their scientific **publications**

provide open access to any other research data of their choice

Horizon 2020 grantees are **encouraged to also share datasets beyond publication**

private data
intellectual property
embargo period

as open as possible, as closed as necessary



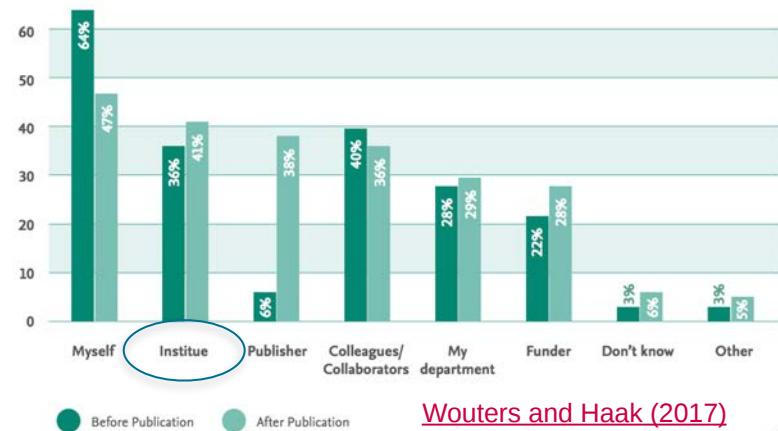
Open data. WHO?



[Auke Herrema \(2014\). 4th RDA plenary meeting](#)

Who owns research data?

Figure 3. Research data ownership before and after publication (% , n=1162)



Data manager / curator / steward

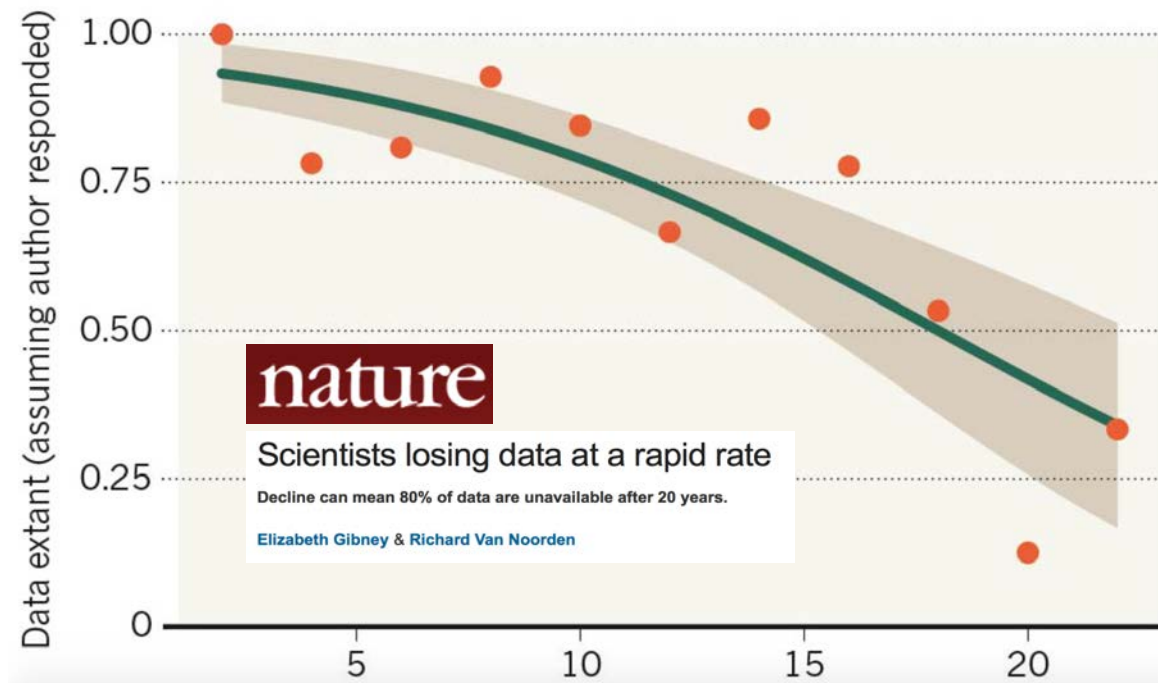


bcdc.no

Open data. WHEN?

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



[Gibney and Noorden \(2013\). doi: 10.1038/nature.2013.14416](https://doi.org/10.1038/nature.2013.14416)

[Vines et al. \(2013\). doi: 10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014)

Open data. WHERE?

Where to **SUBMIT** data:
Data archives

Where to **RETRIEVE** data:
Data archives + portals

National Data Centres



NMDC
norwegian marine
data centre



NOAA NATIONAL CENTERS FOR
ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Institutional archives



**Meteorologisk
institutt**



Others

NIRD RESEARCH DATA ARCHIVE



BJERKNES CENTRE
for Climate Research



SeaDataNet

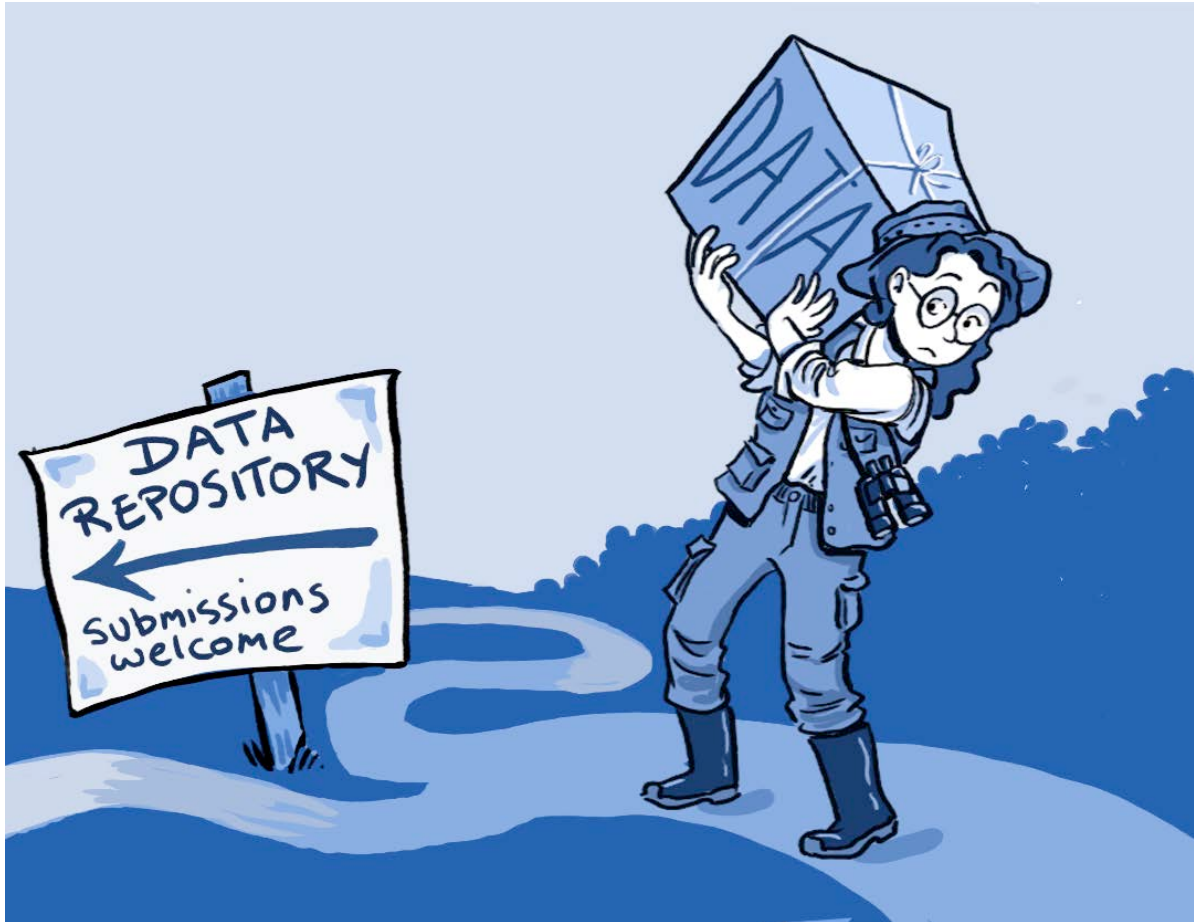


SIOS
SVALBARD INTEGRATED ARCTIC
EARTH OBSERVING SYSTEM



... and **MANY** more

Open data. HOW?

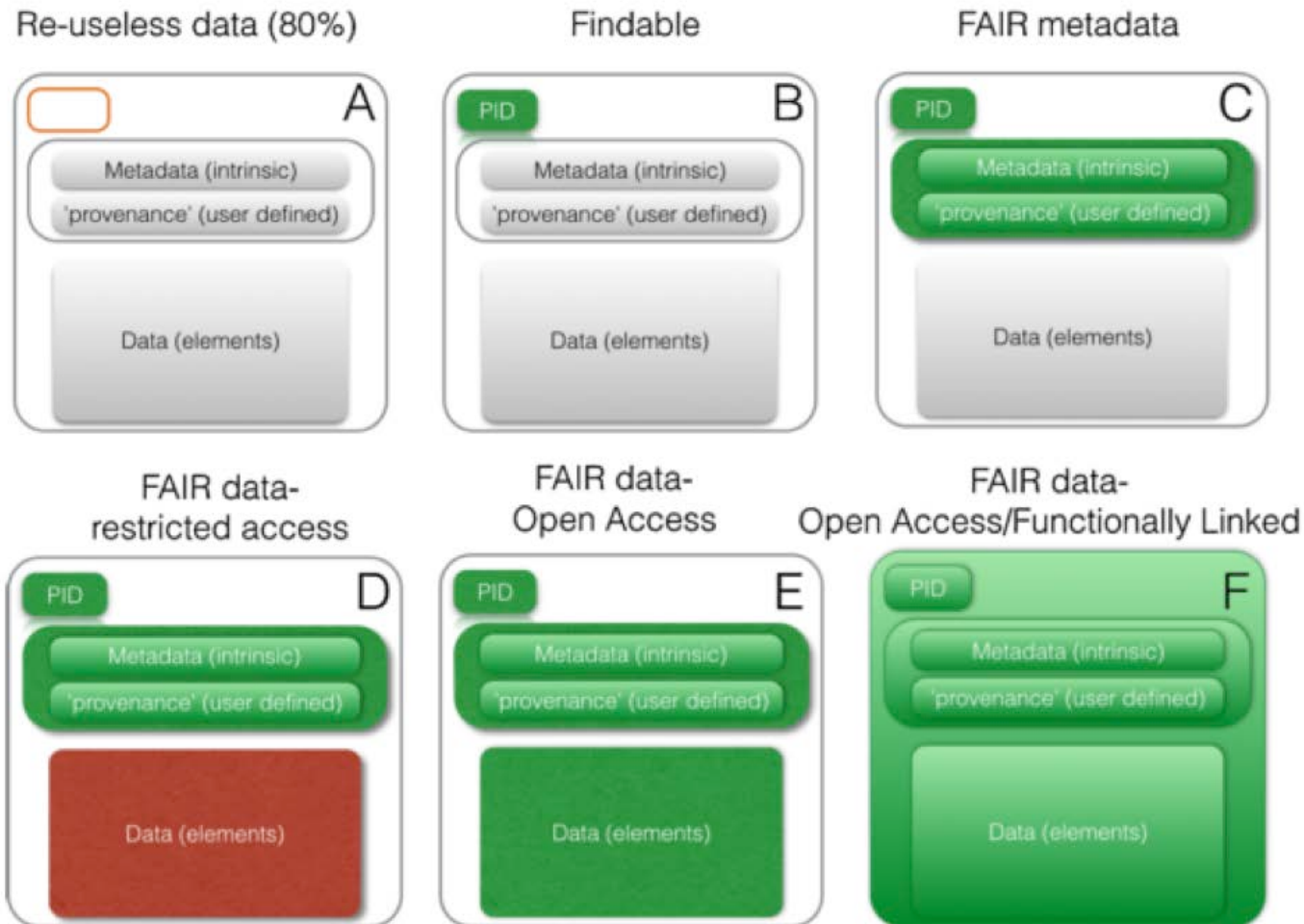


[Ainsley Seago \(2014\). doi: journal.pbio.1001779.g001](https://doi.org/10.1016/j.pbi.2014.07.001)



Open data. HOW?

FAIR principles



[Wilkinson et al. \(2016\). doi: 10.1038/sdata.2016.18](#)

Open data. HOW?

Metadata. Beyond the publication

Discovery metadata

What: Atmospheric Carbon Dioxide Record

Who: R.F. Keeling, S.C. Piper, A.F. Bollenbacher and J.S. Walker. Scripps Institution of Oceanography

Where: 19°32' N, 155°35' W, 3397 m above MSL

When: 1958-2008

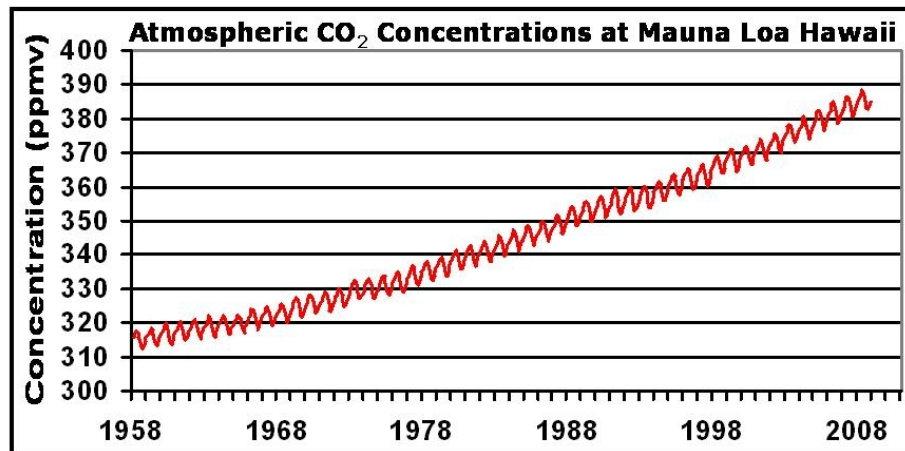
How: Infrared analyzer

Enriched metadata

Data processing: The daily values we report are therefore not simple averages of hourly data, but averages of “baseline data”, i.e. hourly data that have been screened to avoid island effects.

Instrument(s): APC until 2006, followed by Siemens (+1 year overlap). Siemens substituted by CRD in 2012 (+1 year overlap).

Calibration: when, gas concentrations...



PID: DOI: [10.3334/CDIAC/atg.035](https://doi.org/10.3334/CDIAC/atg.035)



Open data. HOW?

Standardization

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS **THE** CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013.II.27. 27½-13 2013.158904109
MMXIII-II-XXVII MMXIII ^{LVII}/_{CCCLXV} 1330300800
 $((3+3) \times ((11+1) - 1) \times 3 / 3 - 1 / 3^3)$ 2013 Mississ
10/11011/1101 02/27/20/13 01²2³3⁴7⁵8 

<https://xkcd.com/1179/>

Cruise	ID	Date	UTC	Time	UTC	Local
08FSH01	11082008	0.549294	11082008			
08FSH01	11082008	0.550949	11082008			
08FSH01	11082008	0.552604	11082008			
08FSH01	11082008	0.554259	11082008			

Location	Date / Time (UTC)	Lat
Harney	22.11.2010 11:37	25.41804
Harney	22.11.2010 11:38	25.41857
Harney	22.11.2010 11:39	25.4188 -8
Harney	22.11.2010 11:39	25.419 -8
Harney	22.11.2010 11:40	25.41965
Harney	22.11.2010 11:41	25.42045

BEN	Year	DOY
	2011.4306	157.6021
	2011.4306	157.6056
	2011.4306	157.6090
	2011.4306	157.6125
	2011.4306	157.6160
	2011.4307	157.6250
	2011.4307	157.6285
	2011.4307	157.6319
	2011.4307	157.6354

GMT	D	M	Y	LAT
23	21	10	57	32
0	22	10	57	32
1	22	10	57	32
2	22	10	57	32

2.0130000e+03	7.3524200e+05	8.0391200e-01
2.0130000e+03	7.3524200e+05	8.0445600e-01
2.0130000e+03	7.3524200e+05	8.0498840e-01
2.0130000e+03	7.3524200e+05	8.0553240e-01
2.0130000e+03	7.3524200e+05	8.0606480e-01
2.0130000e+03	7.3524200e+05	8.0660880e-01
2.0130000e+03	7.3524200e+05	8.0714120e-01
2.0130000e+03	7.3524200e+05	8.0768520e-01

AND:

File format: plain text, NetCDF, HDF...

Standard vocabularies: CF, GCMD, SeaDataNet, Darwin Core...

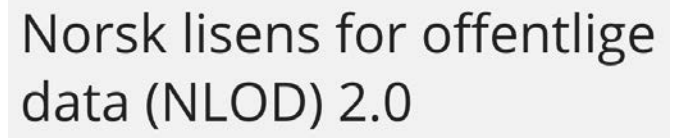
Naming conventions, Versioning

Machine Access: ERDDAP, OPeNDAP

... etc ...



Licenses



*The value of data lies in their use.
Full and open access to scientific data
should be adopted as the
international norm.*

Bits of Power, Issues on Global Access to Scientific Data,
US National Research Council 1997

Thank you!