

discoverability and web-enabled research

kaitlin thaney

@kaythaney ; @mozillascience

#scholarAfrica, nairobi / 10 march 2014

(0)

mozilla

doing good is part of our code



knight-mozilla

OpenNews



Mark Surman

Executive Director, Mozilla Foundation

 Grab The Viking By The Horns <small>MoFo Insider Awards</small> share it!	 Chaotic Ball of Energy <small>MoFo Insider Awards</small> share it!
 Science Journalist <small>National Writing Project</small> share it!	 Javascript Expert <small>Peer 2 Peer University</small> share it!
 Video Editor <small>National Writing Project</small> share it!	 Accessibility Expert <small>Peer 2 Peer University</small> share it!



mozilla

Science Lab

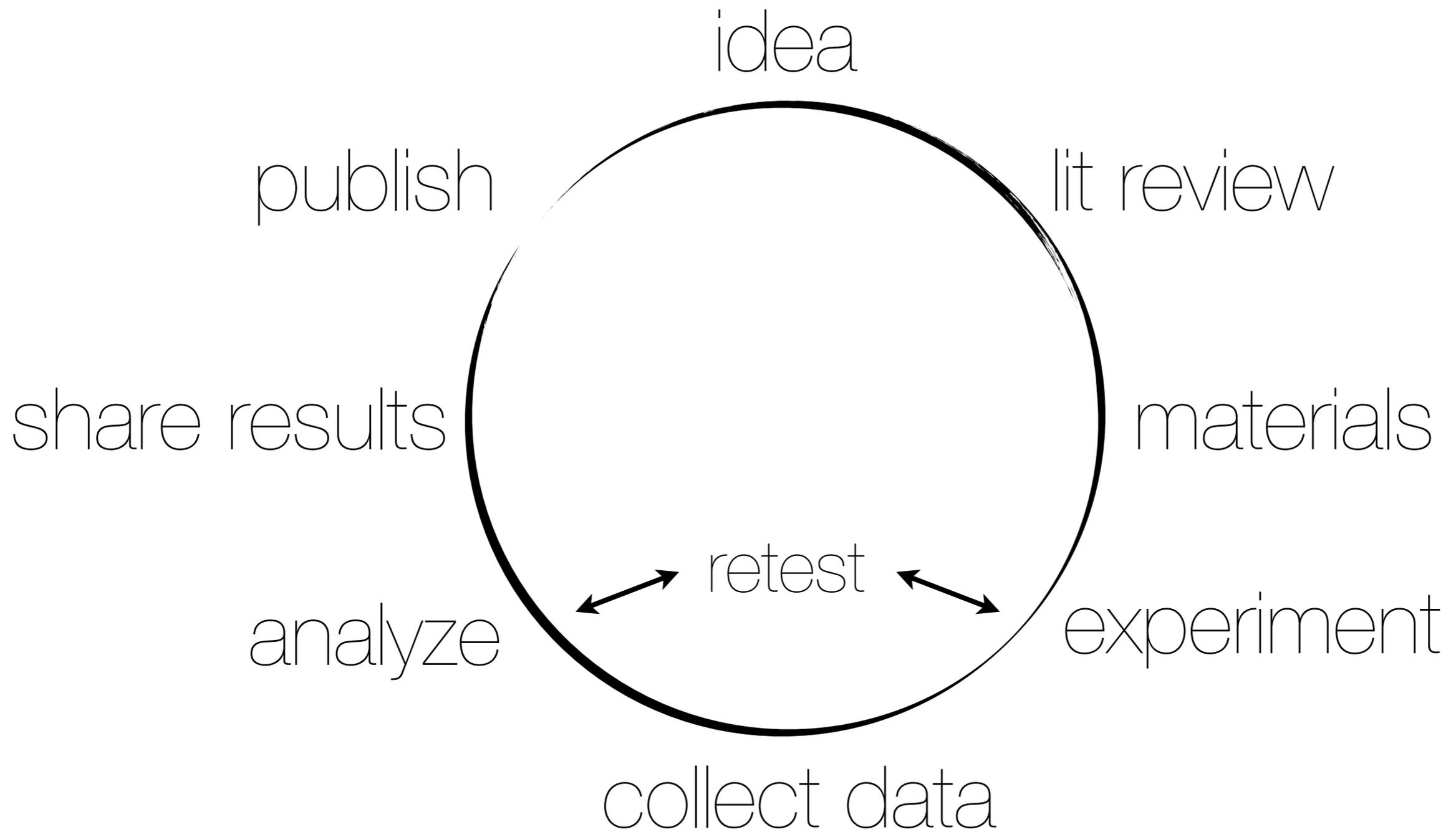
help researchers use the
power of the open web to
change science's future.



discoverability
what the web offers
role of data, sharing
challenges around “open”

(1)

research cycle



types of information

(added complexity)

hypothesis/query

articles

proceedings

negative results

analysis
code

datasets
models

prof activities
mentorship
teaching activities

content

non-digital "stuff"

protocols
parameters

blocking points

(to name a few ...)

idea

publish

access

share results

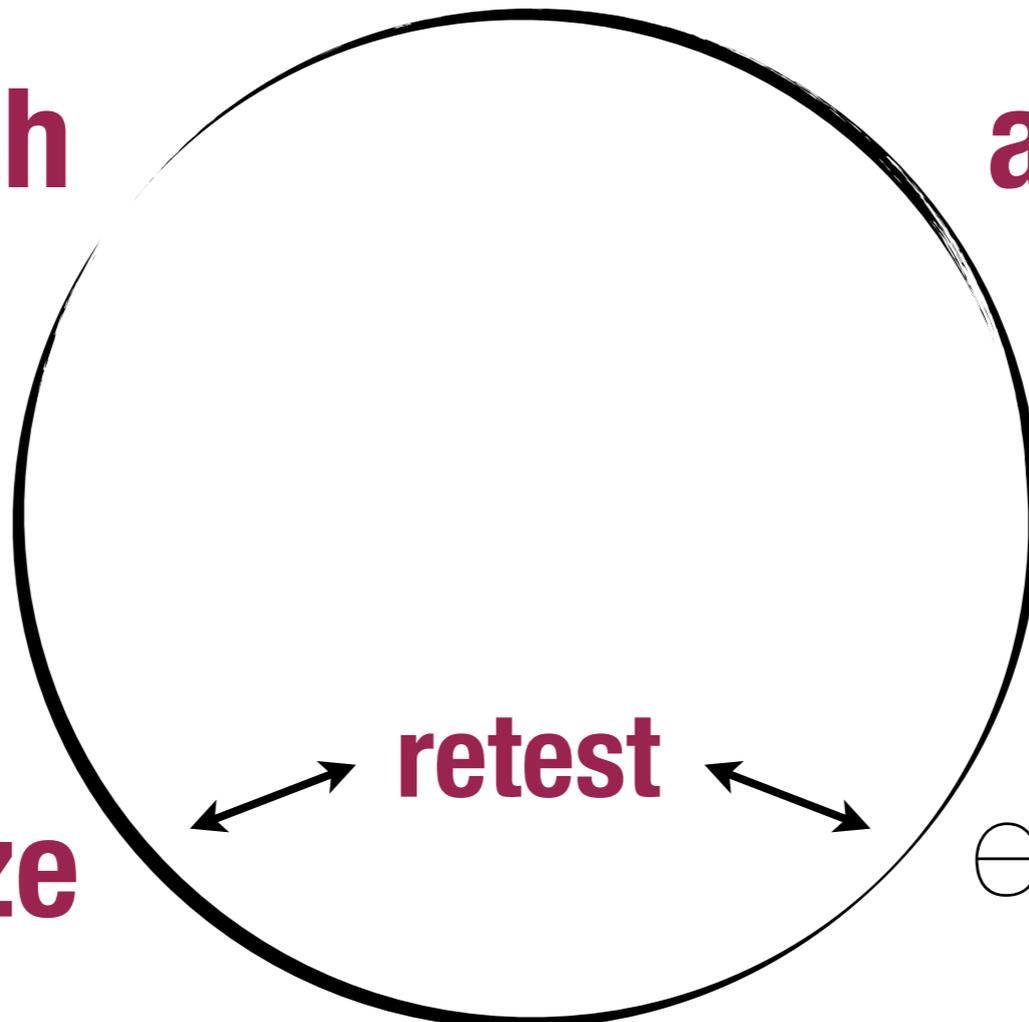
**attaining
materials**

analyze

retest

experiment

collect data



discoverability:

ability to access

(but also to reuse,

disseminate)



technical needs

metadata standards

discoverability services

dissemination platforms

APIs, tagging

rewards, incentives,
reputation



(2)

leveraging the power of
the web for scholarship

“web-enabled science”

- access to content, data, code, materials.
- emergence of “web-native” tools.
- rewards for openness, interoperability, collaboration, sharing.
- push for ROI, reuse, recomputability, transparency.



wasted . . .

\$\$\$

time

opportunity



the web as a platform

communication

access, reuse, scale

community-building



early forms of knowledge sharing

PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD

Vol I.

For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little with-
out Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

The resolution of Cubick equations out of Dr Wallis
in his Meditation before mathematicus computat
suppose $x = y a + z$. $y^3 x^3 = y^3 a^3 + 3 a^2 y z + 3 a y^2 z + z^3$. that is making $a^3 + z^3 = q$
or $x^3 = + 3 a^2 x y + a^3 + z^3$. $y^3 x^3 = y^3 a^3 + 3 a^2 y z + 3 a y^2 z + z^3$.
 ~~$y^3 x^3 = y^3 a^3 + 3 a^2 y z + 3 a y^2 z + z^3$~~
again suppose $y = a - z$. $y^3 x^3 = a^3 - 3 a^2 z + 3 a z^2 - z^3$.
+ is making $a^3 - z^3 = 8 q$. & $3 a z = p$. $y^3 x^3 = - p x + 8 q$.
Then in the first of these $p = 3 a z$. or $\frac{p}{3} = a z$.
or $\frac{p^3}{27} = a^3 z^3 = q - z^3$. Therefore $z^6 = q z^3 - \frac{p^3}{27}$.
 $z^3 = \frac{1}{2} q \pm \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}$. & by $y^3 x^3 = - p x + 8 q$ $a^3 = \frac{1}{2} q \pm \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}$
where y^3 irrationall quantities have divers signes otherwise
 $a^3 + z^3 = q$ would be false. Soe that
 $x = y a + z = \sqrt[3]{\frac{1}{2} q \pm \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}} + \sqrt[3]{\frac{1}{2} q \mp \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}}$.
is a rule for resolving y^3 equation $x^3 + p x + q = 0$. wh
it hath but one roote y^3 is when it may be generated
according to the supposition $x = y a + z$. &c. By y^3
same reason $x^3 + p x + q$ may be resolved by the
rule $x = a - z = \sqrt[3]{\frac{1}{2} q \pm \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}} - \sqrt[3]{\frac{1}{2} q \mp \sqrt{\frac{1}{4} q^2 - \frac{p^3}{27}}}$.

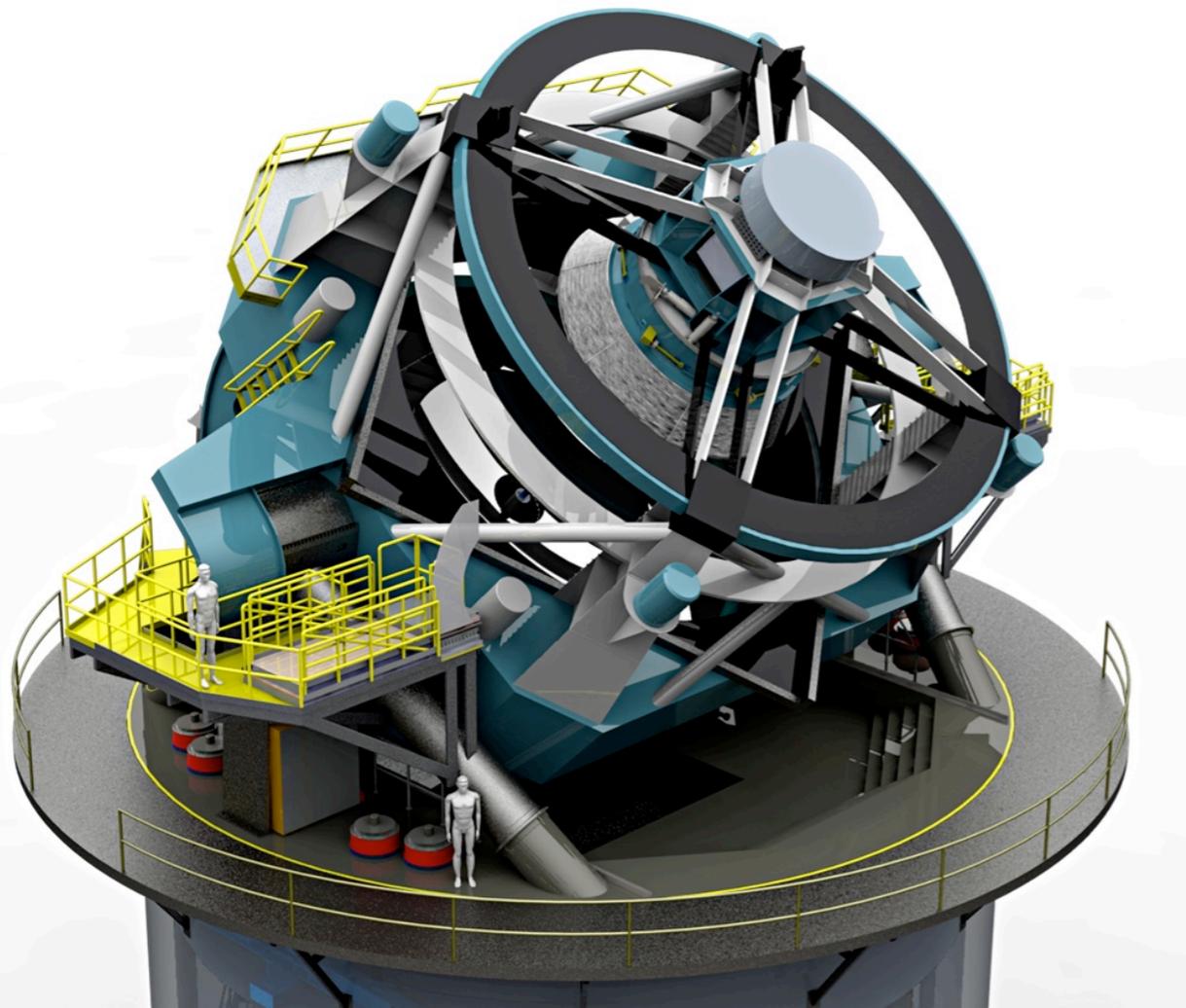
But here observe y^3 Dr Wallis would argue
 y^3 since in the first of these two cases (sometimes
viz when y^3 equation hath 3 roots) y^3 rule fail
as if it were impossible for y^3 equation to have roots
when y^3 it hath, therefore y^3 fault is in algebra
& therefore when y^3 analysis leads us to an
impossibility wee ought not to conclude y^3 thing impos-
sible, untill we have tryed all y^3 ways y^3 may be.
But let our answer y^3 fault is not in y^3 analysis
in this example, but his operation. for when y^3 z
in this example, but his operation. for when y^3 z
quation $x^3 + p x + q = 0$ hath 3 roots we suppose it to
have but one roote viz $x = y a + z$. but since y^3 z
tion cannot be then generated according to y^3 supposition
is impossible it should be resolved by it.

emergence of new
communities, practice



broader access,
engagement,
reach

THE HUMAN ENCYCLOPAEDIA



there's still friction in the
system.

and duplication out of frustration.

“traditions last not because they are excellent, but because influential people are **averse** to change and because of the sheer burdens of **transition to a better state ...**”

Cass Sunstein

our systems need to
talk to one another.

both human and machine, globally.

(3)

thinking beyond the
PDF.

role of data, code, methods.

3 GB/year

9 GB/PhD

3 GB/year

30,000,000 GB/all*

9 GB/PhD

3 GB/year

* roughly

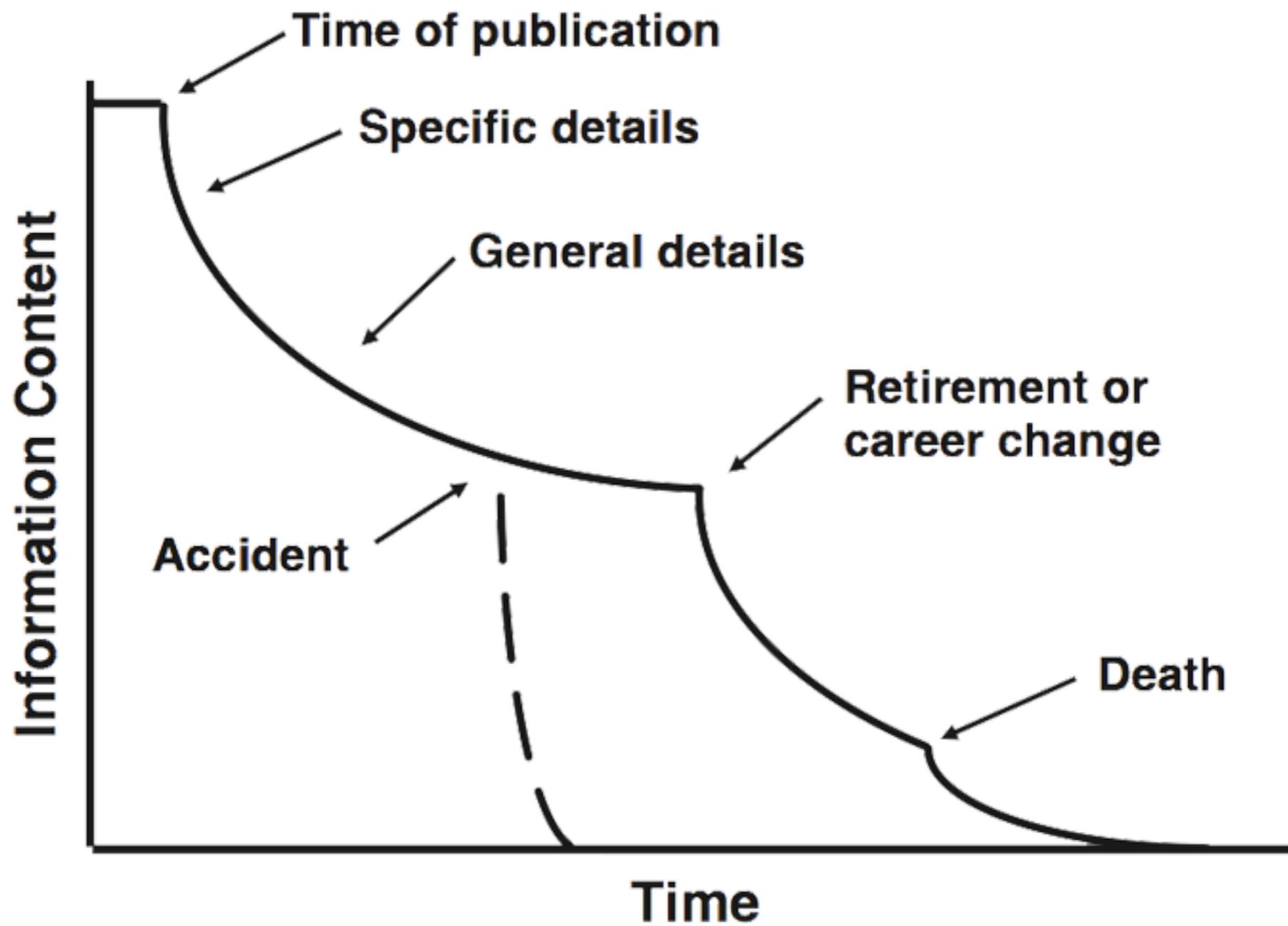
how much made available?

30,000,000 GB/all*

9 GB/PhD

3 GB/year

* roughly



```
# Data inspection
```

```
The codes for variable site are PC=Potter Cove AB= Admiral Bay
```

```
```{r dataInspection, echo=FALSE, message=FALSE, warning=FALSE}
```

```
chla <- read.table("Meteo_stat.txt", header=T)
```

```
pairs(na.omit(chla[,2:6]), panel=function(x,y) { points(x,y); lines(lowess(x,y))})
```

```
#source("panelutils.R")
```

```
#pairs(chlaR, panel=panel.smooth, diag.panel=panel.hist, main="")
```

```
#hist(chlaR$ClorMAX, col="bisque", right=FALSE)
```

```
#hist(sqrt(chlaR$ClorMAX), col="bisque", right=FALSE)
```

```
#hist(chlaR$DegDay_NM, col="bisque", right=FALSE)
```

```
#hist(chlaR$ENSO_NM, col="bisque", right=FALSE)
```

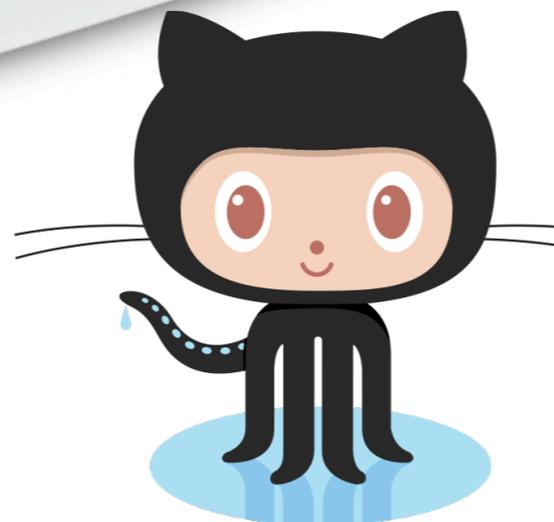
```
#hist(chlaR$SAM_NM, col="bisque", right=FALSE)
```



code as a research object



what's needed  
to reuse ?



OPEN ACCESS PEER-REVIEWED

31,870

VIEWS

81

CITATIONS

289

ACADEMIC BOOKMARKS

66

SOCIAL SHARES

RESEARCH ARTICLE | FEATURED IN PLOS COLLECTIONS

# Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar, Roger S. Day, Douglas B. Fridsma

Article

About the Authors

Metrics

Comments

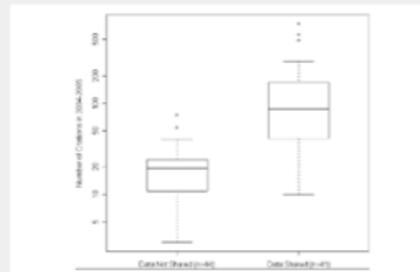
Related Content

Download

Print

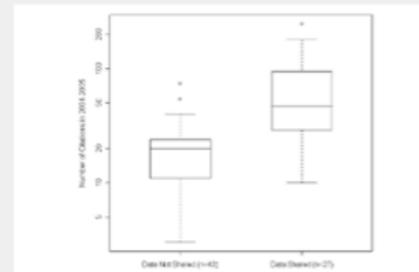
Share

	Total	Data Shared	Data Not Shared	95% CI	95% CI
Total	46	41 (89%)	5 (11%)		
Open Access	32	27 (84%)	5 (16%)		
Peer-Reviewed	14	14 (100%)	0 (0%)		
Published 2013-2015	75	68 (91%)	7 (9%)		
US Authors	36	32 (89%)	4 (11%)		
All US Authors	26	23 (88%)	3 (12%)		



	Percent increase in citation count (95% confidence interval)	p-value
Publish in a journal with twice the impact factor	84% (59 to 109%)	<0.001
Increase the publication date by a month	-3% (-5 to -2%)	<0.001
Include a US author	38% (1 to 89%)	0.049
<b>Make data publicly available</b>	<b>69% (18 to 143%)</b>	<b>0.006</b>

We calculated a multivariate linear regression over the citation counts, including covariates for journal impact factor, date of publication, US authorship, and data availability. The coefficients and p-values for each of the covariates are shown here, representing the contribution of other covariates to the citation count, independent of other covariates.  
doi:10.1371/journal.pone.0000308.t002



Hide Figures

Included in the Following Collection

Open Access Collection

Comments

referencing raw data  
Posted by jcb Bradley

Source: Piwowar, et al. PLOS.



[日本語要約](#)

## Credit for code

*Nature Genetics* **46**, 1 (2014) | doi:10.1038/ng.2869

Published online 27 December 2013



PDF



Citation



Reprints



Rights & permissions



Article metrics

**Moving toward fully transparent research publications, we suggest several approaches to share research that is instantiated in software written for computers and other laboratory machines. Review, replication, reuse and recognition are all incentives to provide code.**

Software for biomedical research ranges from single scripts used to format data to complex suites of analytical tools. The biggest problem often encountered by editors, referees and readers is



# Soil Diversity and Hydration as Observed by ChemCam at Gale Crater, Mars

P.-Y. Meslin,<sup>1,2\*</sup> O. Gasnault,<sup>1,2</sup> O. Forni,<sup>1,2</sup> S. Schröder,<sup>1,2</sup> A. Cousin,<sup>3</sup> G. Berger,<sup>1,2</sup> S. M. Clegg,<sup>3</sup> J. Lasue,<sup>1,2</sup> S. Maurice,<sup>1,2</sup> V. Sautter,<sup>4</sup> S. Le Mouélic,<sup>5</sup> R. C. Wiens,<sup>3</sup> C. Fabre,<sup>6</sup> W. Goetz,<sup>7</sup> D. Bish,<sup>8</sup> N. Mangold,<sup>5</sup> B. Ehlmann,<sup>9,10</sup> N. Lanza,<sup>3</sup> A.-M. Harri,<sup>11</sup> R. Anderson,<sup>12</sup> E. Rampe,<sup>13</sup> T. H. McConnochie,<sup>14</sup> P. Pinet,<sup>1,2</sup> D. Blaney,<sup>10</sup> R. Lèveillé,<sup>15</sup> D. Archer,<sup>13</sup> B. Barraclough,<sup>16</sup> S. Bender,<sup>16</sup> D. Blake,<sup>17</sup> J. G. Blank,<sup>17</sup> N. Bridges,<sup>18</sup> B. C. Clark,<sup>19</sup> L. DeFlores,<sup>10</sup> D. Delapp,<sup>3</sup> G. Dromart,<sup>20</sup> M. D. Dyar,<sup>21</sup> M. Fisk,<sup>22</sup> B. Gondet,<sup>23</sup> J. Grotzinger,<sup>9</sup> K. Herkenhoff,<sup>12</sup> J. Johnson,<sup>18</sup> J.-L. Lacour,<sup>24</sup> Y. Langevin,<sup>23</sup> L. Leshin,<sup>25</sup> E. Lewin,<sup>26</sup> M. B. Madsen,<sup>27</sup> N. Melikechi,<sup>28</sup> A. Mezzacappa,<sup>28</sup> M. A. Mischna,<sup>10</sup> J. E. Moores,<sup>29</sup> H. Newsom,<sup>30</sup> A. Ollila,<sup>30</sup> R. Perez,<sup>31</sup> N. Renno,<sup>32</sup> J.-B. Sirven,<sup>24</sup> R. Tokar,<sup>16</sup> M. de la Torre,<sup>9</sup> L. d'Uston,<sup>1,2</sup> D. Vaniman,<sup>16</sup> A. Yingst,<sup>16</sup> MSL Science Team†

The ChemCam instrument, which provides insight into martian soil chemistry at the submillimeter scale, identified two principal soil types along the Curiosity rover traverse: a fine-grained mafic type and a locally derived, coarse-grained felsic type. The mafic soil component is representative of widespread martian soils and is similar in composition to the martian dust. It possesses a ubiquitous hydrogen signature in ChemCam spectra, corresponding to the hydration of the amorphous phases found in the soil by the CheMin instrument. This hydration likely accounts for an important fraction of the global hydration of the surface seen by previous orbital measurements. ChemCam analyses did not reveal any significant exchange of water vapor between the regolith and the atmosphere. These observations provide constraints on the nature of the amorphous phases and their hydration.

The composition, mineralogy, and volatile inventory of the martian soil constitute an global and usually unrelated to bedrock exposures (6, 15). Therefore, the nature and origin of this

Viking and Phoenix landers (~100-mg samples were analyzed by the Viking Molecular Analysis Experiment) (27, 28). This higher resolution is crucial to unraveling the chemical and physical processes that formed the martian soil. Understanding the soil fine-scale chemistry, including its hydration, is an important objective of the ChemCam instrument onboard the Curiosity rover.

The laser-induced breakdown spectrometer (LIBS) on the ChemCam instrument (29, 30) provides insight on martian soil and dust chemical variability at the submillimeter scale. The small sampling area of the ChemCam laser (~350 to 550  $\mu\text{m}$  depending on distance) allows it to isolate various soil components and identify mixing trends that bulk measurements might average together. The spectroscopic measurement of each individual soil, or “LIBS point” (31), is typically obtained from a series of 30 to 50 laser shots. Because each shot produces a LIBS spectrum of a deeper portion of the soil than the previous shot, it is possible to retrieve a profile of chemical composition to depths of a few millimeters in soils and a few tens of micrometers in rocks. The uncertainty budget of the LIBS measurements is small enough for such types of analyses to be performed. This yields analyses deeper than the probing depth of thermal, near-infrared, and x-ray spectrometers but shallower than GRS

<sup>1</sup>Université de Toulouse, UPS-OMP, IRAP, 31028 Toulouse, France. <sup>2</sup>CNRS, IRAP, 9 Av. Colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France. <sup>3</sup>Los Alamos National Laboratory,



### Soil Diversity and Hydration as Observed by ChemCam at Gale Crater, Mars

P.-Y. Meslin,<sup>1,2\*</sup> O. Gasnaut,<sup>1,2</sup> O. Forni,<sup>1,2</sup> S. Schröder,<sup>1,2</sup> A. Cousin,<sup>3</sup> G. Berger,<sup>1,2</sup> S. M. Clegg,<sup>3</sup> J. Lasue,<sup>1,2</sup> S. Maurice,<sup>1,2</sup> V. Sautter,<sup>4</sup> S. Le Mouélic,<sup>5</sup> R. C. Wiens,<sup>3</sup> C. Fabre,<sup>6</sup> W. Goetz,<sup>7</sup> D. Bish,<sup>8</sup> N. Mangold,<sup>5</sup> B. Ehlmann,<sup>9,10</sup> N. Lanza,<sup>3</sup> A.-M. Harri,<sup>11</sup> R. Anderson,<sup>12</sup> E. Rampe,<sup>13</sup> T. H. McConnochie,<sup>14</sup> P. Pinet,<sup>1,2</sup> B. Blaney,<sup>10</sup> R. Lèveillé,<sup>15</sup> D. Archer,<sup>13</sup> B. Barraclough,<sup>16</sup> S. Bender,<sup>16</sup> D. Blake,<sup>17</sup> J. G. Blank,<sup>18</sup> N. Bridges,<sup>18</sup> B. C. Clark,<sup>19</sup> L. DeFlores,<sup>10</sup> D. Delapp,<sup>3</sup> G. Dromart,<sup>20</sup> M. D. Dyar,<sup>21</sup> M. Fisk,<sup>22</sup> B. Gondet,<sup>23</sup> J. Grotzinger,<sup>9</sup> K. Herkenhoff,<sup>12</sup> J. Johnson,<sup>18</sup> J.-L. Lacour,<sup>24</sup> Y. Langevin,<sup>23</sup> L. Leshin,<sup>25</sup> E. Lewin,<sup>26</sup> M. B. Madsen,<sup>27</sup> N. Melikechi,<sup>28</sup> A. Mezzacappa,<sup>28</sup> M. A. Mischna,<sup>10</sup> J. E. Moores,<sup>29</sup> H. Newsom,<sup>30</sup> A. Ollila,<sup>30</sup> R. Perez,<sup>31</sup> N. Renno,<sup>32</sup> J.-B. Sirven,<sup>24</sup> R. Tokar,<sup>16</sup> M. de la Torre,<sup>9</sup> L. d'Uston,<sup>1,2</sup> D. Vaniman,<sup>16</sup> A. Yingst,<sup>16</sup> MSL Science Team†

The ChemCam instrument, which provides insight into martian soil chemistry at the submillimeter scale, identified two principal soil types along the Curiosity rover traverse: a fine-grained mafic type and a locally derived, coarse-grained felsic type. The mafic soil component is representative of widespread martian soils and is similar in composition to the martian dust. It possesses a ubiquitous hydrogen signature in ChemCam spectra, corresponding to the hydration of the amorphous phases found in the soil by the CheMin instrument. This hydration likely accounts for an important fraction of the global hydration of the surface seen by previous orbital measurements. ChemCam analyses did not reveal any significant exchange of water vapor between the regolith and the atmosphere. These observations provide constraints on the nature of the amorphous phases and their hydration.

Viking and Phoenix landers (~100-mg samples were analyzed by the Viking Molecular Analysis Experiment) (27, 28). This higher resolution is crucial to unraveling the chemical and physical processes that formed the martian soil. Understanding the soil fine-scale chemistry, including its hydration, is an important objective of the ChemCam instrument onboard the Curiosity rover.

The laser-induced breakdown spectrometer (LIBS) on the ChemCam instrument (29, 30) provides insight on martian soil and dust chemical variability at the submillimeter scale. The small sampling area of the ChemCam laser (~350 to 550 μm depending on distance) allows it to isolate various soil components and identify mixing trends that bulk measurements might average together. The spectroscopic measurement of each individual soil, or “LIBS point” (31), is typically obtained from a series of 30 to 50 laser shots. Because each shot produces a LIBS spectrum of a deeper portion of the soil than the previous shot, it is possible to retrieve a profile of chemical composition to depths of a few millimeters in soils and a few tens of micrometers in rocks. The uncertainty budget of the LIBS measurements is small enough for such types of analyses to be performed. This yields analyses deeper than the probing depth of thermal, near-infrared, and x-ray spectrometers but shallower than GRS

The composition, mineralogy, and volatile inventory of the martian soil constitute an global and usually unrelated to bedrock exposures (6, 15). Therefore, the nature and origin of this

<sup>1</sup>Université de Toulouse, UPS-OMP, IRAP, 31028 Toulouse, France. <sup>2</sup>CNRS, IRAP, 9 Av. Colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France. <sup>3</sup>Los Alamos National Laboratory,

(4)

is open enough?

**necessary but not sufficient.**



**access versus deluge**

how to reconcile?

**our systems need to talk to one another.**



# we also need to build capacity.

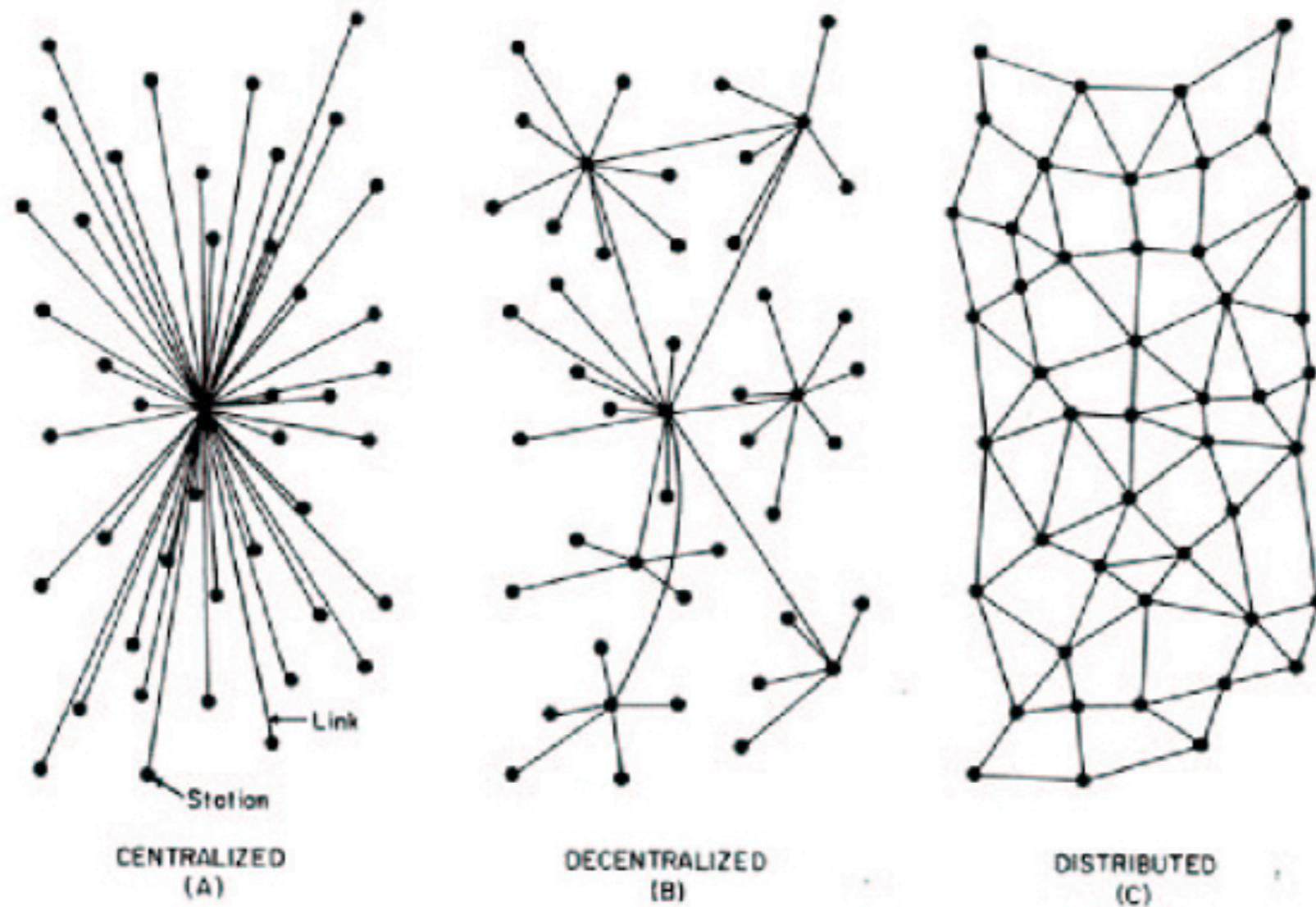


FIG. 1 — Centralized, Decentralized and Distributed Networks

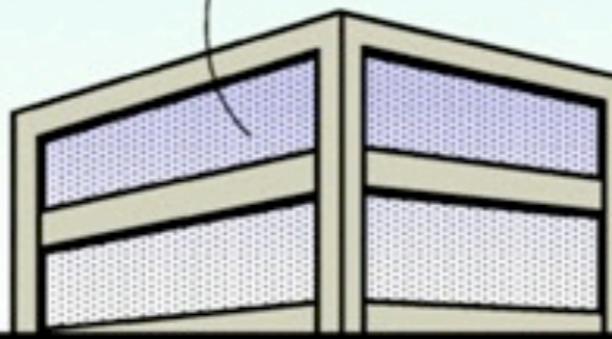
# digital literacy

WE ADDED A NEW PERFORMANCE TEST, BUT LEARNED THAT THE TEST ITSELF IS FLAWED.



Dilbert.com DilbertCartoonist@gmail.com

NOW OUR PRODUCT FAILS OUR OWN TESTS AND OUR CUSTOMERS ARE ASKING TO SEE THE TEST RESULTS.



8-11-10 © 2010 Scott Adams, Inc./Dist. by UFS, Inc.

DO I HAVE PERMISSION TO FAKE THE TEST DATA?



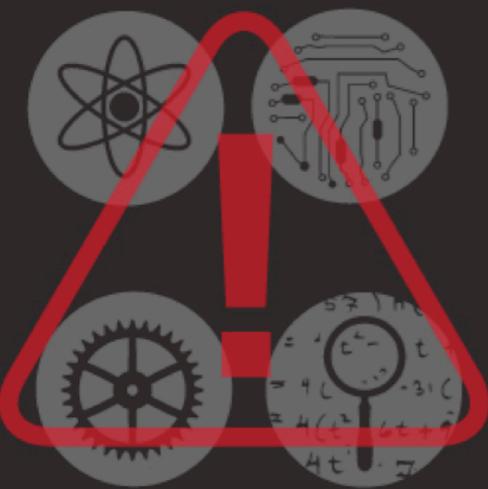
I DIDN'T EVEN KNOW DATA CAN BE REAL.



# Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

## Computational Thinking: A Digital Age Skill for Everyone



### The 21st Century **WORKFORCE:** Skills Gap & The STEM **DILEMMA**

The gap between the skills available in the current workforce and those needed for many 21st century jobs creates a serious challenge



of employers believe new graduates in the workforce are adequately prepared by their colleges or other pre-employment training programs

A workforce prepared to tackle science, technology, engineering and math (STEM) is critical to driving future growth and innovation

WHILE  
ONLY  
**4%**

OF THE NATION'S  
WORKFORCE IS COMPOSED  
OF SCIENTISTS AND  
ENGINEERS, THIS GROUP  
DISPROPORTIONATELY  
CREATES JOBS FOR

THE OTHER  
**96%**



“Reliance on  
ad-hoc, self-  
education  
about what’s  
possible  
doesn’t scale.”

- Selena Decklemann



education as a means of  
building community

**... globally, as well as across disciplines.**

# current activity:

129 instructors

(60+, training)

109 bootcamps

3700+ learners



# R OpenSci

StilettoFiend / OpenScienceTraining

Course materials from the Open Science Training Initiative

28 commits

1 branch

0 releases

1 contributor

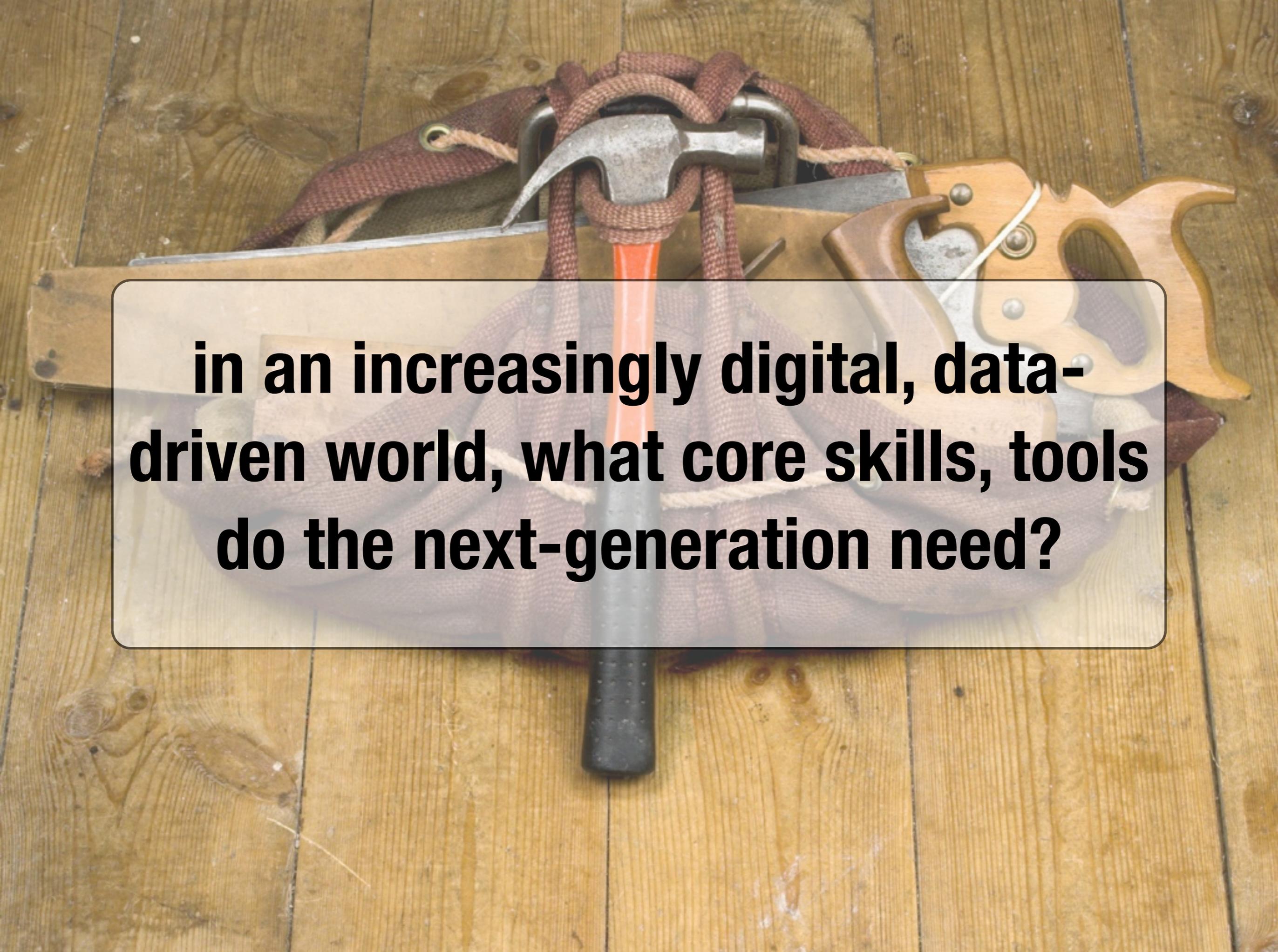
branch: master OpenScienceTraining

Update README.md		
 <b>StilettoFiend</b> authored 3 months ago		latest commit 8c8
 <a href="#">LegoSessions</a>	Slide decks from #solo13lego	3 m
 <a href="#">LICENCE.txt</a>	Rename cc-by-licence to LICENCE.txt	10 m
 <a href="#">Lecture1A-OpenAndRepro-B...</a>	Lecture 1 - beta version of slides added in ODP and PPTX formats	4 m
 <a href="#">Lecture1A-OpenAndRepro-B...</a>	Files added for Lego-based sessions	3 m
 <a href="#">Lecture1B-ExampleRotation...</a>	Lecture 1B: Example handout to accompany RBL assessment	5 m
 <a href="#">Lecture1B-ExampleRotation...</a>	Lecture 1B: Example handout to accompany RBL assessment	5 m
 <a href="#">Lecture1B-RotationOutline.odp</a>	Lecture 1B: ODP, PPT slides for rotation-based learning	5 m
 <a href="#">Lecture1B-RotationOutline.p...</a>	Lecture 1B: ODP, PPT slides for rotation-based learning	5 m
 <a href="#">Lecture2-Licensing.odp</a>	Lecture 2 - slides added in ODP and PPTX formats	5 m
 <a href="#">Lecture2-Licensing.pptx</a>	Lecture 2 - slides added in ODP and PPTX formats	5 m



## Open Science: An Introduction

How open access, data, and research are changing the way we share scientific

A collection of tools including a hammer, a saw, and a wrench, along with a red tool bag, all resting on a wooden surface. The hammer has a red handle and a silver head. The saw has a wooden handle and a metal blade. The wrench is made of wood and metal. The red tool bag is made of canvas and has a metal clasp. The tools are arranged on a wooden surface made of vertical planks.

**in an increasingly digital, data-driven world, what core skills, tools do the next-generation need?**

(5)

shifting practice  
(and getting it to stick)  
is challenging.

**and we're here to help.**

# **what are the necessary components?**

tools and technology

cultural awareness, best practice

connections, open dialogue

skills training

coordination and  
collaboration are key.

design for interoperability.

remember the  
non-technical challenges.

operating in isolation  
doesn't scale.

help shape the  
conversation.

**teach, contribute, learn.**

<http://software-carpentry.org>

<http://mozillascience.org>

questions?

kaitlin@mozillafoundation.org

@kaythane ; @mozillascience