Open Science

What is it? Do we want it?

(And if so, how do we get there?)

Nanofront Retreat 2019

Gary Steele

What is "open science"?

LOTS of things…

General theme: getting science to be more open and transparent

Sub-theme: change the way we work

Why? Science might work better?

Outline of my talk

You

The world

Open data

Open access

Green and Gold?

How much are you willing to share?

Do we need journals?

Is something else possible?

I want to hear from you!

High Tech (Google Form)   **Low tech**

Write down the questions

(1) Send me an email, or:

(2) Write it on paper and give it to me

https://forms.gle/iCskYtUwCjj1sBVh8

You

Open Data

What is open data? *(green)*

Data files associated with your results are published online

Why? *(red)* Looking at your data might be useful for others!

Theorists trying to model it   Other groups trying to reproduce it   Who knows?

How does it work?

Example from my group:

http://dx.doi.org/10.5281/zenodo.1296129
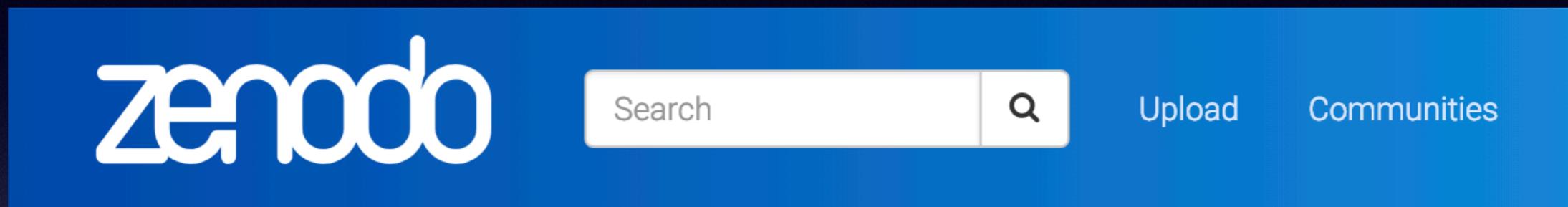
Data availability
All raw and processed data as well as supporting code for processing and figure generation is available in Zenodo with the identifiers https://doi.org/10.5281/zenodo.1296129[43] and https://doi.org/10.5281/zenodo.1408933[44].

43. Schmidt, F. et al. Data and processing for "A ballistic graphene superconducting microwave circuit", Zenodo https://doi.org/10.5281/zenodo.1296129 (2018).
44. Jenkins, M. D. et al. Measurement and Analysis scripts for "A ballistic graphene superconducting microwave circuit", Zenodo https://doi.org/10.5281/zenodo.1408933 (2018).

What does it look like?

June 27, 2018

Data and code for "A ballistic graphene superconducting microwave circuit"

Schmidt, Felix E.; Jenkins, Mark D.; Watanabe, Kenji; Taniguchi, Takashi; Steele, Gary

Data and processing for "A ballistic graphene superconducting microwave circuit"

Requires libraries available at https://github.com/yausern/stlab

Files (1.7 GB)

Name	Size	
gJJ_cavity.tar.gz	1.7 GB	⬇ Download

md5:093a6866364bc60be32f0bce90bf9fd1 ❓

Indexed in

OpenAIRE

Publication date:
June 27, 2018
DOI:

19 downloads…even though it's 1.7 GB!!!

What do I put in it?
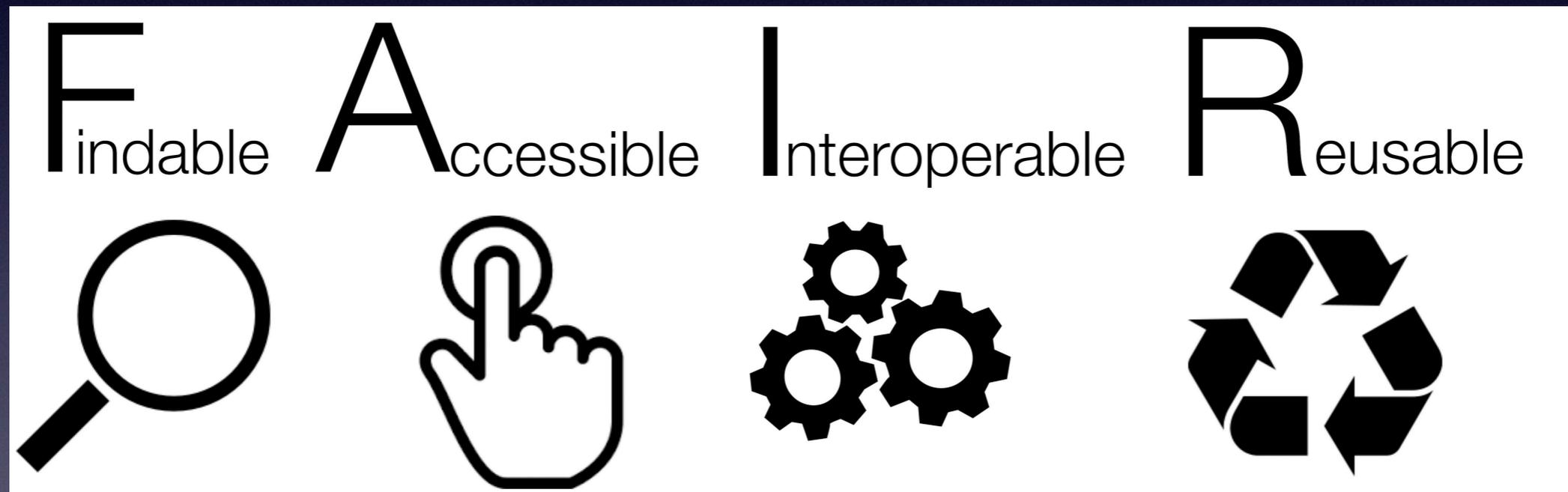
QN Open Data policy

Anton and myself: less philosophy, more concrete

Level 0: Upload the numerical data shown in figures in a format that is readable by others

Level 1: Publication of the raw data and scripts for the full data processing chain that produce the final plots shown in the paper

Approved by QN, now being adopted by TNW

https://doi.org/10.5281/zenodo.2556949

Level 0: Simple

Just the points in your plot
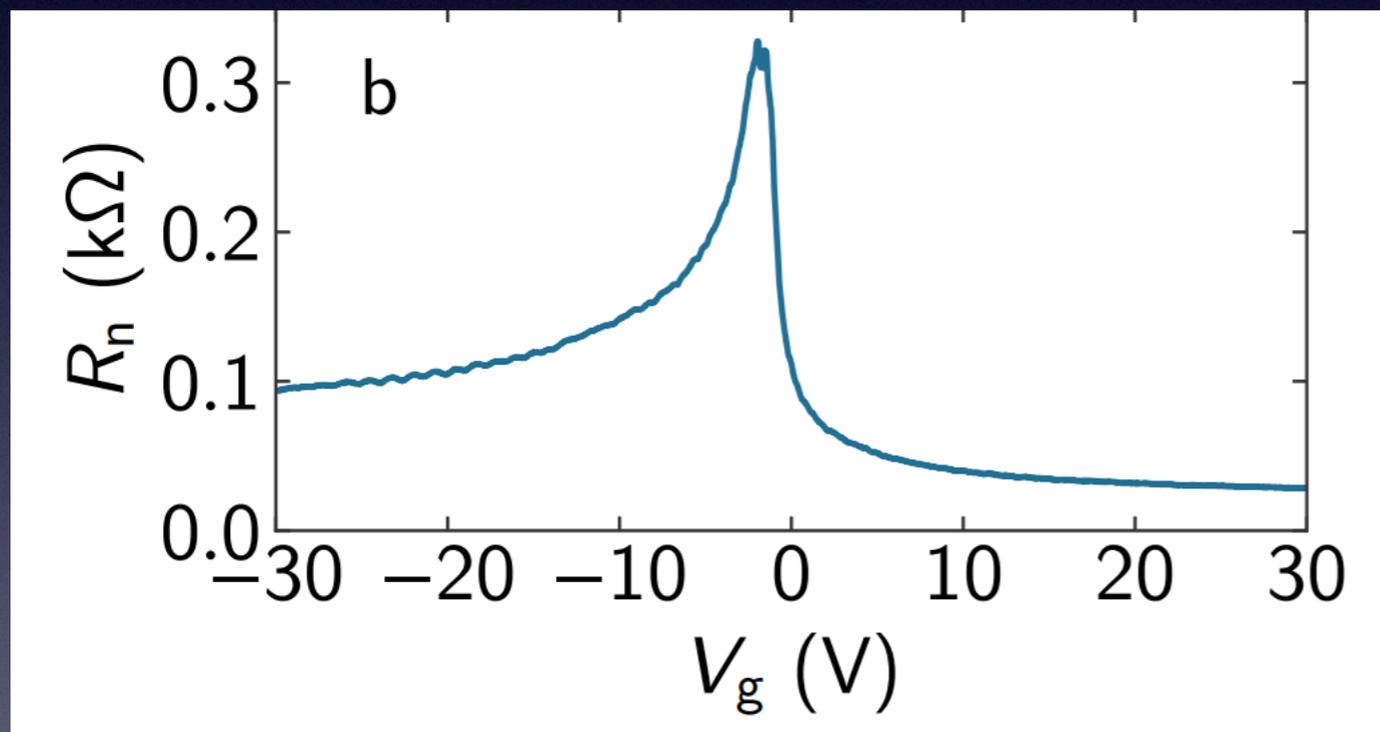
Fig2b.dat

-30 0.1223
-29 0.1133
-28 0.1558
...

What is this useful for?

Other people can analyse you data too

Theorists can fit to it

In principle it's already there

Level 1: A bit more work

Raw data file as recorded during measurement

```
#Frequency (Hz), S21re (), S21im (), S21dB (dB), S21Ph (rad), Po
8.1746000000e+09, -3.8384770000e-02, -2.9692450000e-02, -2.62800
8.1746500000e+09, -3.8401060000e-02, -2.8232300000e-02, -2.64364
8.1747000000e+09, -3.9724860000e-02, -2.6986430000e-02, -2.63707
8.1747500000e+09, -3.9109650000e-02, -2.4865620000e-02, -2.66799
8.1748000000e+09, -3.9544810000e-02, -2.6571220000e-02, -2.64400
8.1748500000e+09, -4.0538970000e-02, -2.6061340000e-02, -2.63402
8.1749000000e+09, -4.2112750000e-02, -2.5065790000e-02, -2.61940
8.1749500000e+09, -4.2064770000e-02, -2.3101750000e-02, -2.63768
8.1750000000e+09, -4.2431660000e-02, -2.3521430000e-02, -2.62824
8.1750500000e+09, -4.2612520000e-02, -2.2390830000e-02, -2.63504
8.1751000000e+09, -4.3512810000e-02, -2.0816920000e-02, -2.63325
8.1751500000e+09, -4.3805420000e-02, -1.9851890000e-02, -2.63582
8.1752000000e+09, -4.3527490000e-02, -1.9960690000e-02, -2.63958
8.1752500000e+09, -4.3660070000e-02, -2.0569940000e-02, -2.63277
8.1753000000e+09, -4.4615900000e-02, -1.7401960000e-02, -2.63951
8.1753500000e+09, -4.5179600000e-02, -1.6710230000e-02, -2.63443
8.1754000000e+09, -4.5898690000e-02, -1.5618080000e-02, -2.62881
8.1754500000e+09, -4.6238390000e-02, -1.5049690000e-02, -2.62626
8.1755000000e+09, -4.5757240000e-02, -1.3835320000e-02, -2.64108
8.1755500000e+09, -4.7169880000e-02, -1.3821450000e-02, -2.61689
8.1756000000e+09, -4.7555410000e-02, -1.1830460000e-02, -2.61952
8.1756500000e+09, -4.7359950000e-02, -1.1975290000e-02, -2.62226
8.1757000000e+09, -4.7504400000e-02, -1.1367820000e-02, -2.62234
8.1757500000e+09, -4.7353720000e-02, -9.9771630000e-03, -2.63042
8.1758000000e+09, -4.7604620000e-02, -9.0821070000e-03, -2.62917
8.1758500000e+09, -4.8968010000e-02, -8.1411660000e-03, -2.60833
8.1759000000e+09, -4.7339280000e-02, -7.1849890000e-03, -2.63966
8.1759500000e+09, -4.7213270000e-02, -5.5391430000e-03, -2.6441
```

Code used to process and plot it

```python
#To concatenate compatible dicts
def join(dict1,dict2):
    result = stlabdict()
    for lab in dict1:
        result[lab] = np.concatenate([dict1[lab],
    return result


Ttrans = {'1K':1.,'300mK':0.3,'15mK':0.015, 'wide
    for each label
##DC DATA
pathlist = glob.glob('../DC_data/*Is_vs_VG.dat')
dataDC = [stlab.readdata.readdat(filename)[0] for
#Add temperature column to each set
for dataset,mystr in zip(dataDC,pathlist):
    for lab in Ttrans:
        if mystr.find(lab) is not -1:
            dataset.addparcolumn('T (K)',Ttrans[l
            break
#Sort sets according to temperature
```

My 5 year prediction

Open data will become the norm

Pressure from peers, universities,
funding agencies, journals

"FAIR" data required for a PhD thesis starting
from 2019 in Delft! (Level 0?)

Better to lead the dialogue and define the standards

Transition from Level 0 to Level 1?
Dialogue about possible next steps

Questions 1-3

Open Data

What type of open data would you be willing to publish?

○ None

○ Level 0

○ Level 1

○ Even more!

Your thoughts: How do you feel about open data? Good idea or bad idea? Do you think it's worth it? How far should we go? Write down what you feel, let your emotions loose!

Your answer

How do you feel about sharing the code you write? Would you be happy to do it? Why / why not?

Your answer

Discuss with your neighbour!

The World

Gold vs. Green

Gold OA	Green OA
Article is published in an Open Access or hybrid journal	Article is self-archived in an online repository (subject or institutional), archive or website after publication
Final published version of work free for anyone to read online immediately after publication	Pre-print, post-print or final published version of work free for anyone to read online; sometimes an embargo period (usually 6-24 months) applies
May include article publishing charges (APCs) paid by the author, or on his/her behalf by institutions or funders (one-time fee)	Publishing expenses paid by subscribers, no additional charges for authors
Less stringent licensing restrictions to maximize access, re-use and dissemination	Flexible licensing options that allow for more author control over his/her work

https://s3.amazonaws.com/libapps/accounts/68137/images/OA_Green_Gold.jpg

Is there something wrong with green?

ArXiv is great! We invented it in 1992!

Some publishers make trouble

Official embargo (often ignored)

Sometimes people forget, or are lazy…

(Making this presentation, I discovered that 11 of my 76 papers are not on the arxiv!)

Most fields of science have no arxiv / no arxiv culture!

Question 4

Open access

What percentage of your papers are on the arxiv?

○ 100%

○ Probably 90%

○ Less than 50%

○ None

BACK   SUBMIT

Green open access

It is a way to make your papers open!
(if your journal allows it)

It requires discipline (or a big EU stick…)

If your field has a culture of it, it really works!

My only source of information about new
papers is the arxiv!

So what's the problem?

19 billion euro

(warning: possibly polarising discussion coming up…)

Academic publishing is an *extremely* profitable business

total global revenues of more than £19bn, it weighs in somewhere between the recording and the film industries in size, but it is far more profitable. In 2010, Elsevier's scientific publishing arm reported profits of £724m on just over £2bn in revenue. It was a 36% margin - higher than Apple, Google, or Amazon posted that year.

Somewhere, there is a missing feedback loop...

https://www.universityaffairs.ca/opinion/in-my-opinion/time-stand-academic-publishing-industry/

https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

https://theconversation.com/could-subscriptions-for-academic-journals-go-the-way-of-pay-phones-63575

It's big business

operating profit	company	industry
7%	Woolworths	supermarkets, poker
12%	BMW	automobiles
22%	Coca-Cola	adding sugar to water
23%	Rio Tinto	mining
36%	Apple	premium computing
34%	Springer	scholarly publishing
36%	Elsevier	scholarly publishing
42%	Wiley	scholarly publishing

http://tagg.org/rants/AcPubPizzo.html

Could we get rid of journals?

Could we just go "cold turkey"?
What would we miss?

Question 5

Purpose of Journals

In the modern era, with the arxiv and google, what do we need journals for?

Your answer

Discuss with your neighbour!

What purpose do journals serve?

Idea for nut pap exp → PRL IF=50   IF=2

You need a system that includes a peer review that is recognized by the international community

recognition by the community — reason why people
at one point thought
unisons agreed idea
blind peer review

peer review — system understands

modern needs

Older days needs:

distribute & print (completely irrelevant for 20 years)

→ original purpose
letters publish after 1870?

why?

everyone reads it

Select for me what I should read (topic)

in physics not necessarily you

this needs to develop as culture in other disciplines

This can easily be replaced.

In physics, it already has been by our own community

(arxiv)

(1992)

all physicists put their papers on the arxiv

pre-select for me what I should consider "high quality work"

Nature paper!

Gary thinks! this is probably a pretty good paper

Why do I think this?

Because my experience is that it is often true

I already read arxiv every day looking at papers,

I already make a decision

I don't actually read the paper

system I trust or points = my subfield → could twitter

a measure for assessing the quality of scientists

→ PDF your submit
→ proposal new

→ not 100%
h factor
→ not 100%
nut pap

(clonky methods and communicates)

major advance that is new?

Because? go to recruit

→ does nature know how to do this?

95 5
50 50
5% → 95%

What do journals do?

Let me read papers

Arxiv can do this 100%

Verify (to a degree)
that results are
scientifically valid
(part of peer review)

Arxiv does not do
this at all

Inform me

Arxiv can do this 100%

What do journals do?

Let me read papers

Arxiv can do this 100%

Verify (to a degree) that results are scientifically valid (part of peer review)

Arxiv does not do this at all

Inform me

Arxiv can do this 100%

Assess the (perceived) importance of a scientific work (editor and peer review)

Arxiv does not do this at all

Scientific validity

Assess Scientific Validity

This one is not hard (and we do it already anyway!)

We would need a "coordinator"…

Assess Scientific Validity

This one is not hard (and we do it already anyway!)

We would need a "coordinator"…

Sci|Post

ArXiv overlay journal that coordinates peer review

My opinion: SciPost is cool! But not enough…

"Importance"

Question 6

Nature / Science / Cell

Do you want to publish in high impact journals like Nature / Science / Cell / Baby Nature?

◯ Yes

◯ No

I do! Why?

One reason: it is my currency!

Next grant   (A job!)

Big paper
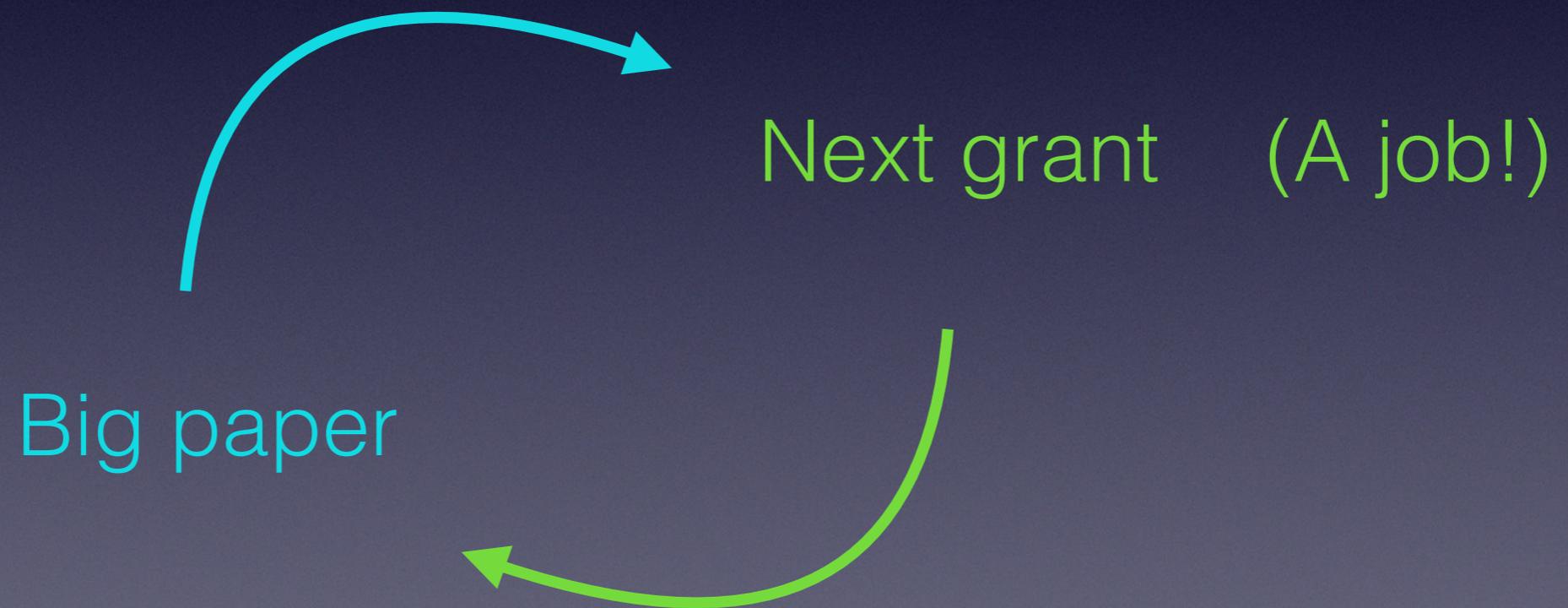
Why else?

Why else?

It feels great!

Why else?

Why?

It feels great!

Why else?

Why?

It feels great!

Recognition

You've done a great job!

And it comes (at least partly) from your peers that you respect!

Recognition

We need it!   We crave it!   It motivates us

Striving for excellence is not a bad thing!

And we are impatient people. We need to know!

Not 5 years from now, we need recognition and some form of closure now!

How do we get recognition?

The Moriond Plan? (Plan-M?)

(Freeform-after-dinner-brainstorming in hallway with anonymous colleague on Monday…)

Who do we want recognition from?

Question 7

Recognition

Who do you want recognition from?

Your answer

How do we get recognition?

The Moriond Plan? (Plan-M?)

(Freeform-after-dinner-brainstorming in hallway with
anonymous colleague on Monday...)

Who do we want recognition from?

Our answer was: Our peers

We want to know that a our peers find our work
exciting and important

Could we set this up?

A panel of scientists from the field that make a selection of the "HOT" arxiv postings each month?

Whatever we do, key concepts for me:

Trust, transparency, recognition

Through transparency and effort from people in the field, could such a system establish trust and recognition?

The future of science?

arXiv.org

Content

Sci|Post

Peer review of validity

The HOT list?*

Recognition: something to strive for!

*(or some kind of "official" panel? Or panels?)

Thanks for your attention!

Let's think together about a bright future for science

(add your thoughts in the last section of the form!)

Open success story

We gave away all of our secrets!
30 pages of pictures of exactly how to peel stuff
Thorlabs parts list!

Result: 560 citations in 5 years!
From a simple, non-sexy technical paper!

Plan-S

Plan devised by former European Commissioner

Short summary:

Force gold open access with a 2000 euro APC
Ban publication in non-gold OA

My personal analysis

Best case: Kill the journals

Best case risk: Sacrifice a generation or two of scientists while the system adjusts

Worst case: Create a new species of profitable predatory journals with poor quality control

Worst case risk: Replace a monster with an even worse monster?

https://sites.google.com/view/plansopenletter

Eight routes towards Plan-S compliance

routes to Plan S compliance	A. existing/new APC gold journal / platform	B. existing/new non-APC gold journal / platform (diamond)	C. flipping journals to APC gold (by publishers or editors)	D. soft-flipping journals to APC gold (leaving hybrid intact)	E. flipping journals to non-APC gold (diamond), by publishers or eds.	F. archiving publisher version, on publication	G. archiving AAM, on publication	H. sharing preprints and using overlay PR
1. compliant?	yes	yes	yes	yes	yes	yes	unsure	yes
2. example	PLoS, 1000s more	Open Library of Humanities, 1000s more	Epidemiology & Infection, 100s more	Water Research X, J. of Hydrology X	Scoap[3]	MNRAS, APS journals	Emerald journals	SciPost
3. current use / availability	sizeable amount	limited	very limited	rare	very limited	limited	limited	very limited
4. effect on publishers	gold publ. win, evt. decreasing subscriptions	more competition / perhaps evt. decreasing subscriptions	change in business model / probl. for high rejection	matching APC levels, evt. decreasing subscriptions?	new partnerships or loose journals to funders/institutions	evt. decreasing subscriptions, need to solve sustainability?	keep large part of perceived value	change publishing model or loose out
5 effect on researchers	away from trad. venues and IF-thinking	away from trad. venues and IF-thinking	depends on (funding for) APC	accept new journal as similar to sister	none	small effort	small effort, accept limitations	adapt to new idea
6. effect on libraries	away from hybrid deals & IF-thinking	away from hybrid deals & IF-thinking, pot. role in funding	limited	limited	potential role in funding	role insofar as hosted in IR / cancel subs evt.	continued role, esp. hosting in inst. repo	chance to play role in curation
7. effect on funders	supporting (own) platforms / lower APC levels	supporting (own) platforms / lower APC levels?	depends on APC levels	depends on APC levels	lower average APC levels? / pot. role in funding	no financial gain	no financial gain	adapt to new idea, change assessment
8. effect on societies	big, because of subscription dependance	big, because of subscription dependance	change in business model / probl. for high rejection	matching APC levels / income?	change in business model / probl. for high rejection	evt. decreasing subscriptions?	evt. decreasing subscriptions?	limited role, perhaps in quality assurance?
9. effect on editors of trad. jrnls.	fewer submissions, lower status	fewer submissions, lower status	none (or big role in leading flip)	none	none (or big role in leading flip)	none	none	new role in overlay journals?
10. overall pub cost	depends on cap / market	depends on cap / market / funding sources	depends on cap / market	depends on cap / market	depends on cap / market / funding sources	remains high	remains high	substantially lower?
11. fits changes in assessment	+	+	-	- (theoretically)	-	-	-	++

https://101innovations.wordpress.com/2018/10/22/eight-routes-towards-plan-s-compliance/

Open Source

FOSS: Free and open-source software

Free to use   Source code for free

Idea: you share your code, others share their code

Open Hardware?
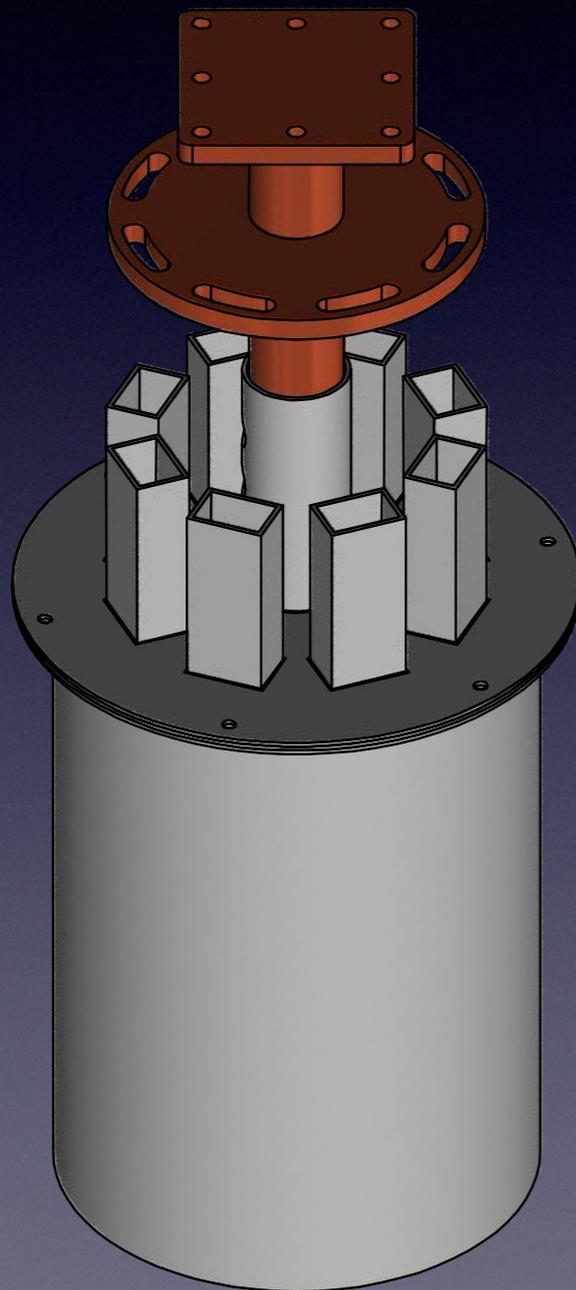
We just spent 4 months designing multilayer quit shielding

Why should someone else have to do this again?

Solution: publish the design files online! We can and will

Level 1: is it scary?

A lot of work…but if you do it properly from the start, it will make your work flow more efficient and more robust!

My code works but is not pretty…I'm don't want people to see it… but it means that people (including your advisor) can know exactly what you did! (And maybe it motivates you to tidy it up…)

I'm giving it away all my hard work for free! But assuming others are not evil, they will appreciate it and maybe cite it. And eventually, it will become the norm (by social pressure or by funding agency police…) and you will benefit from their code too.