FAIR Data Stewardship

Erik Schultes, Barbara Magagna, Alessa Gambardella

5 December 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell’Università e della Ricerca
Italiadomani

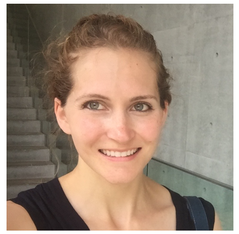
Welcome - Recap

13:00-13:10
(Erik & Barbara)

Who we are…

Erik Schultes
FAIR Implementation Lead
GO FAIR Foundation | Senior Researcher
Leiden Academic Center for Drug Research

Barbara Magagna
Lead FAIR Development
GO FAIR Foundation | Twente University

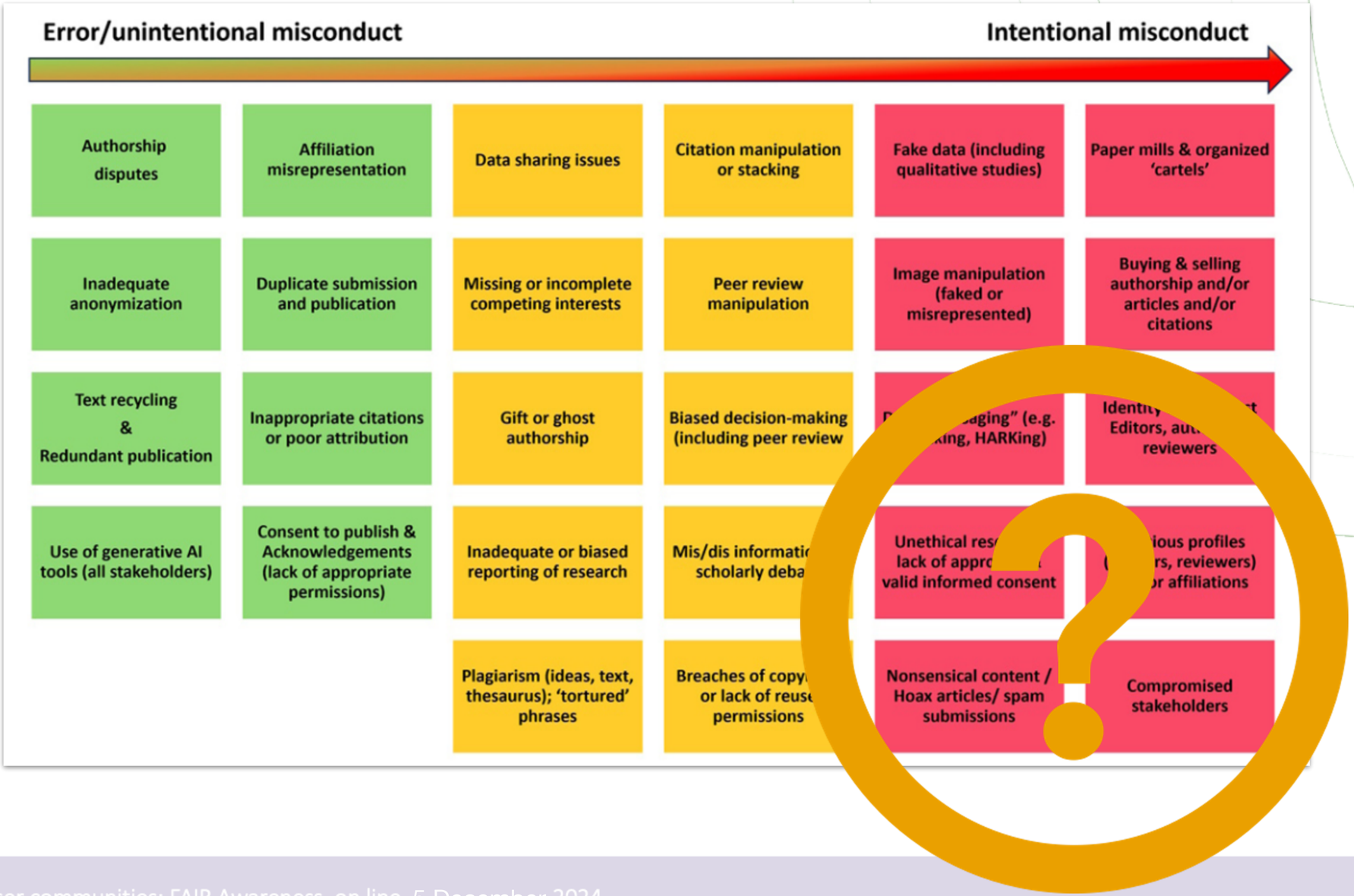


Alessa Gambardella
Data Steward
Faculty of Science, Leiden University | GO FAIR Foundation Fellow, 3PFF Facilitator Assistant

Yesterday…

Yesterday…

Research Integrity



Agenda

13:00-13.10 (10 minutes)	**Welcome (Erik)**
13:10-13:20 (10 minutes)	Discuss homework (Erik)
13:20-13:40 (20 minutes)	Eriks personal data management story (Erik)
13:40-14:00 (20 minutes)	Data Stewardship & Research Objects (Erik)
14.00-14:45 (45 minutes)	Data life cycle & Research objects (interactive)
14:45-15:00 (15 minutes)	**Break**
15:00-15:20 (20 minutes)	Data Management Planning (Alessa)
15:20-15:40 (20 minutes)	Machine Actionable DMPs (Barbara)
15:40-15:50 (10 minutes)	Q&A
15:50-16:00 (10 minutes)	**Break**
16:00-16:45 (45 minutes)	Try out services and discuss (breakout)
16:45-16:55 (10 minutes)	Report from breakout
16:55-17:00 (5 minutes)	Wrap up
17:00	**Adjourn**

Common Notes: https://bit.ly/FA-5-T-1_ITINERIS

Erik’s Data Stewardship Catastrophe

13:20-13:40
(Erik)

UCLA 1992

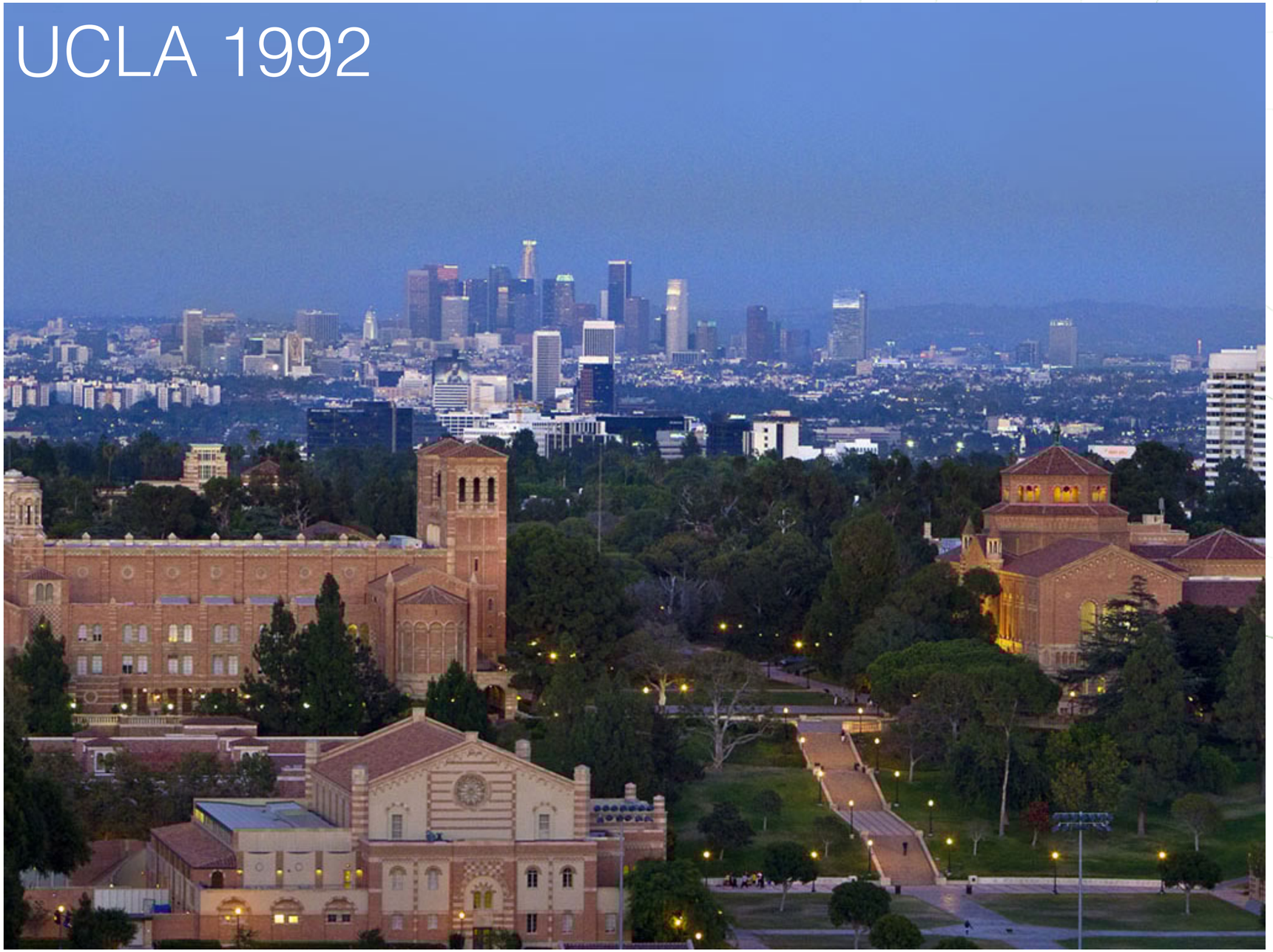


TABLE 1. Phylogenetically and functionally representative compilation of ssRNA nucleotide composition.

RNA	Taxon	n^a	$\langle N \rangle^b$	$\langle G+C \rangle^c$	$\langle G+A \rangle^c$	$\langle G+U \rangle^c$
Comprehensive	—	2,800	287.0 ± 662.0	0.509 ± 0.200	0.516 ± 0.0310	0.523 ± 0.0462
23S rRNA	Archaea	15	2,968.9 ± 65.8	0.588 ± 0.0525	0.567 ± 0.00597	0.509 ± 0.0212
	Bacteria	39	2,915.6 ± 86.3	0.526 ± 0.0383	0.570 ± 0.00720	0.517 ± 0.0102
	Eucarya	33	3,615.0 ± 470.3	0.530 ± 0.0832	0.540 ± 0.0139	0.520 ± 0.0143
18S rRNA	Metazoa	20	1,821.0 ± 43.2	0.494 ± 0.0297	0.517 ± 0.0123	0.535 ± 0.0113
16S rRNA	Archaea	15	1,530.7 ± 184.9	0.611 ± 0.0427	0.562 ± 0.00500	0.507 ± 0.00681
	Bacteria	85	1,511.8 ± 30.9	0.550 ± 0.0387	0.568 ± 0.00770	0.520 ± 0.0118
	Eucarya	47	1,823.6 ± 57.8	0.486 ± 0.0255	0.524 ± 0.00671	0.527 ± 0.00955
5S rRNA	Archaea	26	124.5 ± 3.9	0.598 ± 0.0726	0.508 ± 0.0272	0.498 ± 0.0409
	Bacteria	123	117.6 ± 3.9	0.575 ± 0.0574	0.520 ± 0.0308	0.495 ± 0.0344
	Eucarya	234	119.3 ± 1.4	0.557 ± 0.0369	0.517 ± 0.0119	0.505 ± 0.0176
P RNA	Archaea	7	400.9 ± 62.0	0.644 ± 0.0966	0.564 ± 0.0140	0.469 ± 0.0296
	Bacteria	37	389.0 ± 39.7	0.569 ± 0.114	0.570 ± 0.0168	0.496 ± 0.0147
Group I ribozymes	—	13	757.15 ± 517.5	0.434 ± 0.105	0.561 ± 0.0302	0.492 ± 0.0268
Group II ribozymes	—	6	734.8 ± 853.8	0.355 ± 0.0933	0.566 ± 0.0259	0.495 ± 0.0276
HDV Ribozymes	—	2	43.5 ± 43.1	0.619 ± 0.00212	0.512 ± 0.0368	0.519 ± 0.0269
snRNA, U1	Eucarya	24	162.4 ± 4.2	0.557 ± 0.0228	0.487 ± 0.00957	0.548 ± 0.0152
snRNA, U2	Eucarya	16	187.8 ± 11.6	0.455 ± 0.0357	0.456 ± 0.0310	0.543 ± 0.0257
snRNA, U3	Eucarya	8	220.6 ± 14.6	0.468 ± 0.0684	0.477 ± 0.0417	0.561 ± 0.0207
snRNA, U4	Eucarya	11	143.6 ± 12.3	0.484 ± 0.0384	0.493 ± 0.0192	0.536 ± 0.0303
snRNA, U5	Eucarya	11	125.6 ± 29.7	0.411 ± 0.0345	0.448 ± 0.0279	0.531 ± 0.0220
snRNA, U6	Eucarya	14	101.5 ± 7.7	0.451 ± 0.0233	0.544 ± 0.0304	0.469 ± 0.0352
tRNA	Comprehensive	2,011	74.3 ± 6.3	0.496 ± 0.120	0.510 ± 0.0280	0.528 ± 0.0509
	Archaea	121	77.0 ± 4.2	0.633 ± 0.0457	0.503 ± 0.0215	0.516 ± 0.0351
	Bacteria	371	78.3 ± 5.2	0.580 ± 0.0522	0.506 ± 0.0222	0.522 ± 0.0317
	Eucarya	436	75.8 ± 4.2	0.572 ± 0.0438	0.510 ± 0.0282	0.544 ± 0.0365
	Chloroplast	291	75.6 ± 5.0	0.526 ± 0.0531	0.510 ± 0.0212	0.540 ± 0.0341
	Mitochondria	742	70.3 ± 6.4	0.372 ± 0.0939	0.514 ± 0.0328	0.519 ± 0.0685
Artificial ribozymes[d]	Class I					
	b1	1	274	0.471	0.511	0.493
	b1-207	1	119	0.513	0.546	0.445
	Class II					
	c2	1	271	0.568	0.535	0.546
	d1	1	273	0.487	0.502	0.524
	f1	1	273	0.546	0.524	0.502
	Class III					
	e3	1	272	0.529	0.511	0.563
	g1	1	274	0.544	0.522	0.518
Random	—	500	25	0.498 ± 0.104	0.495 ± 0.0979	0.494 ± 0.130
	—	500	74	0.500 ± 0.0560	0.502 ± 0.0587	0.498 ± 0.0548
	—	500	120	0.500 ± 0.0471	0.501 ± 0.0457	0.500 ± 0.0446
	—	500	400	0.501 ± 0.0260	0.501 ± 0.0247	0.500 ± 0.0254
	—	500	1,500	0.500 ± 0.0134	0.500 ± 0.0132	0.500 ± 0.0129
	—	500	3,000	0.500 ± 0.00924	0.500 ± 0.00954	0.500 ± 0.00888
	—	500	5,000	0.500 ± 0.00692	0.500 ± 0.00758	0.500 ± 0.00703

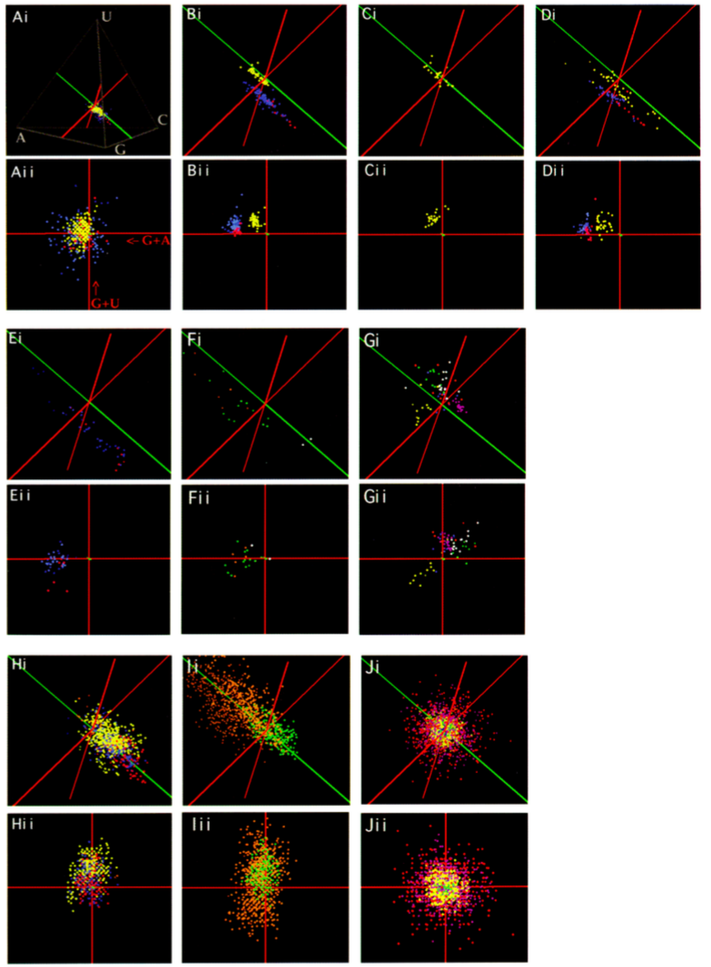


FIGURE 2. (*Legend on facing page.*)

Erik’s PhD Data Stewardship Plan 1997

3.5 inch floppy disks, 1.4 MB per disk

MIT 1997

minus the background levels observed in the HSP in the control (Sar1-GDP–containing) incubation that prevents COPII vesicle formation. In the microsome control, the level of p115-SNARE associations was less than 0.1%.
46. C. M. Carr, E. Grote, M. Munson, F. M. Hughson, P. J. Novick, *J. Cell Biol.* **146**, 333 (1999).
47. C. Ungermann, B. J. Nichols, H. R. Pelham, W. Wickner, *J. Cell Biol.* **140**, 61 (1998).
48. E. Grote and P. J. Novick, *Mol. Biol. Cell* **10**, 4149 (1999).
49. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
50. GST-SNARE proteins were expressed in bacteria and purified on glutathione-Sepharose beads using standard methods. Immobilized GST-SNARE protein (0.5 μM) was incubated with rat liver cytosol (20 mg) or purified recombinant p115 (0.5 μM) in 1 ml of NS buffer containing 1% BSA for 2 hours at 4°C with rotation. Beads were briefly spun (3000 rpm for 10 s) and sequentially washed three times with NS buffer and three times with NS buffer supplemented with 150 mM NaCl. Bound proteins were eluted three times in 50 μl of 50 mM tris-HCl (pH 8.5), 50 mM reduced glutathione, 150 mM NaCl, and 0.1% Triton X-100 for 15 min at 4°C with intermittent mixing, and elutes were pooled. Proteins were precipitated by MeOH/CH_3Cl and separated by SDS–polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting using p115 mAb 13F12.
51. V. Rybin *et al.*, *Nature* **383**, 266 (1996).
52. K. G. Hardwick and H. R. Pelham, *J. Cell Biol.* **119**, 513 (1992).
53. A. P. Newman, M. E. Groesch, S. Ferro-Novick, *EMBO J.* **11**, 3609 (1992).
54. A. Spang and R. Schekman, *J. Cell Biol.* **143**, 589 (1998).
55. M. F. Rexach, M. Latterich, R. W. Schekman, *J. Cell Biol.* **126**, 1133 (1994).
56. A. Mayer and W. Wickner, *J. Cell Biol.* **136**, 307 (1997).
57. M. D. Turner, H. Plutner, W. E. Balch, *J. Biol. Chem.* **272**, 13479 (1997).
58. A. Price, D. Seals, W. Wickner, C. Ungermann, *J. Cell Biol.* **148**, 1231 (2000).
59. X. Cao and C. Barlowe, *J. Cell Biol.* **149**, 55 (2000).
60. G. G. Tall, H. Hama, D. B. DeWald, B. F. Horazdovsky, *Mol. Biol. Cell* **10**, 1873 (1999).
61. C. G. Burd, M. Peterson, C. R. Cowles, S. D. Emr, *Mol. Biol. Cell* **8**, 1089 (1997).
62. M. R. Peterson, C. G. Burd, S. D. Emr, *Curr. Biol.* **9**, 159 (1999).
63. M. G. Waters, D. O. Clary, J. E. Rothman, *J. Cell Biol.* **118**, 1015 (1992).
64. D. M. Walter, K. S. Paul, M. G. Waters, *J. Biol. Chem.* **273**, 29565 (1998).
65. N. Hui *et al.*, *Mol. Biol. Cell* **8**, 1777 (1997).
66. T. E. Kreis, *EMBO J.* **5**, 931 (1986).
67. H. Plutner, H. W. Davidson, J. Saraste, W. E. Balch, *J. Cell Biol.* **119**, 1097 (1992).
68. D. S. Nelson *et al.*, *J. Cell Biol.* **143**, 319 (1998).
69. We thank G. Waters for p115 cDNA and p115 mAbs; G. Warren for p97 and p47 antibodies; R. Scheller for rbet1, membrin, and sec22 cDNAs; H. Plutner for excellent technical assistance; and P. Tan for help during the initial phase of this work. Supported by NIH grants GM 33301 and GM42336 and National Cancer Institute grant CA58689 (W.E.B.), a NIH National Research Service Award (B.D.M.), and a Wellcome Trust International Traveling Fellowship (B.B.A.).

20 March 2000; accepted 22 May 2000

One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds

Erik A. Schultes and David P. Bartel*

We describe a single RNA sequence that can assume either of two ribozyme folds and catalyze the two respective reactions. The two ribozyme folds share no evolutionary history and are completely different, with no base pairs (and probably no hydrogen bonds) in common. Minor variants of this sequence are highly active for one or the other reaction, and can be accessed from prototype ribozymes through a series of neutral mutations. Thus, in the course of evolution, new RNA folds could arise from preexisting folds, without the need to carry inactive intermediate sequences. This raises the possibility that biological RNAs having no structural or functional similarity might share a common ancestry. Furthermore, functional and structural divergence might, in some cases, precede rather than follow gene duplication.

Related protein or RNA sequences with the same folded conformation can often perform very different biochemical functions, indicating that new biochemical functions can arise from preexisting folds. But what evolutionary mechanisms give rise to sequences with new macromolecular folds? When considering the origin of new folds, it is useful to picture, among all sequence possibilities, the distribution of sequences with a particular fold and function. This distribution can range very far in sequence space (*1*). For example, only seven nucleotides are strictly conserved among the group I self-splicing introns, yet secondary (and presumably tertiary) structure within the core of the ribozyme is preserved (*2*). Because these disparate isolates have the same fold and function, it is thought that they descended from a common ancestor through a series of mutational variants that were each functional. Hence, sequence heterogeneity among divergent isolates implies the existence of paths through sequence space that have allowed neutral drift from the ancestral sequence to each isolate. The set of all possible neutral paths composes a "neutral network," connecting in sequence space those widely dispersed sequences sharing a particular fold and activity, such that any sequence on the network can potentially access very distant sequences by neutral mutations (*3*–*5*).

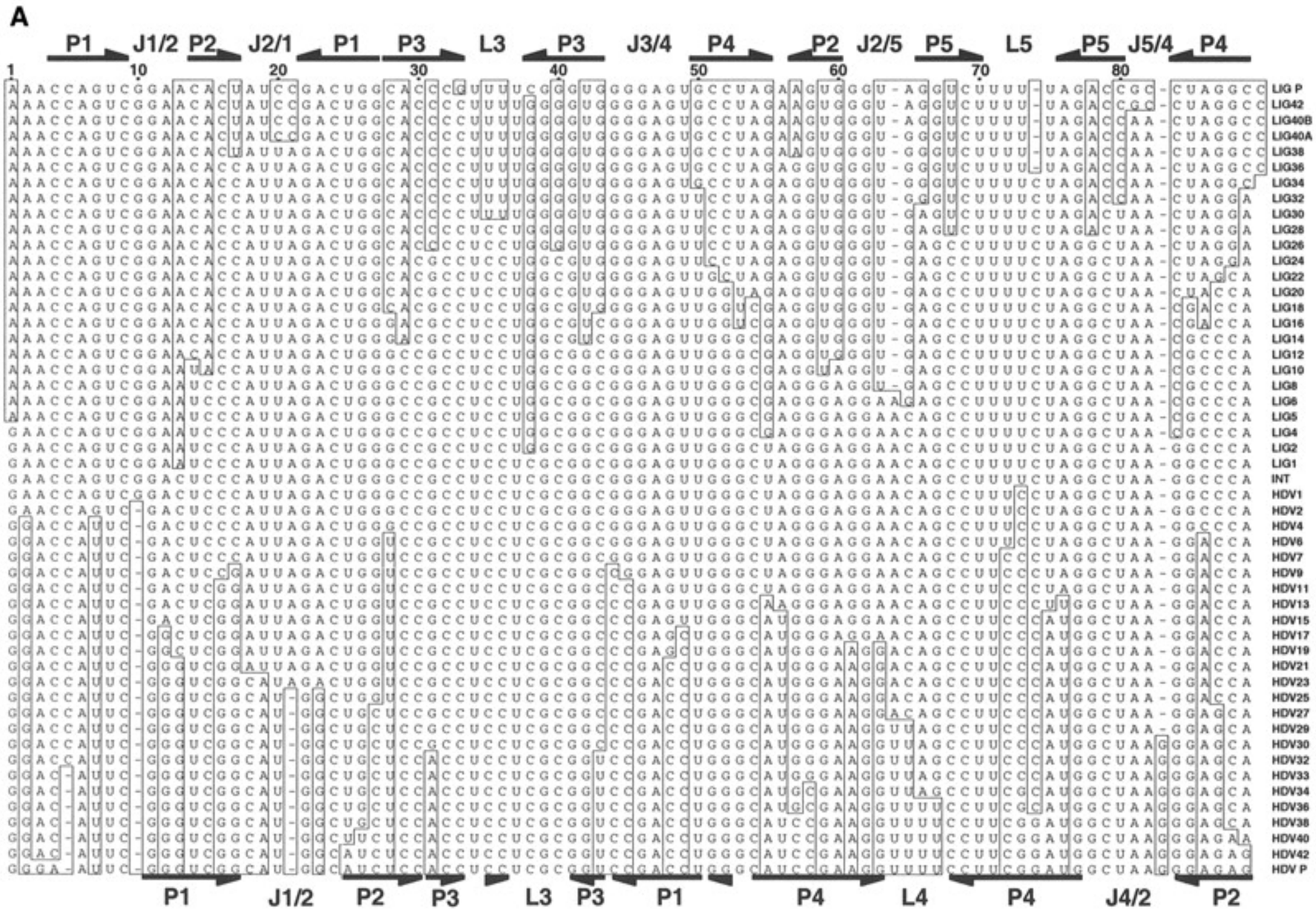
Theoretical analyses using algorithms for predicting RNA secondary structure have suggested that different neutral networks are interwoven and can approach each other very closely (*3*, *5*–*8*). Of particular interest is whether ribozyme neutral networks approach each other so closely that they intersect. If so, a single sequence would be capable of folding into two different conformations, would have two different catalytic activities, and could access by neutral drift every sequence on both networks. With intersecting networks, RNAs with novel structures and activities could arise from previously existing ribozymes, without the need to carry nonfunctional sequences as evolutionary intermediates. Here, we explore the proximity of neutral networks experimentally, at the level of RNA function. We describe a close apposition of the neutral networks for the hepatitis delta virus (HDV) self-cleaving ribozyme and the class III self-ligating ribozyme.

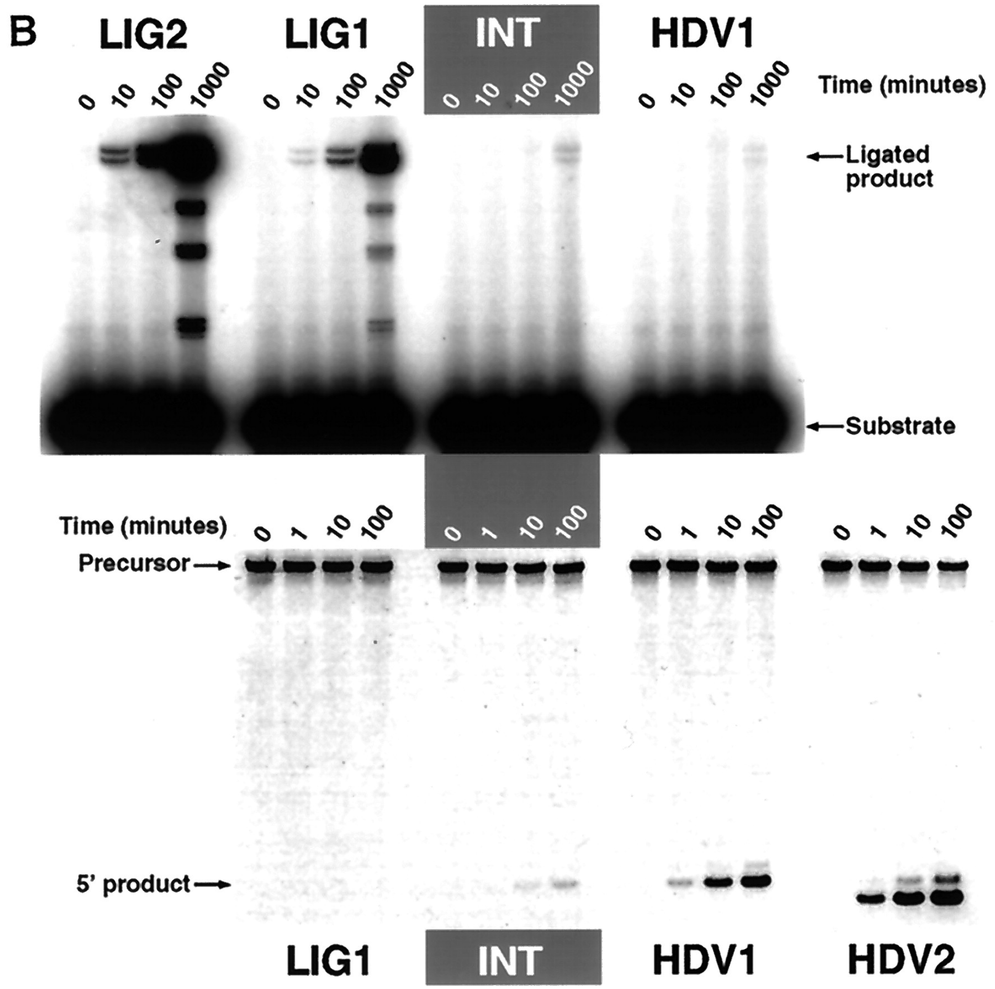
In choosing the two ribozymes for this investigation, an important criterion was that they share no evolutionary history that might confound the evolutionary interpretations of our results. Choosing at least one artificial ribozyme ensured independent evolutionary histories. The class III ligase is a synthetic ribozyme isolated previously from a pool of random RNA sequences (*9*). It joins an oligonucleotide substrate to its 5′ terminus. The prototype ligase sequence (Fig. 1A) is a shortened version of the most active class III variant isolated after 10 cycles of in vitro selection and evolution. This minimal construct retains the activity of the full-length isolate (*10*). The HDV ribozyme carries out the site-specific self-cleavage reactions needed during the life cycle of HDV, a satellite virus of hepatitis B with a circular, single-stranded RNA genome (*11*). The prototype HDV construct for our study (Fig. 1B) is a shortened version of the antigenomic HDV ribozyme (*12*), which undergoes self-cleavage at a rate similar to that reported for other antigenomic constructs (*13*, *14*).

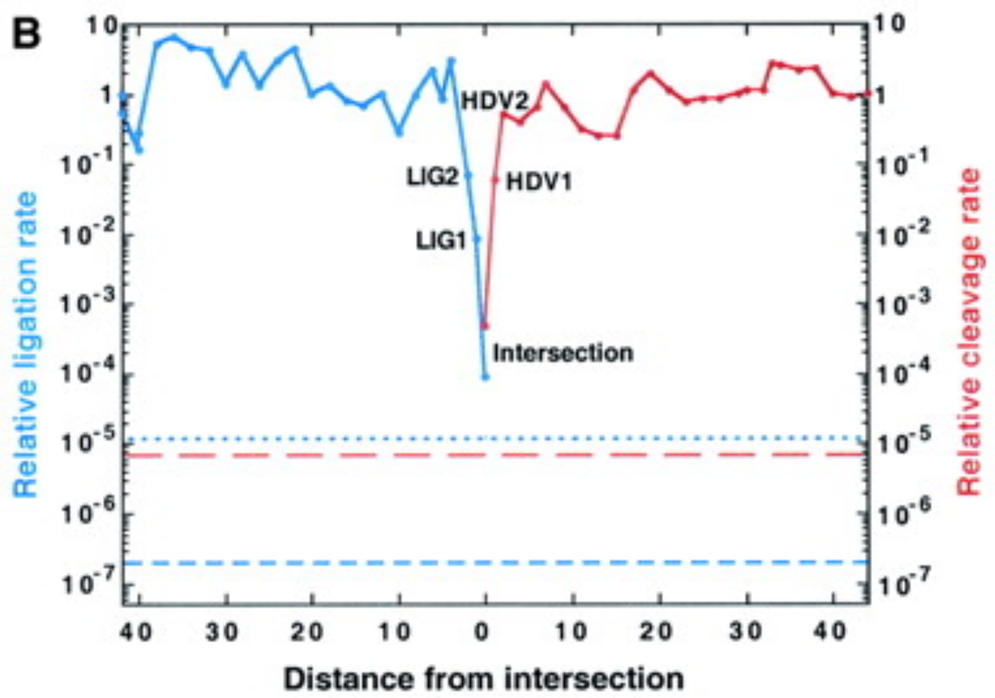
The prototype class III and HDV ribozymes have no more than the 25% sequence identity expected by chance and no fortuitous structural similarities that might favor an intersection of their two neutral networks. Nevertheless, sequences can be designed that simultaneously satisfy the base-pairing requirements

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu


A
P1 J1/2 P2 J2/1 P1 P3 L3 P3 J3/4 P4 P2 J2/5 P5 L5 P5 J5/4 P4
1 10 20 30 40 50 60 70 80
LIG P
LIG42
LIG40B
LIG40A
LIG38
LIG36
LIG34
LIG32
LIG30
LIG28
LIG26
LIG24
LIG22
LIG20
LIG18
LIG16
LIG14
LIG12
LIG10
LIG8
LIG6
LIG5
LIG4
LIG2
LIG1
INT
HDV1
HDV2
HDV4
HDV6
HDV7
HDV9
HDV11
HDV13
HDV15
HDV17
HDV19
HDV21
HDV23
HDV25
HDV27
HDV29
HDV30
HDV32
HDV33
HDV34
HDV36
HDV38
HDV40
HDV42
HDV P
P1 J1/2 P2 P3 L3 P3 P1 P4 L4 P4 J4/2 P2


B
LIG2
LIG1
INT
HDV1
0 10 100 1000
Time (minutes)
Ligated product
Substrate
Time (minutes)
0 1 10 100
Precursor
5' product
LIG1
INT
HDV1
HDV2


B
HDV2
LIG2
HDV1
LIG1
Intersection
Relative ligation rate
Relative cleavage rate
Distance from intersection

Erik’s Postdoc Data Stewardship Plan 2004

Optical disk, 100 MB per disk

three, at which point the pool was cloned and sequenced. Each of the 20 isolates had a different nucleotide sequence, indicating that there are many highly active sequences in the vicinity of the intersection sequence. Sequences of active clones informed the design of LIG1, LIG2, and LIG4.
24. M. A. Tanner and T. R. Cech, *RNA* **2**, 74 (1996).
25. Supplemental data showing the predicted secondary structures of each construct (Fig. 3) and explaining the ligation activity of truncated ribozymes (Fig. 2B) are available at *Science* Online at www.sciencemag.org/feature/data/1050240.shl.
26. J. Maynard Smith, *Nature* **225**, 563 (1970).
27. The Intersection Theorem (*5*) states that there exists at least one intersection sequence for every pair of RNA secondary structures (usually as suboptimal conformations).
28. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
29. D. L. Minor Jr. and P. S. Kim, *Nature* **380**, 730 (1996).
30. M. H. J. Cordes, N. P. Walsh, C. J. McKnight, R. T.

Log in
My account
Contact us

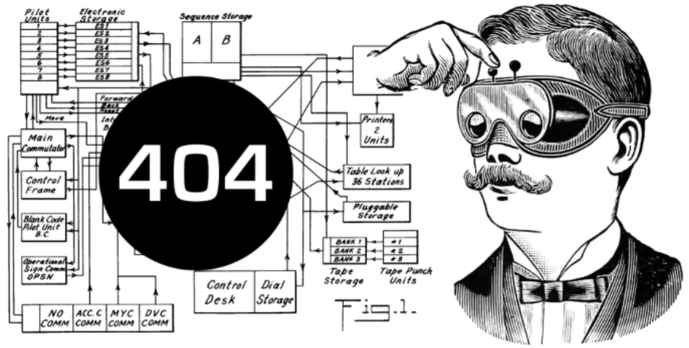
Become a member
Renew my subscription
Sign up for newsletters

Science AAAS

Authors | Members | Librarians | Advertisers

Home News Journals Topics Careers

Search



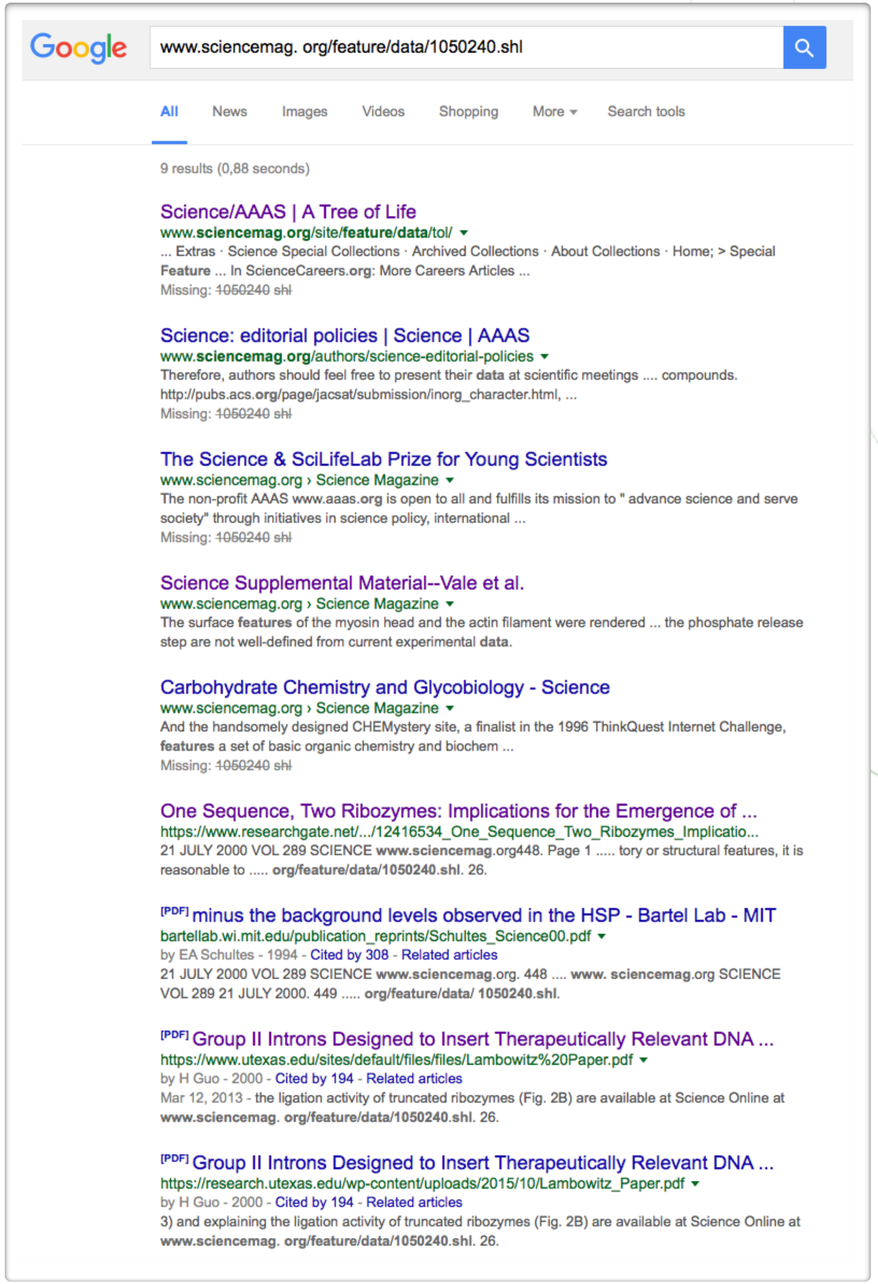
Hmmm...

This doesn't *look* like science.

It seems you're in search of a page that doesn't exist, or may have moved. You can use the Back button in your browser to return to the page that brought you here, or **search for your missing page**.

If you'd like to visit a page that has *plenty* of science on it, please visit our home page.

Visit the Home Page


Google
www.sciencemag. org/feature/data/1050240.shl
All
News
Images
Videos
Shopping
More
Search tools
9 results (0,88 seconds)
Science/AAAS | A Tree of Life
www.sciencemag.org/site/feature/data/tol/
... Extras · Science Special Collections · Archived Collections · About Collections · Home; > Special Feature ... In ScienceCareers.org: More Careers Articles ...
Missing: 1050240 shl
Science: editorial policies | Science | AAAS
www.sciencemag.org/authors/science-editorial-policies
Therefore, authors should feel free to present their data at scientific meetings .... compounds. http://pubs.acs.org/page/jacsat/submission/inorg_character.html, ...
Missing: 1050240 shl
The Science & SciLifeLab Prize for Young Scientists
www.sciencemag.org › Science Magazine
The non-profit AAAS www.aaas.org is open to all and fulfills its mission to " advance science and serve society" through initiatives in science policy, international ...
Missing: 1050240 shl
Science Supplemental Material--Vale et al.
www.sciencemag.org › Science Magazine
The surface features of the myosin head and the actin filament were rendered ... the phosphate release step are not well-defined from current experimental data.
Carbohydrate Chemistry and Glycobiology - Science
www.sciencemag.org › Science Magazine
And the handsomely designed CHEMystery site, a finalist in the 1996 ThinkQuest Internet Challenge, features a set of basic organic chemistry and biochem ...
Missing: 1050240 shl
One Sequence, Two Ribozymes: Implications for the Emergence of ...
https://www.researchgate.net/.../12416534_One_Sequence_Two_Ribozymes_Implicatio...
21 JULY 2000 VOL 289 SCIENCE www.sciencemag.org448. Page 1 ..... tory or structural features, it is reasonable to ..... org/feature/data/1050240.shl. 26.
[PDF] minus the background levels observed in the HSP - Bartel Lab - MIT
bartellab.wi.mit.edu/publication_reprints/Schultes_Science00.pdf
by EA Schultes - 1994 - Cited by 308 - Related articles
21 JULY 2000 VOL 289 SCIENCE www.sciencemag.org. 448 .... www. sciencemag.org SCIENCE VOL 289 21 JULY 2000. 449 ..... org/feature/data/ 1050240.shl.
[PDF] Group II Introns Designed to Insert Therapeutically Relevant DNA ...
https://www.utexas.edu/sites/default/files/files/Lambowitz%20Paper.pdf
by H Guo - 2000 - Cited by 194 - Related articles
Mar 12, 2013 - the ligation activity of truncated ribozymes (Fig. 2B) are available at Science Online at www.sciencemag. org/feature/data/1050240.shl. 26.
[PDF] Group II Introns Designed to Insert Therapeutically Relevant DNA ...
https://research.utexas.edu/wp-content/uploads/2015/10/Lambowitz_Paper.pdf
by H Guo - 2000 - Cited by 194 - Related articles
3) and explaining the ligation activity of truncated ribozymes (Fig. 2B) are available at Science Online at www.sciencemag. org/feature/data/1050240.shl. 26.

From: **Erik Schultes** erik.schultes@dtls.nl
Subject: Supplemental Data
Date: October 1, 2016 at 11:20 AM
To: science_editors@aaas.org

ES

To whom it may concern -

I authored this article in 2000:
http://science.sciencemag.org/content/289/5478/448.article-info

At the time, we had submitted Supplemental Data.

However, on the webpage supporting the paper, I no longer see a link to the Supplemental Data.

In endnote 25 of the paper, I see this link to the Supplemental Data:
www.sciencemag.org/feature/data/l050240.shl

However, this link appears to be broken, and redirects to a 404 page:
http://www.sciencemag.org/feature/data/l050240.shl

So I have a few questions:

(1) Is it possible, one way or another, to obtain or access the original Supplemental Data ?

(2) My immediate interest is not so much at the content level, but more at the 'Data Stewardship' level. Recently, my colleagues and I conducted a LERU Summer School on Data Stewardship (in Leiden, the Netherlands, see http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/). I actually used this 'broken link' as a case study for the kinds of problems we all face in scholarly data stewardship. So I'm wondering, can you tell me when and why the link was broken ? It is instructive for us to better understand the challenges that publishers face these days.

Many thanks in advance,

Best regards,

Erik

--
Erik Schultes PhD
FAIR Data Scientific Projects Lead

Dutch Techcentre for Life Sciences
Visiting address: Catharijnesingel 54 | 3511 GC Utrecht
Postal address: Postbus 19245 | 3501 DE Utrecht

E-mail: erik.schultes@dtls.nl
Skype: easchultes
Tel: +31 642448027
Website: www.dtls.nl

http://www.sciencemag.org/site/feature/data/1050240.xhtml
October 10, 2016 12:45 CET

ABSTRACT FULL TEXT **One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds**
Erik A. Schultes and David P. Bartel

Supplementary Material

Supplemental Figure 1. Secondary structures of the ribozyme variants aligned in Fig. 3A. Accumulated changes from the intersection sequence are indicated (red and blue residues, with blue identifying residues changed at the step indicated by an arrow; dashes mark sites of point deletions). Self-ligation and self-cleavage rates (min^{-1}) are listed below the name of each construct.

Larger, segmented version of this image (new window – 160K JPEG images) | Full version of this image (827K JPEG file)] | Download PDF of this image (9MB)

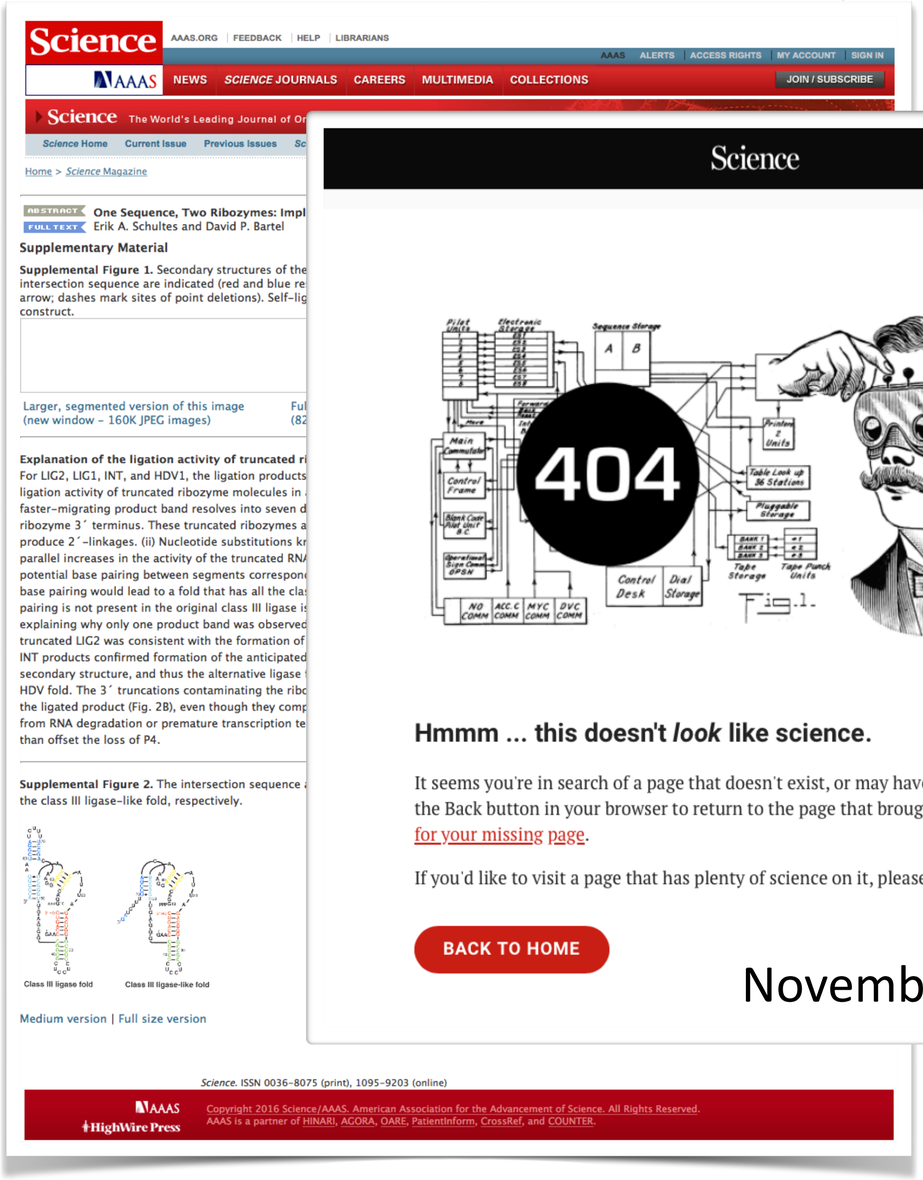
Explanation of the ligation activity of truncated ribozymes.
For LIG2, LIG1, INT, and HDV1, the ligation products are represented by two bands in Fig. 2B (top panel), due to unanticipated ligation activity of truncated ribozyme molecules in addition to the activity of full-length RNA. On lower percentage gels, the faster-migrating product band resolves into seven distinct bands, corresponding to a loss of 9 to 15 nucleotides at the ribozyme 3′ terminus. These truncated ribozymes appear to adopt a class III–like fold as evidenced by the following: (i) They produce 2′–linkages. (ii) Nucleotide substitutions known to favor the class III ligase function and fold (C13A and C38G) lead to parallel increases in the activity of the truncated RNAs (Fig 2). (iii) Constructs LIG2, LIG1, INT, and HDV1 have fortuitous potential base pairing between segments corresponding to G51–U55 and A66–U70 in Fig. 2A. For the truncated molecules, this base pairing would lead to a fold that has all the class III features except the P4 stem–loop (Web fig. 2). The fortuitous potential pairing is not present in the original class III ligase isolates or in the constructs of the ligase neutral path (LIG4–LIG P in Fig. 3), explaining why only one product band was observed for LIG4–LIG P. (iv) The Pb(II)–cleavage pattern of the ligation product of truncated LIG2 was consistent with the formation of G51–U55:A66–U70 pairing, whereas probing of full–length LIG2, LIG1, and INT products confirmed formation of the anticipated P4 and P5 stems. None of the fortuitous pairs are present in the HDV secondary structure, and thus the alternative ligase fold, like the prototype ligase fold, has no base pairs in common with the HDV fold. The 3′ truncations contaminating the ribozyme preparations of LIG2, LIG1, INT, and HDV1 can generate over half of the ligated product (Fig. 2B), even though they compose only a minor fraction of the ribozyme preparation (presumably arising from RNA degradation or premature transcription termination). For these truncated molecules, a more optimal P5 might more than offset the loss of P4.

Supplemental Figure 2. The intersection sequence and a truncated form of this sequence assuming the class III ligase fold and the class III ligase–like fold, respectively.

Medium version | Full size version

Science. ISSN 0036–8075 (print), 1095–9203 (online)

http://www.sciencemag.org/site/feature/data/1050240.xhtml
October 10, 2016 12:45 CET



November 5, 2024 09:33 CET

Data Stewardship: FAIR or otherwise

13:40-14:00
(Erik)

What is Data Stewardship ?

What is Data Stewardship ?

Stewardship is an ethic that embodies the responsible planning and management of resources. Can be applied to nature, economics, health, property, information, theology, etc.

https://en.wikipedia.org/wiki/Stewardship

What is Data Stewardship ?

Data Stewardship: responsible planning and management of data

What is Data Stewardship ?

Data Stewardship: responsible planning and management of data

Data Stewardship: The entire process that deals responsibly with one's own and other peoples data throughout and after the scientific discovery process (First HLEG EOSC 2016).

What is Data Stewardship ?

Data Stewardship: responsible planning and management of data

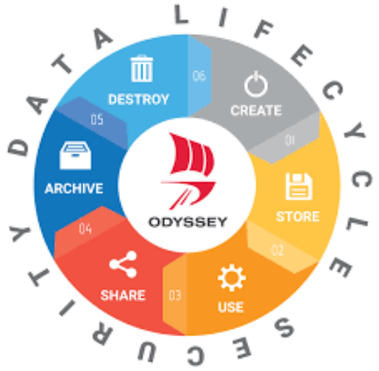
- Project and post-project
- Resource balance (mandate and funded mandate)
- Future-proof, tech-proof
- Choices and hard-choices

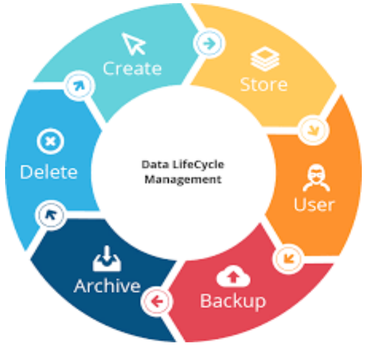
What is Data Stewardship ?

Data Stewardship: responsible planning and management of data

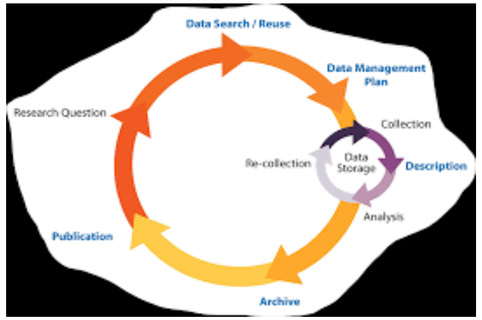
- Project and post-project
- Resource balance (mandate and funded mandate)
- Future-proof, tech-proof
- Choices and hard-choices

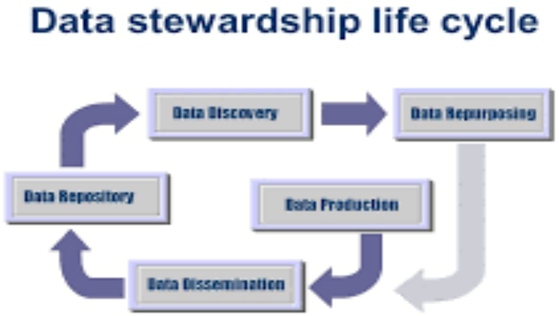
Data Stewardship: getting systematic

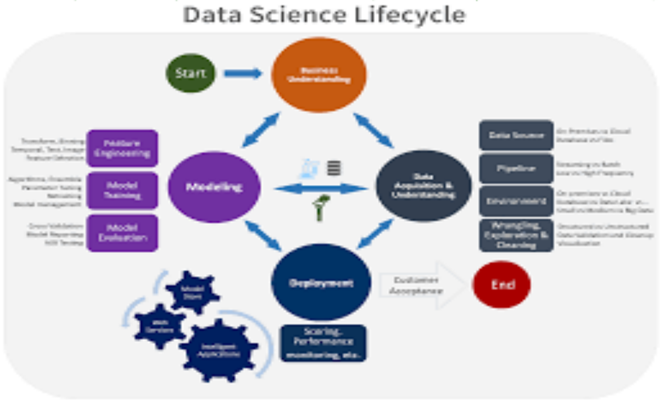


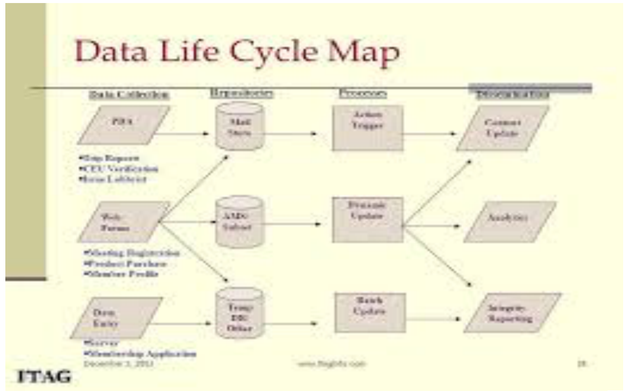


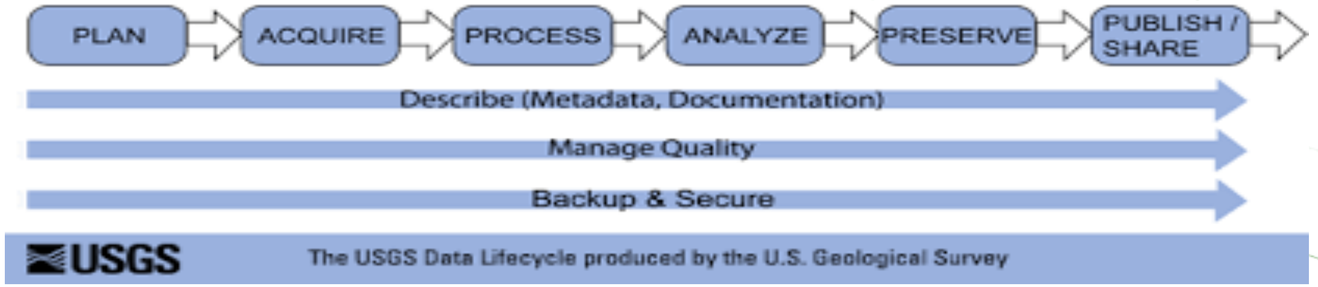
Data Stewardship: getting systematic

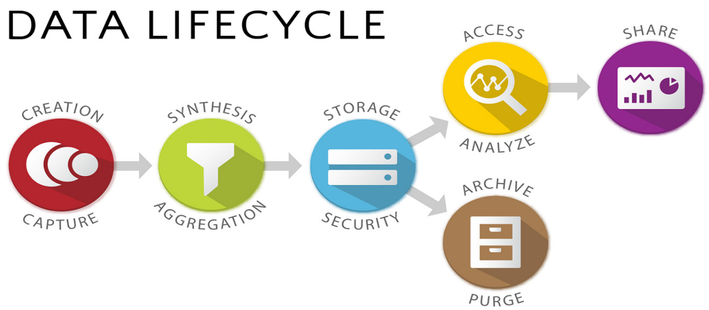




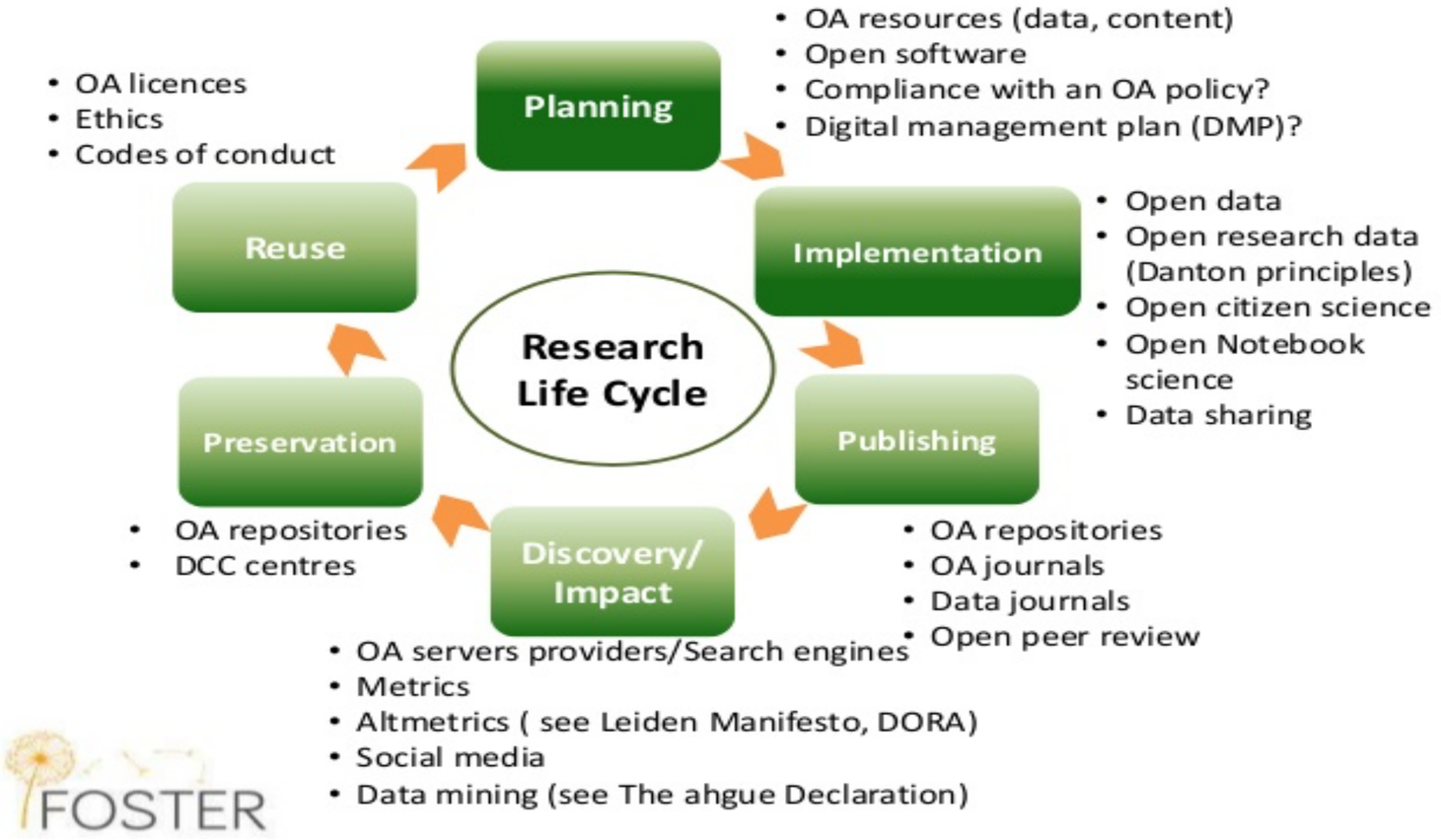








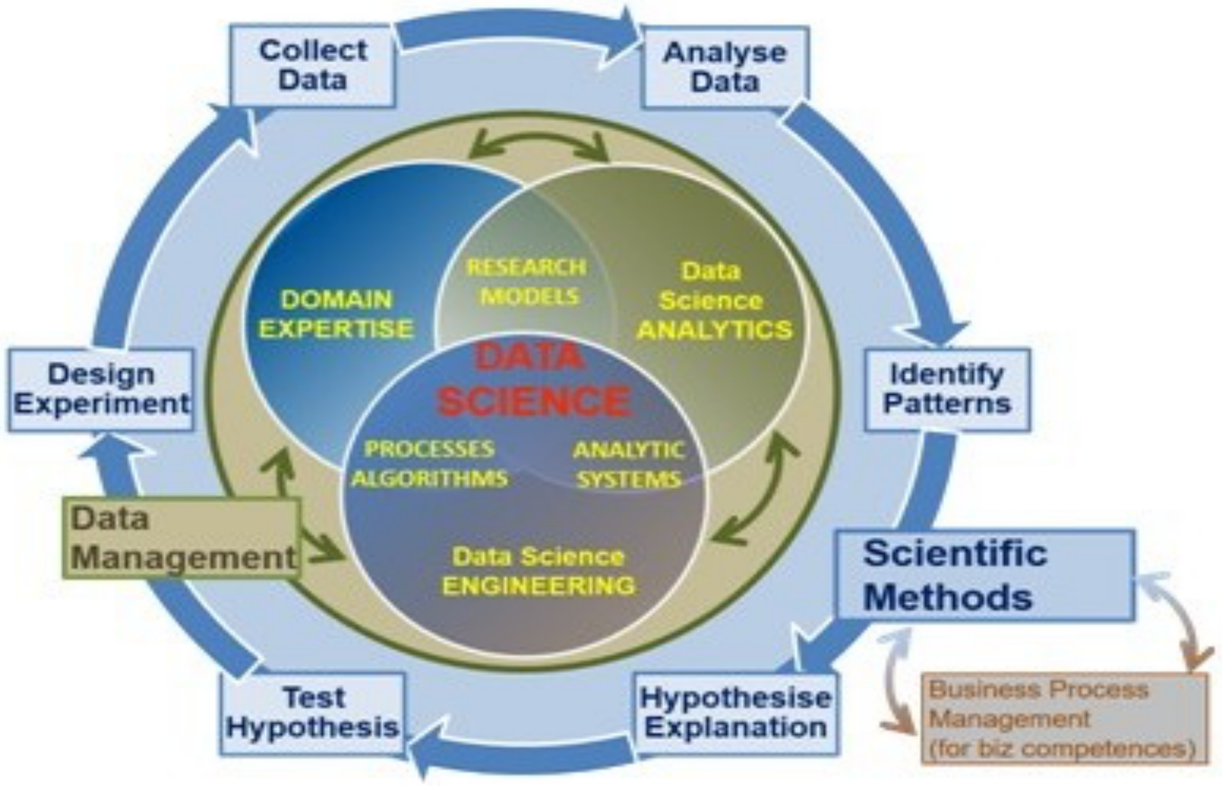
Data Stewardship: getting systematic



FOSTER: practical implementation of Open Science in Horizon 2020 and beyond

https://www.fosteropenscience.eu/

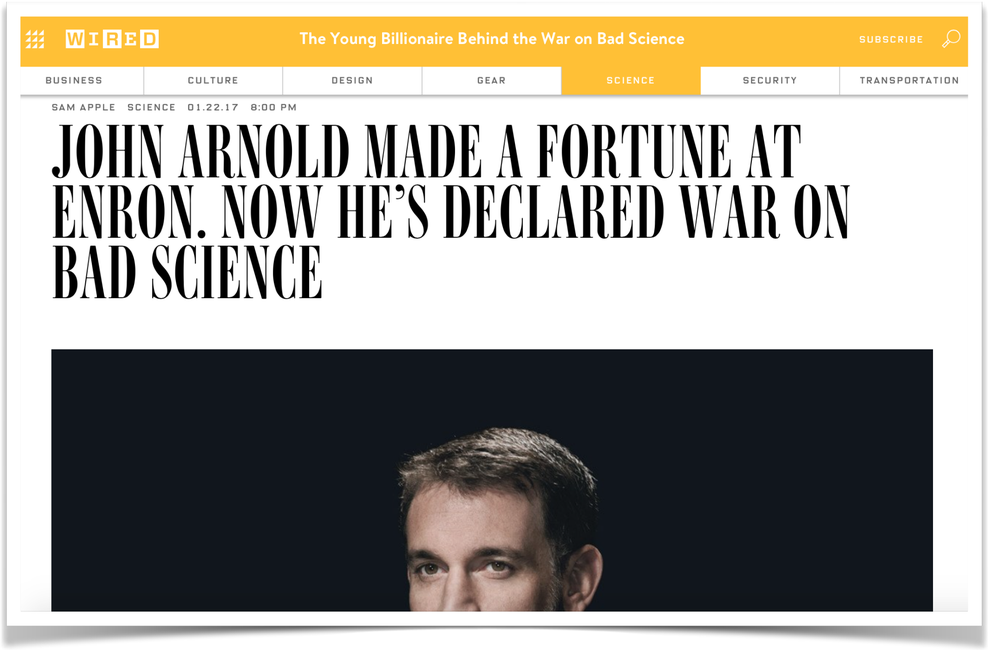
Data Stewardship: getting systematic

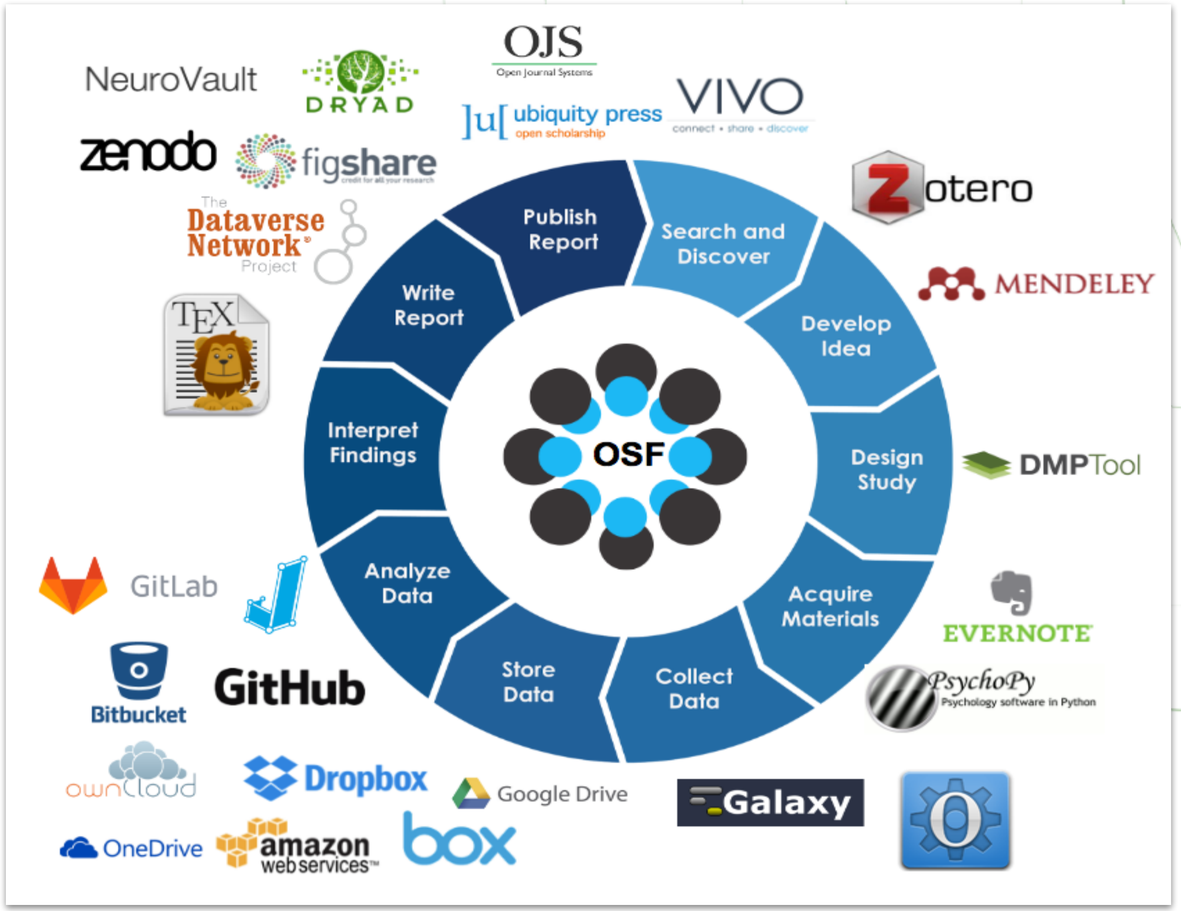


EDISON Data Science Framework to define the Data Science Profession

http://edison-project.eu/edison

Data Stewardship: getting systematic

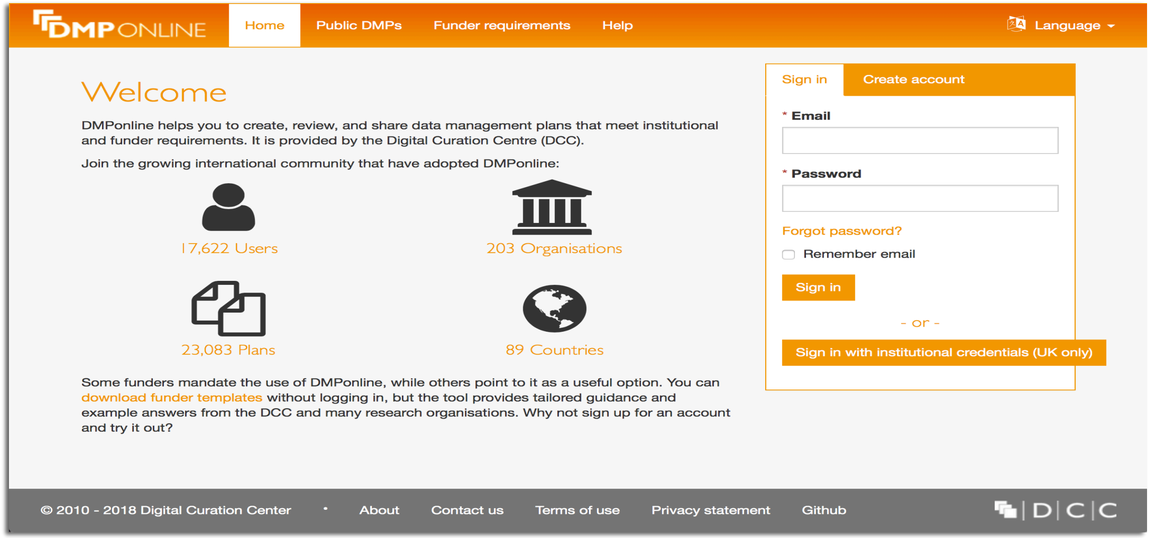




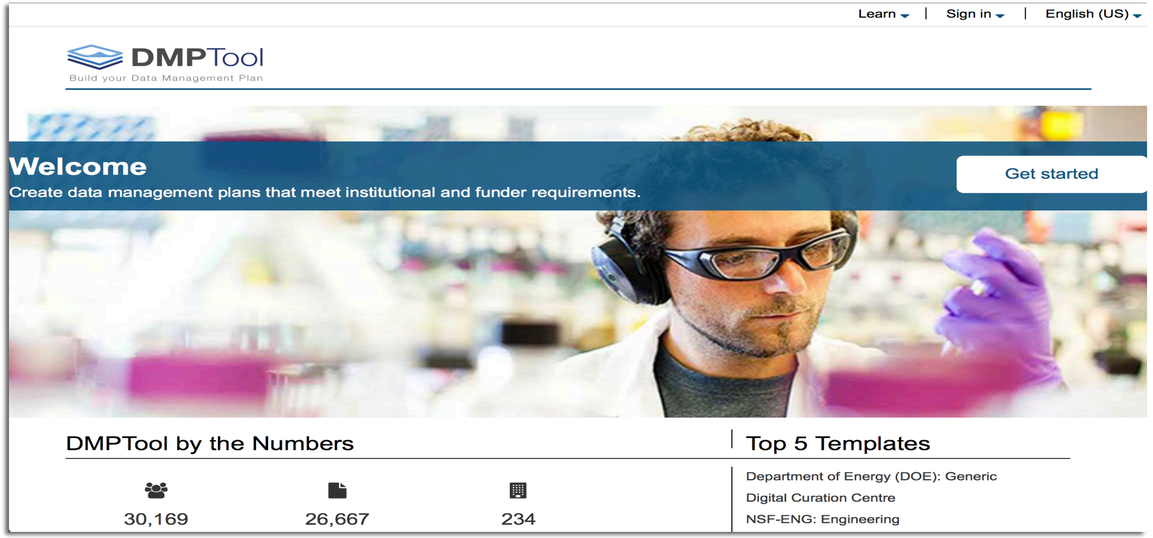
Center of Open Science, providing tools, training, support and advocacy for changes in incentives to make research investment go farther, faster.

https://cos.io

Data Stewardship: getting systematic

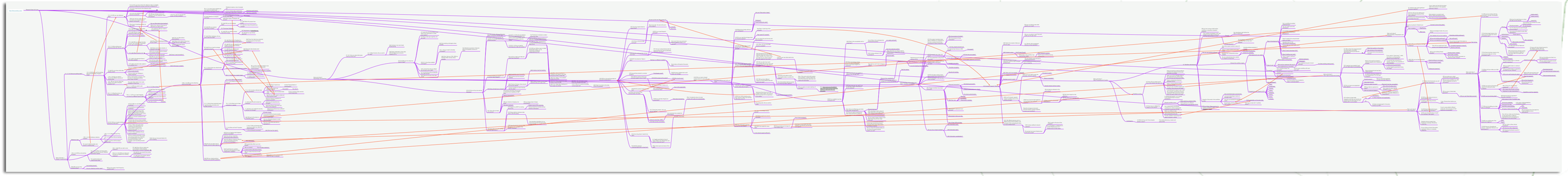


https://dmponline.dcc.ac.uk



https://dmptool.org

Data Stewardship: getting systematic

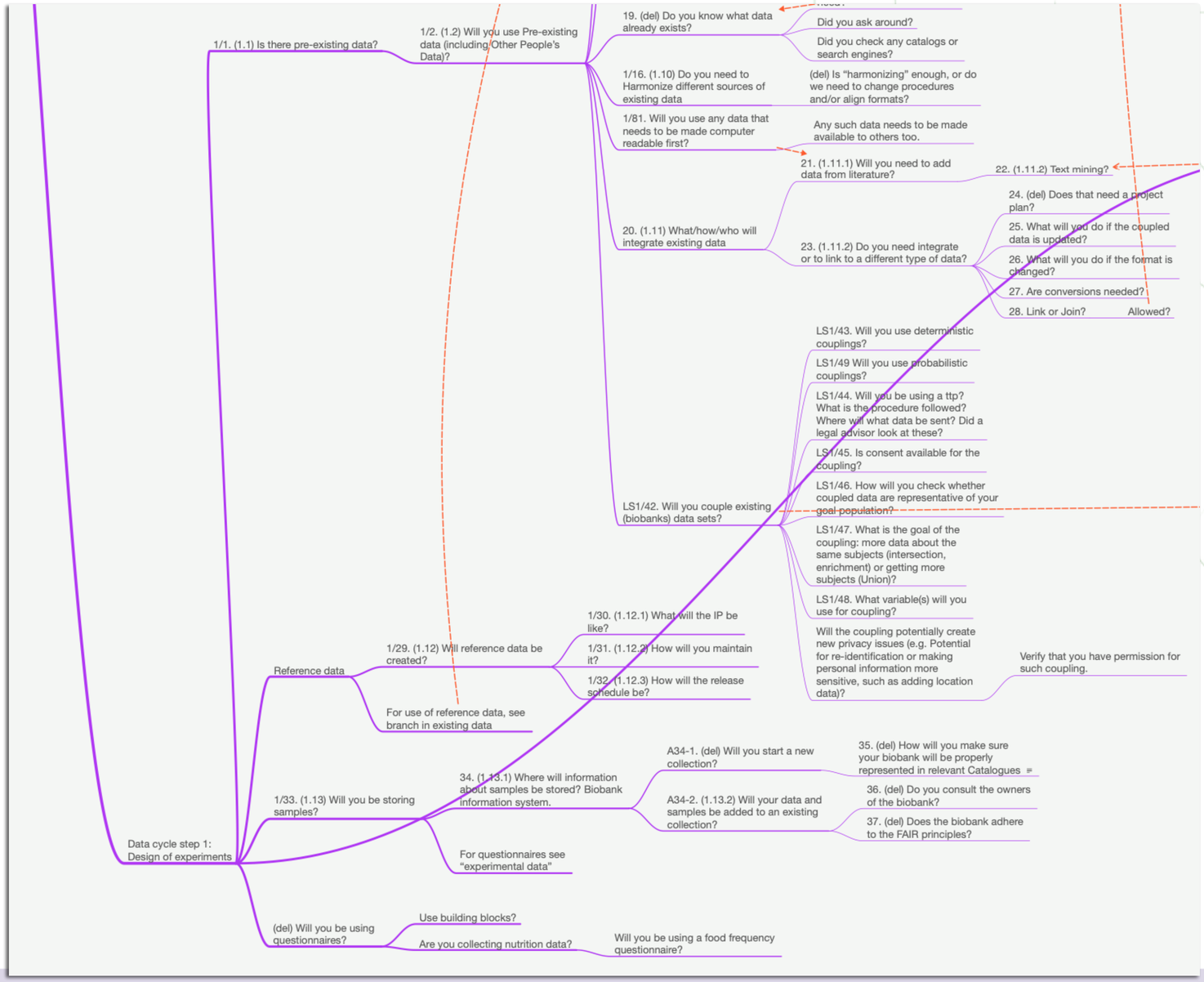


4 meters

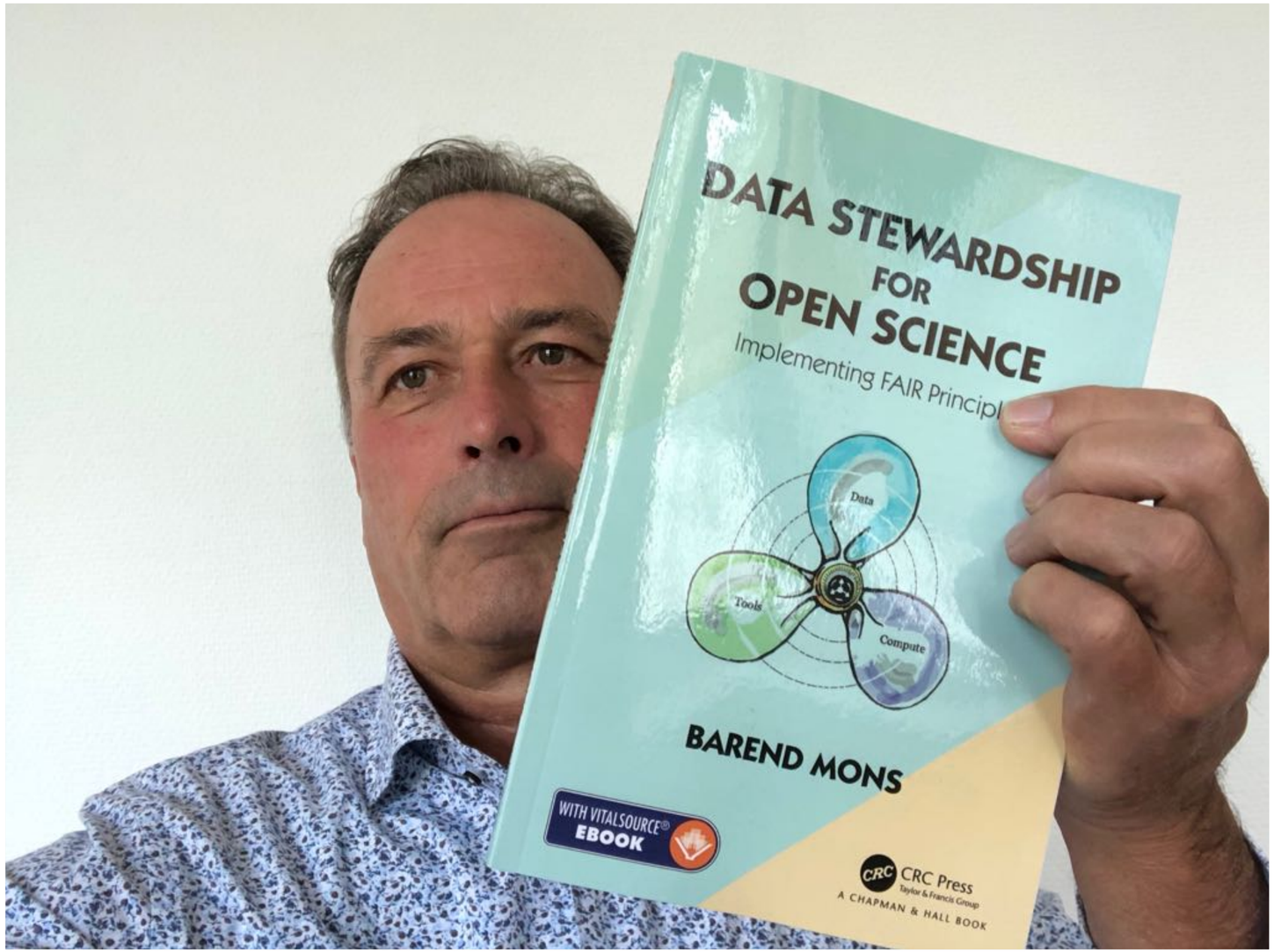
Rob W.W. Hooft. (2019). Data Stewardship Mindmap (Version 20190725). Zenodo. https://doi.org/10.5281/zenodo.3351949

Data Cycle Step 1: Design of Experiment
Data Cycle Step 2: Data Design and Planning
Data Cycle Step 3: Data Capture (Equipment)
Data Cycle Step 4: Data Processing and Curation
Data Cycle Step 5: Data Linking and Integration
Data Cycle Step 6: Data Analysis and Interpretation
Data Cycle Step 7: Publishing

Data Stewardship: getting systematic



Data Stewardship: getting systematic

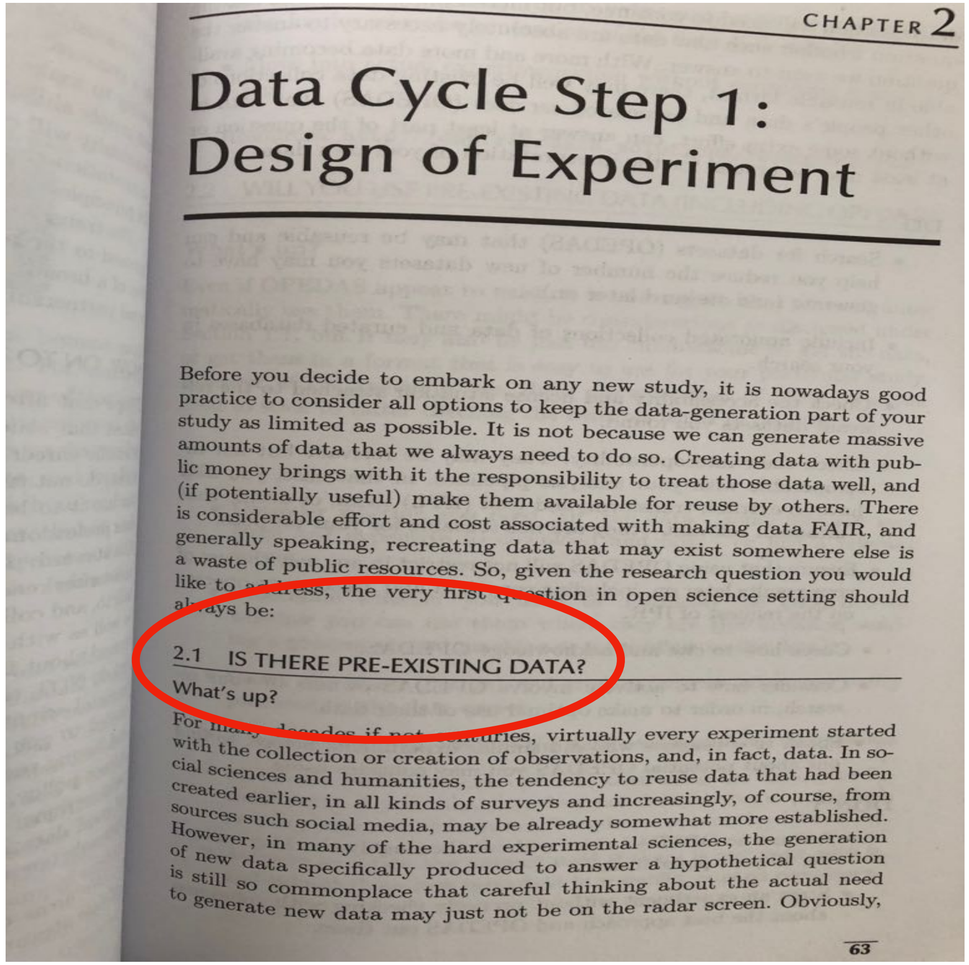


Data Stewardship for Open Science Implementing FAIR Principles 2018

Data Stewardship for Open Science: Implementing FAIR Principles has been written with the intention of making scientists, funders, and innovators in all disciplines and stages of their professional activities broadly aware of the need, complexity, and challenges associated with open science, modern science communication, and data stewardship. The FAIR principles are used as a guide throughout the text, and this book should leave experimentalists consciously incompetent about data stewardship and motivated to respect data stewards as representatives of a new profession, while possibly motivating others to consider a career in the field.

The ebook, avalable for no additional cost when you buy the paperback, will be updated every 6 months on average (providing that significant updates are needed or avaialble). Readers will have the opportunity to contribute material towards these updates, and to develop their own data management plans, via the free Data Stewardship Wizard.

Data Stewardship: getting systematic



CHAPTER 2

Data Cycle Step 1: Design of Experiment

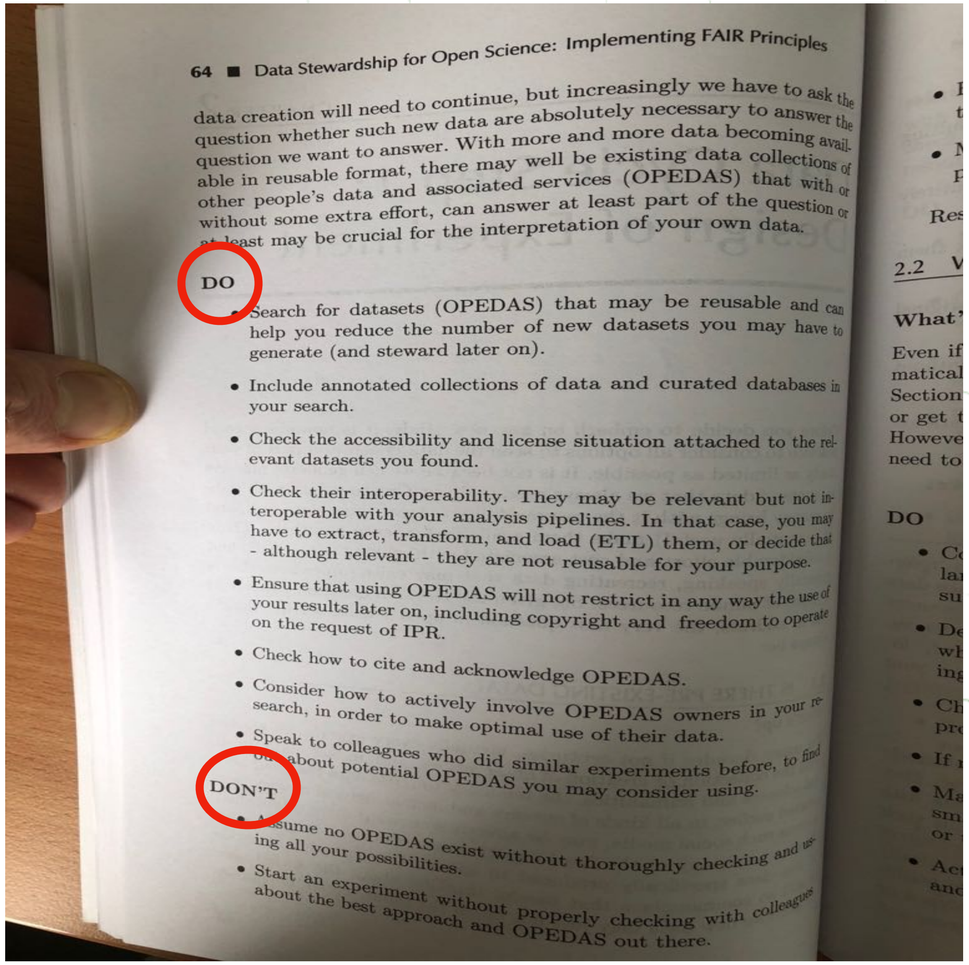
Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data-generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money brings with it the responsibility to treat those data well, and (if potentially useful) make them available for reuse by others. There is considerable effort and cost associated with making data FAIR, and generally speaking, recreating data that may exist somewhere else is a waste of public resources. So, given the research question you would like to address, the very first question in open science setting should always be:

2.1 IS THERE PRE-EXISTING DATA?

What's up?

For many decades if not centuries, virtually every experiment started with the collection or creation of observations, and, in fact, data. In social sciences and humanities, the tendency to reuse data that had been created earlier, in all kinds of surveys and increasingly, of course, from sources such social media, may be already somewhat more established. However, in many of the hard experimental sciences, the generation of new data specifically produced to answer a hypothetical question is still so commonplace that careful thinking about the actual need to generate new data may just not be on the radar screen. Obviously,

63



64 ■ Data Stewardship for Open Science: Implementing FAIR Principles

data creation will need to continue, but increasingly we have to ask the question whether such new data are absolutely necessary to answer the question we want to answer. With more and more data becoming available in reusable format, there may well be existing data collections of other people's data and associated services (OPEDAS) that with or without some extra effort, can answer at least part of the question or at least may be crucial for the interpretation of your own data.

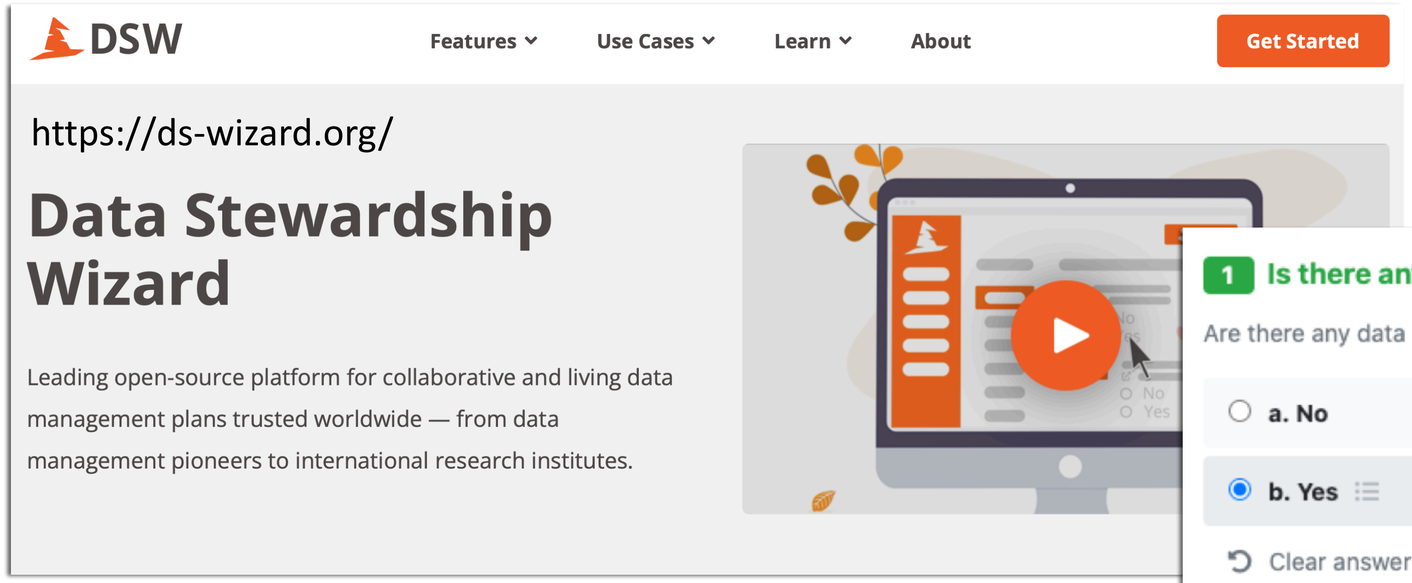
DO

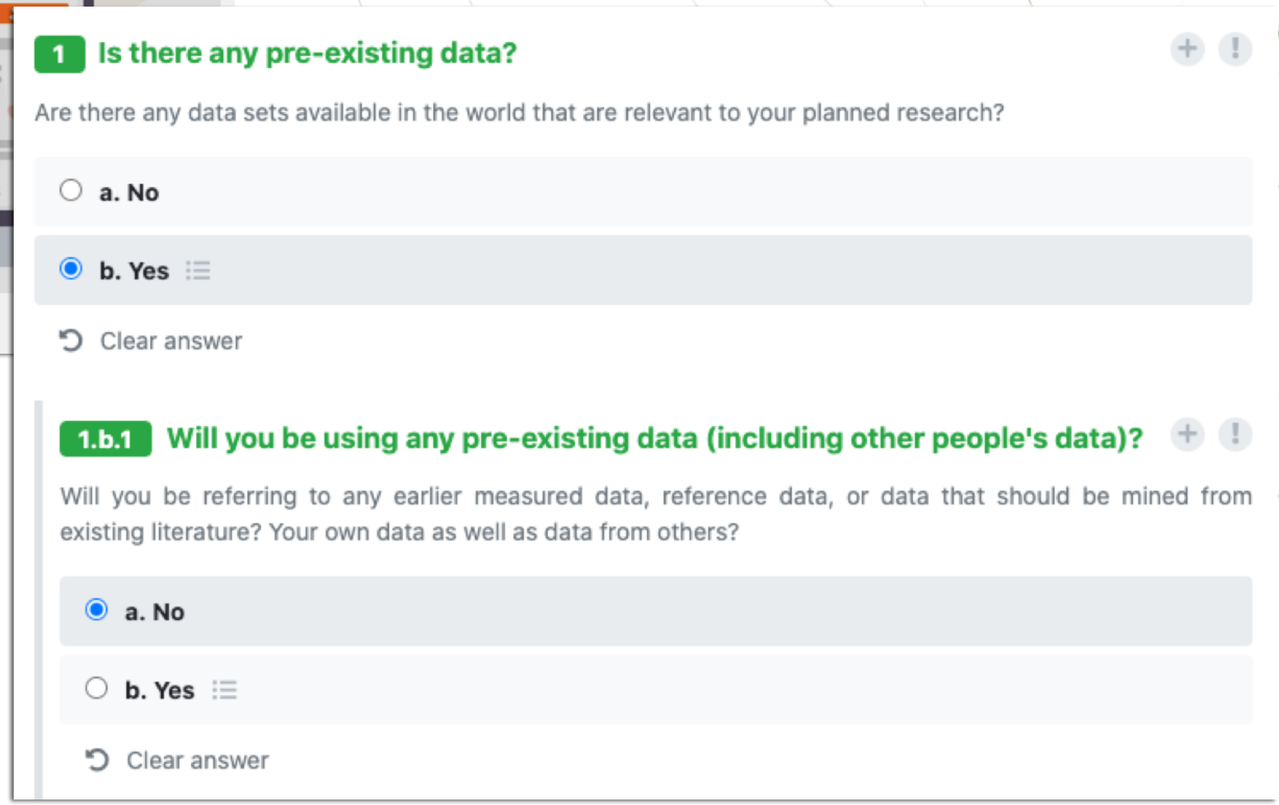
- Search for datasets (OPEDAS) that may be reusable and can help you reduce the number of new datasets you may have to generate (and steward later on).
- Include annotated collections of data and curated databases in your search.
- Check the accessibility and license situation attached to the relevant datasets you found.
- Check their interoperability. They may be relevant but not interoperable with your analysis pipelines. In that case, you may have to extract, transform, and load (ETL) them, or decide that - although relevant - they are not reusable for your purpose.
- Ensure that using OPEDAS will not restrict in any way the use of your results later on, including copyright and freedom to operate on the request of IPR.
- Check how to cite and acknowledge OPEDAS.
- Consider how to actively involve OPEDAS owners in your research, in order to make optimal use of their data.
- Speak to colleagues who did similar experiments before, to find out about potential OPEDAS you may consider using.

DON'T

- Assume no OPEDAS exist without thoroughly checking and using all your possibilities.
- Start an experiment without properly checking with colleagues about the best approach and OPEDAS out there.

Data Stewardship: getting systematic

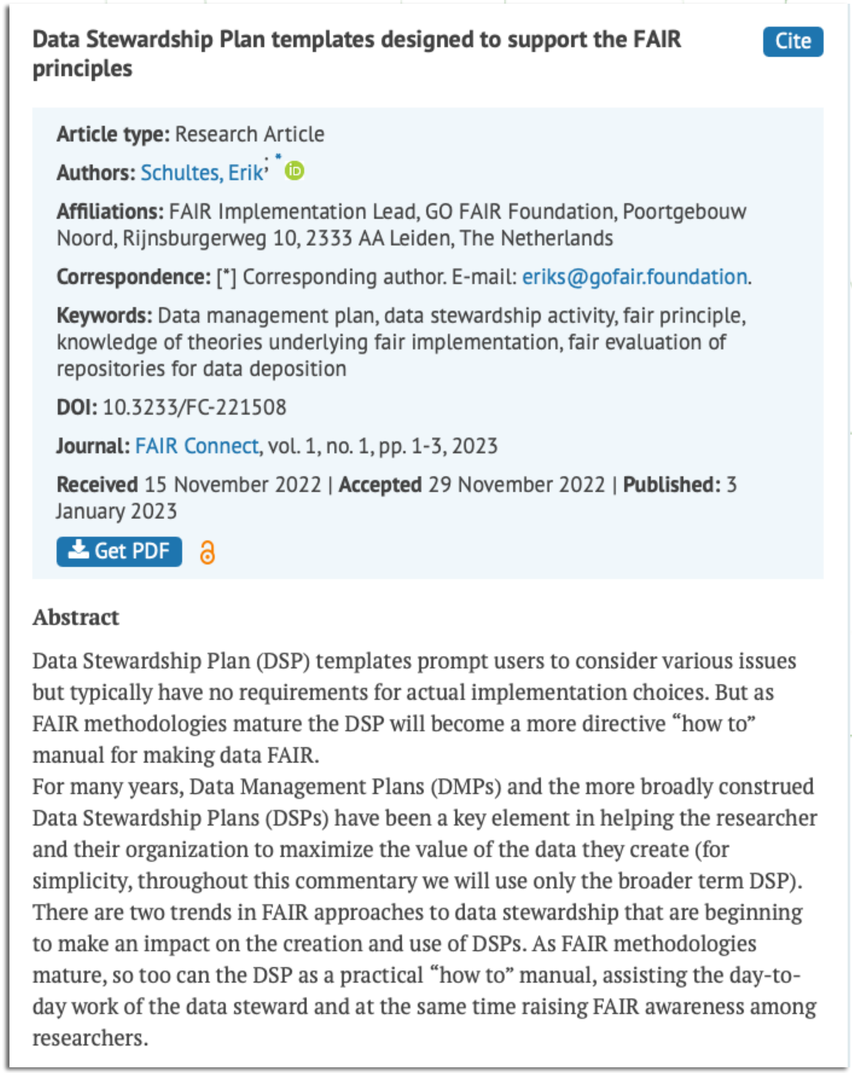




Hooft, Rob, Suchánek, Marek, and Pergl, Robert. 'Data Stewardship Wizard'. 1 Jan. 2023 : 41 – 43. https://content.iospress.com/articles/fair-connect/fc230501

Data Stewardship: getting systematic

Schultes, Erik. ‘Data Stewardship Plan Templates Designed to Support the FAIR Principles’. 1 Jan. 2023 : 1 – 3. https://content.iospress.com/articles/fair-connect/fc221508



Data Stewardship Plan templates designed to support the FAIR principles

Cite

Article type: Research Article

Authors: Schultes, Erik[*]

Affiliations: FAIR Implementation Lead, GO FAIR Foundation, Poortgebouw Noord, Rijnsburgerweg 10, 2333 AA Leiden, The Netherlands

Correspondence: [*] Corresponding author. E-mail: eriks@gofair.foundation.

Keywords: Data management plan, data stewardship activity, fair principle, knowledge of theories underlying fair implementation, fair evaluation of repositories for data deposition

DOI: 10.3233/FC-221508

Journal: FAIR Connect, vol. 1, no. 1, pp. 1-3, 2023

Received 15 November 2022 | **Accepted** 29 November 2022 | **Published:** 3 January 2023

Get PDF

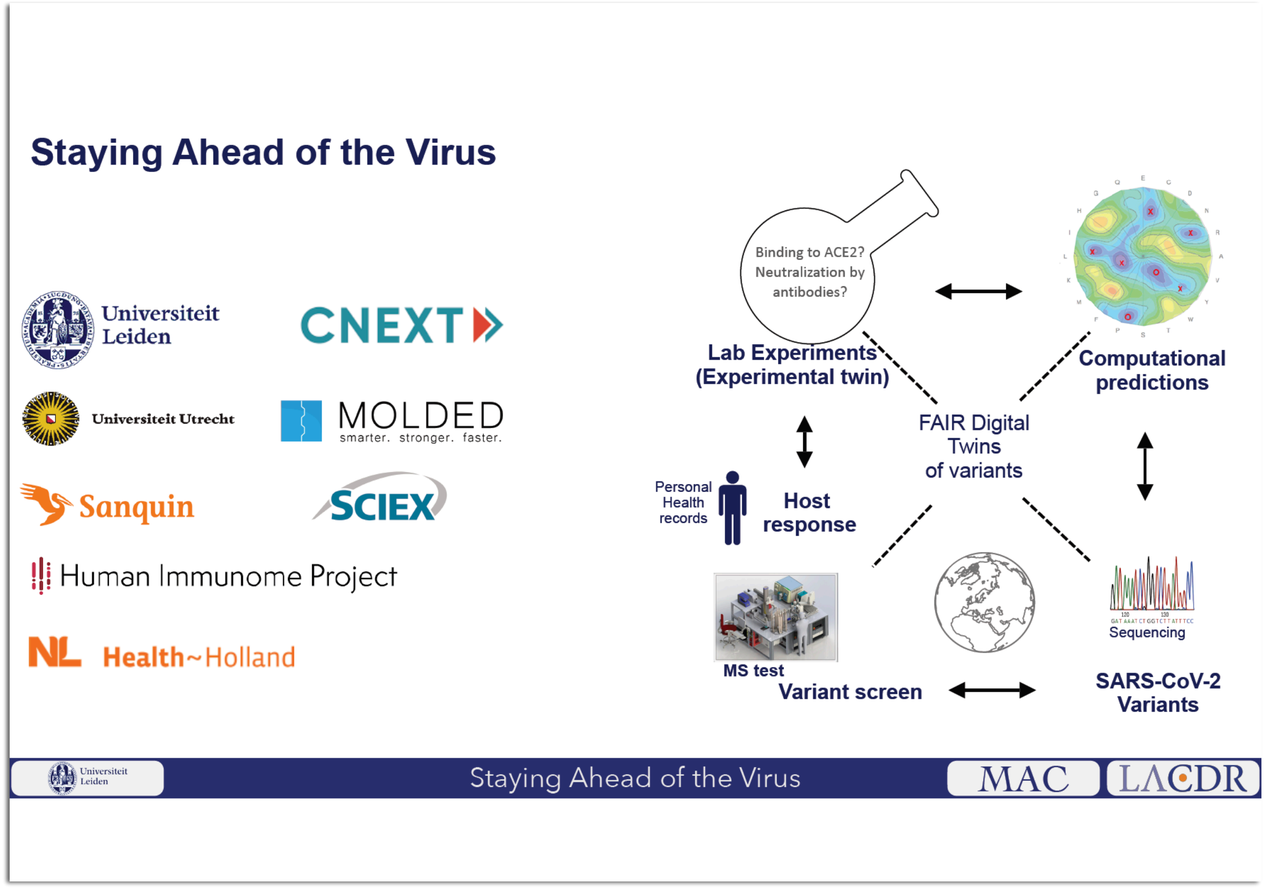
Abstract

Data Stewardship Plan (DSP) templates prompt users to consider various issues but typically have no requirements for actual implementation choices. But as FAIR methodologies mature the DSP will become a more directive “how to” manual for making data FAIR.

For many years, Data Management Plans (DMPs) and the more broadly construed Data Stewardship Plans (DSPs) have been a key element in helping the researcher and their organization to maximize the value of the data they create (for simplicity, throughout this commentary we will use only the broader term DSP). There are two trends in FAIR approaches to data stewardship that are beginning to make an impact on the creation and use of DSPs. As FAIR methodologies mature, so too can the DSP as a practical “how to” manual, assisting the day-to-day work of the data steward and at the same time raising FAIR awareness among researchers.

Data Stewardship: getting systematic

Health~Holland
SHARED CHALLENGES, SMART SOLUTIONS



https://www.health-holland.com/project/2023/2022/staying-ahead-virus

Outline of workplan

The work plan will comprise seven interconnected work packages. The tasks and deliverables (see GANTT chart, Figure 7) from WP1-4 aim to facilitate WP5. The tasks and deliverables from WP1-5 aim to prove the combined approach and scalability of the solutions, and to facilitate a draft roadmap to further develop and implement our ambition (WP6).

WP1: FAIR Hourglass for data management and FAIR Digital Twins of protein variants

As a governing principle for this project, we adopt emerging best practices for FAIR compliant data stewardship[13] (ref, ref). Because this project should also demonstrate the feasibility and global scalability of our approach, where multiple data formats and standards must be accommodated, we adopt the FAIR Hourglass model of data management (Figure 4): the shape is meant to indicate expanding "freedom to operate" (top and bottom) with only absolute, minimal standards required at the center[14]. The FAIR Hourglass model makes data available to humans and machines and ensures that the data created in this project are technology-proof, vendor-proof, and future-proof.

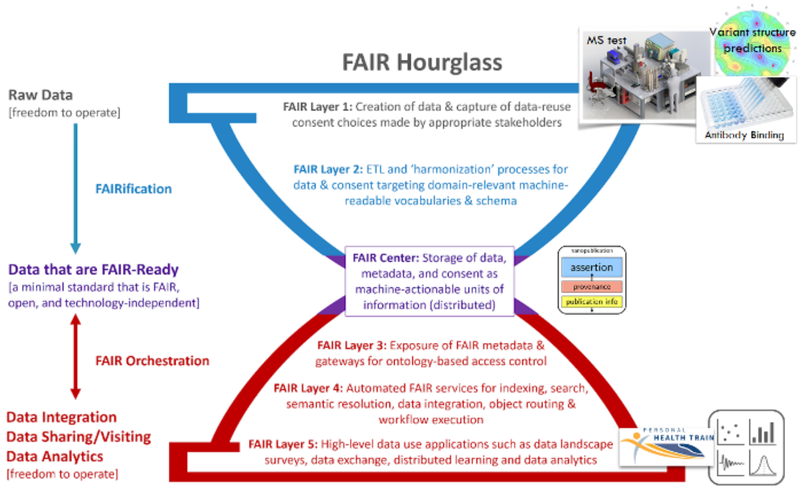


Figure 4. FAIR hourglass model consists of five operational layers that enable maximum freedom of operation to data producers (Layer 1) and data users (Layer 5) through the commitment to a minimal data standard at the center (Nanopublication). The automated workflows that support data conversion to nanopublications described in Layer 2 are the "FAIRification" processes (blue). On the bottom (red), automated workflows that enable data access, integration, and analytics described in layers 3-5 are "FAIR Orchestration".

Paul Ayris
Jean-Yves Berthou
Rachel Bruce (Rapporteur)
Stefanie Lindstaedt
Anna Monreale
Barend Mons (Chair)
Yasuhiro Murayama (Observer, Japan),
Caj Södergård
Klaus Tochtermann
Ross Wilkinson (Observer, Australia).

- 2016
- 24 pages
- “FAIR” mentioned 22 times
- “Data steward” mentioned 27 time
- Hourglass architecture (page 14)

Paul Ayris
Jean-Yves Berthou
Rachel Bruce (Rapporteur)
Stefanie Lindstaedt
Anna Monreale
Barend Mons (Chair)
Yasuhiro Murayama (Observer, Japan),
Caj Södergård
Klaus Tochtermann
Ross Wilkinson (Observer, Australia).

Some recommendations:

Make adequate data stewardship mandatory for all research proposals.

data stewardship and (re-) analysis to support the final aim of science: knowledge discovery.

Consolidate and further develop material and tools for the construction and review of Data Management Plans (including budgeting for re-use of data) and Data Stewardship plans (including budgeting for data publication and long-term preservation in FAIR status)

By 2020, to have in each Member State and for each discipline at least one certified institute to support implementation of Data Stewardship per discipline

Also …. https://www.openscience.nl/sites/open_science/files/media-files/professionalising_datastewardship.pdf

https://interoperable-europe.ec.europa.eu/collection/open-government/news/500000-data-scientists-need

https://www.nature.com/articles/d41586-020-00505-7

WORLD VIEW | 25 February 2020

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

By Barend Mons

Many of the world's hardest problems can be tackled only with data-intensive, computer-assisted research. And I'd speculate that the vast majority of research data are never published. Huge sums of taxpayer funds go to waste because such data cannot be reused. Policies for data reuse are falling into place, but fixing the situation will require more resources than the scientific community is willing to face.

In 2013, I was part of a group of Dutch experts from many disciplines that called on our national science funder to support data stewardship. Seven years later, policies that I helped to draft are starting to be put into practice. These require data created by machines and humans to meet the FAIR principles (that is, they are findable, accessible, interoperable and reusable). I now direct an international Global Open FAIR office tasked with helping communities to implement the guidelines, and I am convinced that doing so will require a large cadre of professionals, about one for every 20 researchers.

Even when data are shared, the metadata, expertise, technologies and infrastructure necessary for reuse are lacking. Most published data sets are scattered into 'supplemental files' that are often impossible for machines or even humans to find. These and other sloppy data practices keep researchers from building on each other's work. In cases of disease outbreaks, for instance, this might even cost lives.

"Funders hold the stick: they should disburse no further funding without a data-stewardship plan."

nature

nature

Explore content · About the journal · Publish with us · Subscribe

nature › world view › article

WORLD VIEW | 25 February 2020

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

By Barend Mons

Many of the world's hardest problems can be tackled only with data-intensive, computer-assisted research. And I'd speculate that the vast majority of research data are never published. Huge sums of taxpayer funds go to waste because such data cannot be reused. Policies for data reuse are falling into place, but fixing the situation will require more resources than the scientific community is willing to face.

In 2013, I was part of a group of Dutch experts from many disciplines that called on our national science funder to support data stewardship. Seven years later, policies that I helped to draft are starting to be put into practice. These require data created by machines and humans to meet the FAIR principles (that is, they are findable, accessible, interoperable and reusable). I now direct an international Global Open FAIR office tasked with helping communities to implement the guidelines, and I am convinced that doing so will require a large cadre of professionals, about one for every 20 researchers.

Even when data are shared, the metadata, expertise, technologies and infrastructure necessary for reuse are lacking. Most published data sets are scattered into 'supplemental files' that are often impossible for machines or even humans to find. These and other sloppy data practices keep researchers from building on each other's work. In cases of disease outbreaks, for instance, this might even cost lives.

https://www.nature.com/articles/d41586-020-00505-7

ITINERIS

…on average, 5% of overall research costs should go towards data stewardship. With €300 billion (US$325 billion) of public money spent on research in the European Union, we should expect to spend €15 billion on data stewardship. Scientists, especially more experienced ones, are often upset when I say this. They see it as 5% less funding for research.

Bunk.

1) taking care of data is an ethical duty, and should be part of good research practice.
2) if data are treated properly, researchers will have significantly more time to do research… Students in PhD programmes spend up to 80% of their time on 'data munging', fixing formatting and minor mistakes to make data suitable for analysis — wasting time and talent.

Data stewardship offers excellent returns on investment.

In preparing for battle I have always found that plans are useless, but planning is indispensable.

Dwight D. Eisenhower

Data life cycle & Research objects

14:00-14:45
(Interactive)

Resume 15:25 CET

Data Management Planning

15:00-15:50
(Alessa)

DMP and the FAIR Hourglass - the same thing?

Outline of workplan
The work plan will comprise seven interconnected work packages. The tasks and deliverables (see GANTT chart, Figure 7) from WP1-4 aim to facilitate WP5. The tasks and deliverables from WP1-5 aim to prove the combined approach and scalability of the solutions, and to facilitate a draft roadmap to further develop and implement our ambition (WP6).

WP1: FAIR Hourglass for data management and FAIR Digital Twins of protein variants

As a governing principle for this project, we adopt emerging best practices for FAIR compliant data stewardship[13] (ref, ref). Because this project should also demonstrate the feasibility and global scalability of our approach, where multiple data formats and standards must be accommodated, we adopt the FAIR Hourglass model of data management (Figure 4): the shape is meant to indicate expanding "freedom to operate" (top and bottom) with only absolute, minimal standards required at the center[14]. The FAIR Hourglass model makes data available to humans and machines and ensures that the data created in this project are technology-proof, vendor-proof, and future-proof.

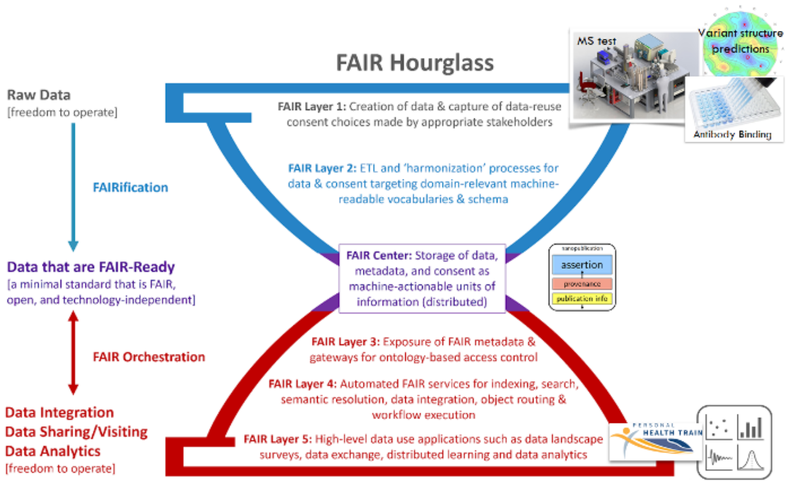


Figure 4. FAIR hourglass model consists of five operational layers that enable maximum freedom of operation to data producers (Layer 1) and data users (Layer 5) through the commitment to a minimal data standard at the center (Nanopublication). The automated workflows that support data conversion to nanopublications described in Layer 2 are the "FAIRification" processes (blue). On the bottom (red), automated workflows that enable data access, integration, and analytics described in layers 3-5 are "FAIR Orchestration".

DMP and the FAIR Hourglass - the same thing?

Outline of workplan
The work plan will comprise seven interconnected work packages. The tasks and deliverables (see GANTT chart, Figure 7) from WP1-4 aim to facilitate WP5. The tasks and deliverables from WP1-5 aim to prove the combined approach and scalability of the solutions, and to facilitate a draft roadmap to further develop and implement our ambition (WP6).

WP1: FAIR Hourglass for data management and FAIR Digital Twins of protein variants

As a governing principle for this project, we adopt emerging best practices for FAIR compliant data stewardship[13] (ref, ref). Because this project should also demonstrate the feasibility and global scalability of our approach, where multiple data formats and standards must be accommodated, we adopt the FAIR Hourglass model of data management (Figure 4): the shape is meant to indicate expanding "freedom to operate" (top and bottom) with only absolute, minimal standards required at the center[14]. The FAIR Hourglass model makes data available to humans and machines and ensures that the data created in this project are technology-proof, vendor-proof, and future-proof.

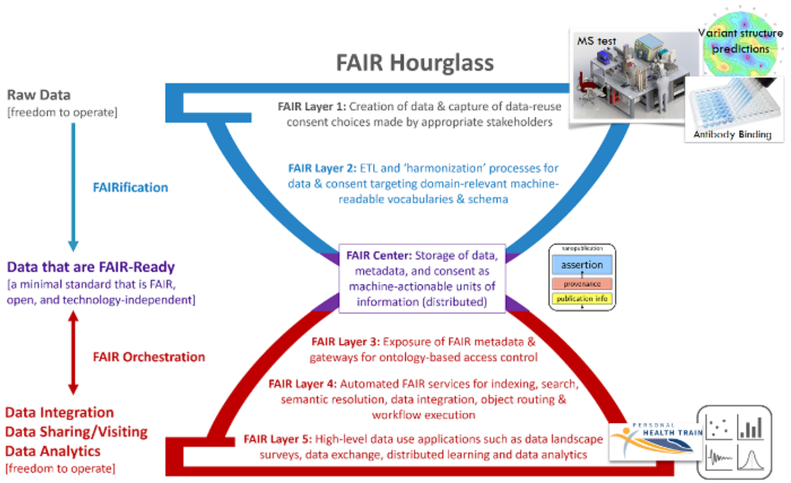
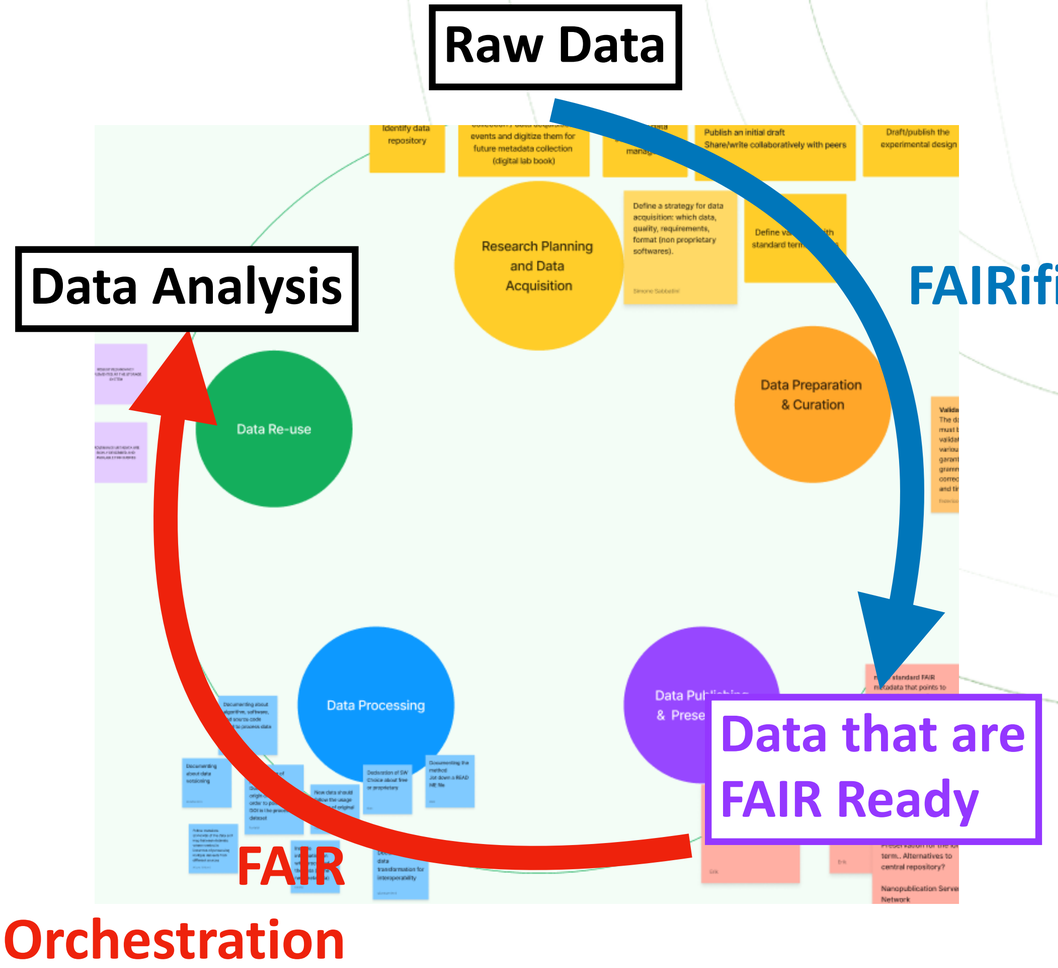


Figure 4. FAIR hourglass model consists of five operational layers that enable maximum freedom of operation to data producers (Layer 1) and data users (Layer 5) through the commitment to a minimal data standard at the center (Nanopublication). The automated workflows that support data conversion to nanopublications described in Layer 2 are the "FAIRification" processes (blue). On the bottom (red), automated workflows that enable data access, integration, and analytics described in layers 3-5 are "FAIR Orchestration".



Machine-Actionable DMPs

15:20-15:40
(Barbara)

Resume 16:00 CET

Try out services and discuss

16:00-16:45
(Interactive)

Wrap up

16:55-17:00

Homework

- **5%:** https://eoscsecretariat.eu/sites/default/files/invest_5_of_research_funds_in_ensuring_data_are_reusable.pdf
- **10 simple rules maDMPs:** https://zenodo.org/records/1434938
- **DMP templates for FAIR:** https://content.iospress.com/articles/fair-connect/fc221508

FAIR Well!

THANKS!

Erik Schultes, Barbara Magagna, Alessa Gambardella
5 December 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani