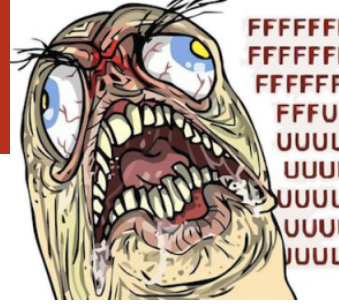
Reproducible Research: Where Do We Stand?

Arnaud Legrand
CNRS, Inria, University of Grenoble

March 22, 2017 – Data Science Institute, Grenoble

Outline

Frustration



As an Author

- Advisor: "Did you take care of setting this?" Me: "Uh?"
- I thought I used the same parameters but I'm getting different results! I swear it worked yesterday!
- A new student wants to compare with the method I proposed last year
- The damned fourth reviewer asked for a major revision and wants me to change figure 3. ☹ Which code and which data set did I use to generate this figure?
- 6 months later: Why did I do that?

As a Reviewer

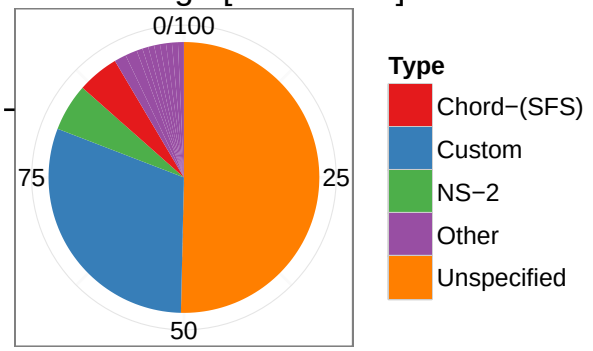
This may be an interesting contribution but:

- There is no label/legend/... What is the meaning of this graph? If only I could access the generation script and get rid of the logscale
- This average value must hide something. As usual, no confidence interval... I wonder whether the difference is significant at all
- That can't be true, I'm sure they removed some points or decided to show only a subset of the data. I wonder what the rest looks like
- It this improvement solely the result of this naive idea?

A Few Edifying Examples in my Domain

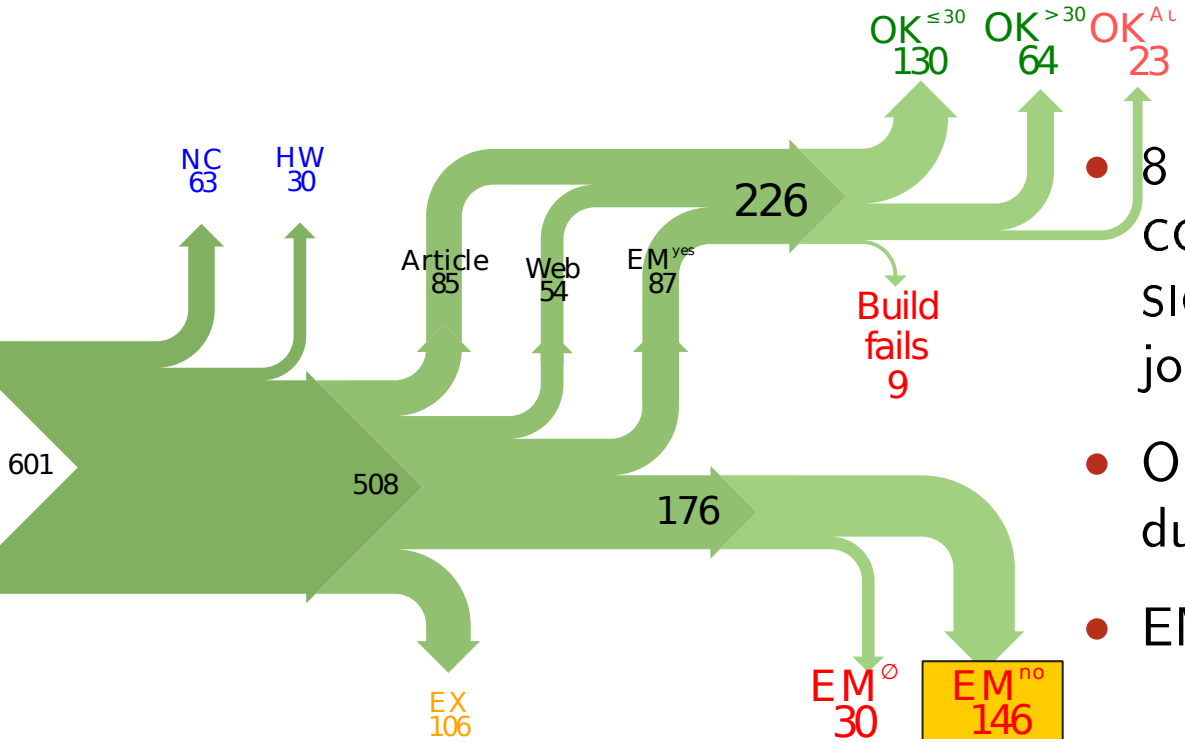
Naicken, Stephen *et Al.*, *Towards Yet Another Peer-to-Peer Simulator*, HET-NETs'06.

Simulator usage [Naicken06]



From 141 P2P sim.papers, 30% use a custom tool, 50% don't report used tool

Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/ 2014,2015



- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- Original study = 80% of non reproducible work
- EM^{no} = the code cannot be provided

The Dog Ate my Homework !!!

- Versioning Problems

Thanks for your interest in the implementation of our paper. The good news is that I was able to find some code. I am just hoping that it is a stable working version of the code, and matches the implementation we finally used for the paper. Unfortunately, I have lost some data when my laptop was stolen last year. The bad news is that the code is not commented and/or clean.

Attached is the ⟨system⟩ source code of our algorithm. I'm not very sure whether it is the final version of the code used in our paper, but it should be at least 99% close. Hope it will help.

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices

Unfortunately, the server in which my implementation was stored had a disk crash in April and three disks crashed simultaneously. While the help desk made significant effort to save the data, my entire implementation for this paper was not found.

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon

Unfortunately the current system is not mature enough at the moment, so it's not yet publicly available. We are actively working on a number of extensions and things are somewhat volatile. However, once things stabilize we plan to release it to outside users. At that point, we would be happy to send you a copy.

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

I am afraid that the source code was never released. The code was never intended to be released so is not in any shape for general use.

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release
- Programmer Left

> *⟨STUDENT⟩ was a graduate student in our program but he left a while back so I am responding instead. For the paper we used a prototype that included many moving pieces that only ⟨STUDENT⟩ knew how to operate and we did not have the time to integrate them in a ready-to-share implementation before he left. Still, I hope you can build on the ideas/technique of the paper.*
>
> *Unfortunately, the author who has done most of the coding for this paper has passed away and the code is no longer maintained.*

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release
- Programmer Left
- Commercial Code

Since this work has been done at ⟨COMPANY⟩ we don't open-source code unless there is a compelling business reason to do so. So unfortunately I don't think we'll be able to share it with you.

The code owned by ⟨COMPANY⟩, and AFAIK the code is not open-source. Your best bet is to reimplement :(Sorry.

The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release
- Programmer Left
- Commercial Code
- Proprietary Academic Code

Unfortunately, the ⟨SYSTEM⟩ sources are not meant to be opensource (the code is partially property of ⟨UNIVERSITY 1⟩, ⟨UNIVERSITY 2⟩ and ⟨UNIVERSITY 3⟩.)

If this will change I will let you know, albeit I do not think there is an intention to make the ⟨SYSTEM⟩ sources opensource in the near future.

If you're interested in obtaining the code, we only ask for a description of the research project that the code will be used in (which may lead to some joint research), and we also have a software license agreement that the University would need to sign.

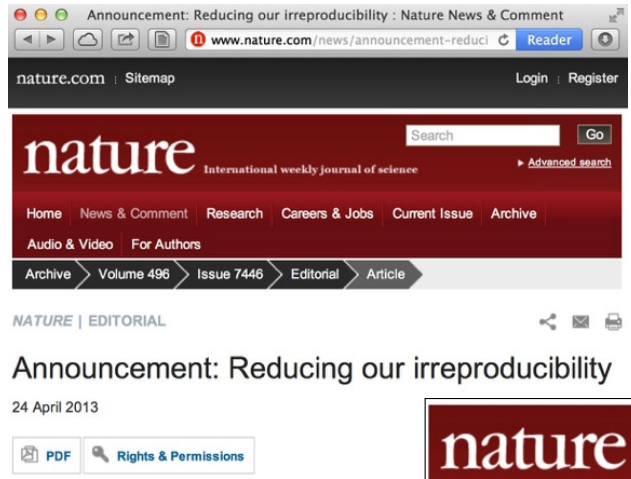
The Dog Ate my Homework !!!

- Versioning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release
- Programmer Left
- Commercial Code
- Proprietary Academic Code
- Research vs. Sharing
- ...
- ...

In the past when we attempted to share it, we found ourselves spending more time getting outsiders up to speed than on our own research. So I finally had to establish the policy that we will not provide the source code outside the group.

Public evidence for a Lack of Reproducibility

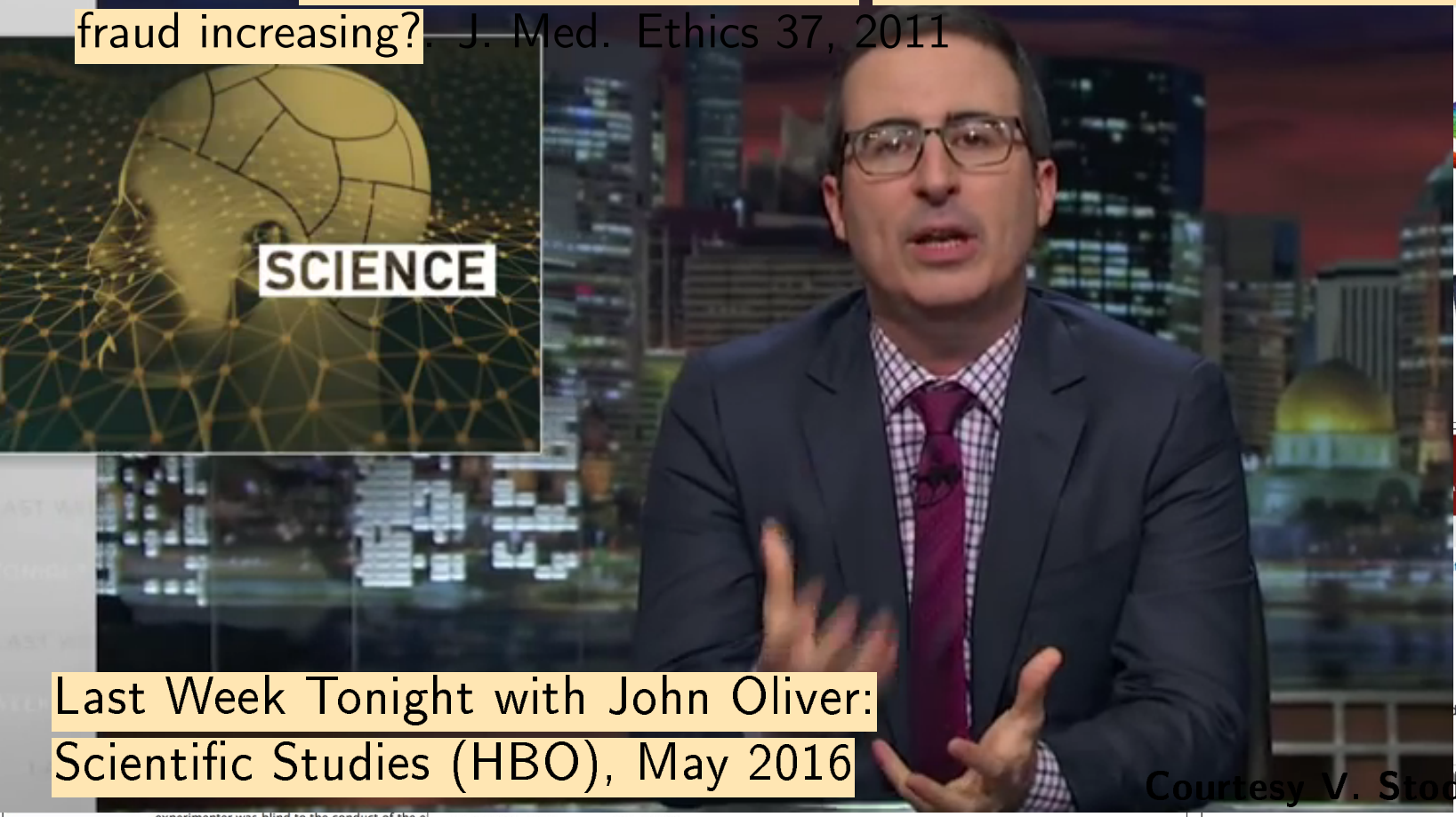
- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010
- *Reproducibility: A tragedy of errors*, Nature, Feb 2016.
- Steen RG, Retractions in the scientific literature: is the incidence of research fraud increasing?. J. Med. Ethics 37, 2011



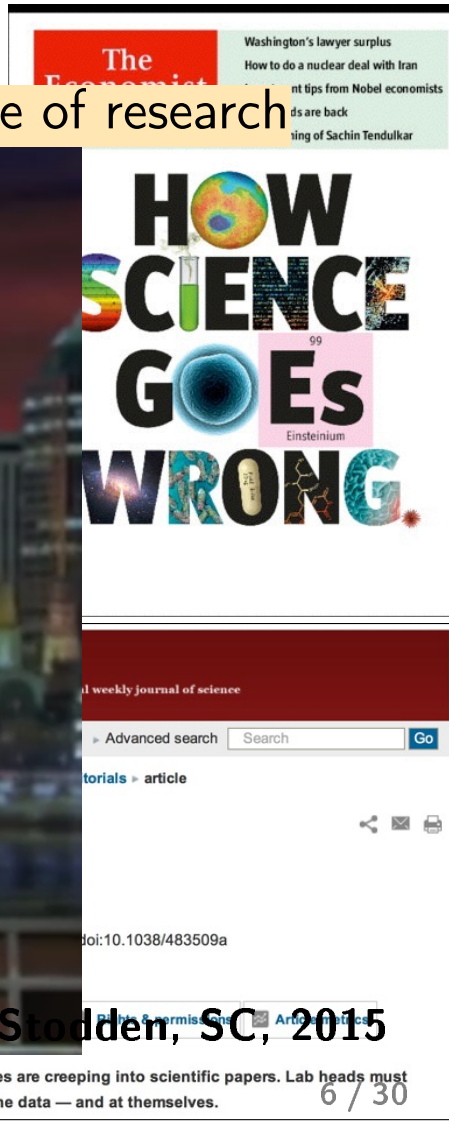
Courtesy V. Stodden, SC, 2015

Public evidence for a Lack of Reproducibility

- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010
- *Reproducibility: A tragedy of errors*, Nature, Feb 2016.
- Steen RG, Retractions in the scientific literature: is the incidence of research fraud increasing?. J. Med. Ethics 37, 2011



Last Week Tonight with John Oliver: Scientific Studies (HBO), May 2016



Courtesy V. Stodden, SC, 2015

Quick poll

1. Have you ever tried to reproduce some research results ?

Quick poll

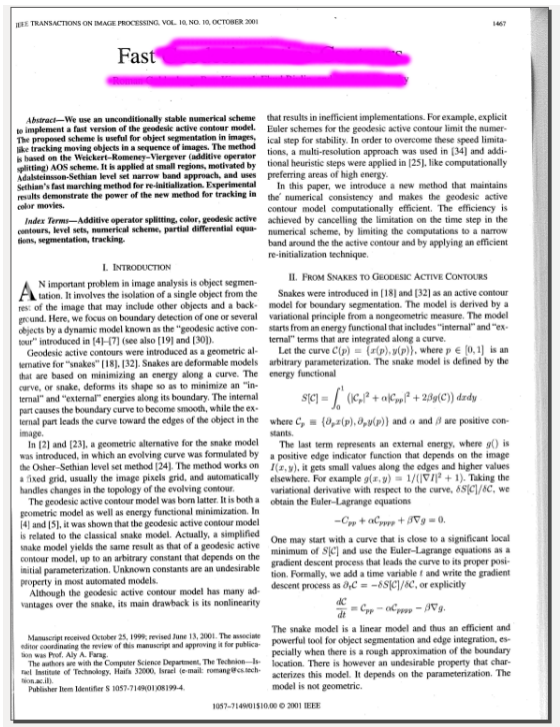
1. Have you ever tried to reproduce some research results ?
2. Have you ever failed ?

Quick poll

1. Have you ever tried to reproduce some research results ?
2. Have you ever failed ?

Article typique en traitement d'image

fichier PDF



IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 10, NO. 10, OCTOBER 2001 1467

Fast [redacted]

Abstract—**We use an unconditionally stable numerical scheme to implement a fast version of the geodesic active contour model. The proposed scheme is useful for object segmentation in images, like tracking moving objects in a sequence of images. The method is based on the Weickert–Romeney–Viergever (additive operator splitting) AOS scheme. It is applied at small regions, motivated by Adalsteinsson-Sethian level set narrow band approach, and uses Sethian's fast marching method for re-initialization. Experimental results demonstrate the power of the new method for tracking in color movies.**

Index Terms—**Additive operator splitting, color, geodesic active contours, level sets, numerical scheme, partial differential equations, segmentation, tracking.**

I. Introduction

An important problem in image analysis is object segmentation. It involves the isolation of a single object from the rest of the image that may include other objects and a background. Here, we focus on boundary detection of one or several objects by a dynamic model known as the "geodesic active contour" introduced in [4]–[7] (see also [19] and [30]).

Geodesic active contours were introduced as a geometric alternative for "snakes" [18], [32]. Snakes are deformable models that are based on minimizing an energy along a curve. The curve, or snake, deforms its shape so as to minimize an "internal" and "external" energies along its boundary. The internal part causes the boundary curve to become smooth, while the external part leads the curve toward the edges of the object in the image.

In [2] and [23], a geometric alternative for the snake model was introduced, in which an evolving curve was formulated by the Osher–Sethian level set method [24]. The method works on a fixed grid, usually the image pixels grid, and automatically handles changes in the topology of the evolving contour.

The geodesic active contour model was born latter. It is both a geometric model as well as energy functional minimization. In [4] and [5], it was shown that the geodesic active contour model is related to the classical snake model. Actually, a simplified snake model yields the same result as that of a geodesic active contour model, up to an arbitrary constant that depends on the initial parameterization. Unknown constants are an undesirable property in most automated models.

Although the geodesic active contour model has many advantages over the snake, its main drawback is its nonlinearity that results in inefficient implementations. For example, explicit Euler schemes for the geodesic active contour limit the numerical step for stability. In order to overcome these speed limitations, a multi-resolution approach was used in [34] and additional heuristic steps were applied in [25], like computationally preferring areas of high energy.

In this paper, we introduce a new method that maintains the numerical consistency and makes the geodesic active contour model computationally efficient. The efficiency is achieved by cancelling the limitation on the time step in the numerical scheme, by limiting the computations to a narrow band around the active contour and by applying an efficient re-initialization technique.

II. From Snakes to Geodesic Active Contours

Snakes were introduced in [18] and [32] as an active contour model for boundary segmentation. The model is derived by a variational principle from a nongeometric measure. The model starts from an energy functional that includes "internal" and "external" terms that are integrated along a curve.

Let the curve $\mathcal{C}(p) = \{x(p), y(p)\}$, where $p \in [0,1]$ is an arbitrary parameterization. The snake model is defined by the energy functional

$$S[\mathcal{C}] = \int_0^1 \left(|\mathcal{C}_p|^2 + \alpha|\mathcal{C}_{pp}|^2 + 2\beta g(\mathcal{C})\right) dxdy$$

where $\mathcal{C}_p \equiv \{\partial_p x(p), \partial_p y(p)\}$ and α and β are positive constants.

The last term represents an external energy, where $g()$ is a positive edge indicator function that depends on the image $I(x,y)$, it gets small values along the edges and higher values elsewhere. For example $g(x,y) = 1/(|\nabla I|^2 + 1)$. Taking the variational derivative with respect to the curve, $\delta S[\mathcal{C}]/\delta\mathcal{C}$, we obtain the Euler–Lagrange equations

$$-\mathcal{C}_{pp} + \alpha\mathcal{C}_{pppp} + \beta\nabla g = 0.$$

One may start with a curve that is close to a significant local minimum of $S[\mathcal{C}]$ and use the Euler–Lagrange equations as a gradient descent process that leads the curve to its proper position. Formally, we add a time variable t and write the gradient descent process as $\partial_t\mathcal{C} = -\delta S[\mathcal{C}]/\delta\mathcal{C}$, or explicitly

$$\frac{d\mathcal{C}}{dt} = \mathcal{C}_{pp} - \alpha\mathcal{C}_{pppp} - \beta\nabla g.$$

The snake model is a linear model and thus an efficient and powerful tool for object segmentation and edge integration, especially when there is a rough approximation of the boundary location. There is however an undesirable property that characterizes this model. It depends on the parameterization. The model is not geometric.

Manuscript received October 25, 1999; revised June 13, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aly A. Farag.

The authors are with the Computer Science Department, The Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: romang@cs.technion.ac.il).

Publisher Item Identifier S 1057-7149(01)08199-4.

ce qu'on peut faire avec :

- ✔ lire les formules
- ✔ croire les résultats
- ✘ vérifier les résultats
- ✘ reproduire les résultats
- ✘ voir les images en détail
- ✘ voir les graphes en détail

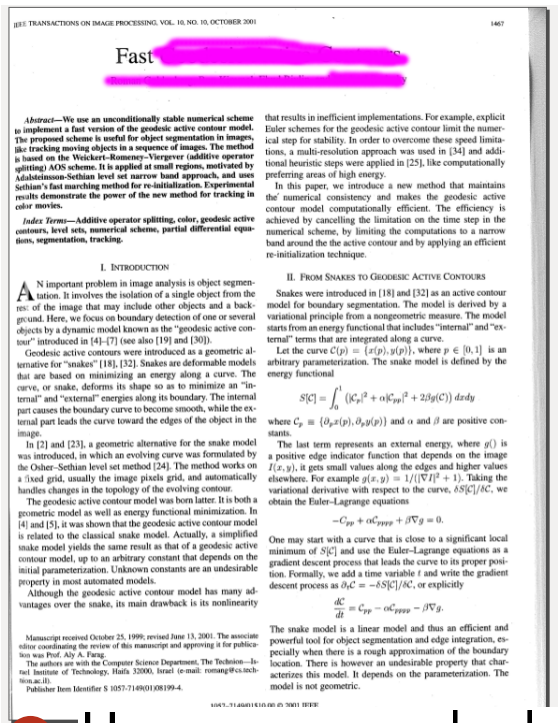
Courtesy of Enric Meinhardt-Llopis, CANUM 2016

Quick poll

1. Have you ever tried to reproduce some research results ?
2. Have you ever failed ?

Article typique en traitement d'image

fichier PDF



IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 10, NO. 10, OCTOBER 2001 1467

Fast [redacted]

Abstract—**We use an unconditionally stable numerical scheme to implement a fast version of the geodesic active contour model. The proposed scheme is useful for object segmentation in images, like tracking moving objects in a sequence of images. The method is based on the Weickert–Romeney–Viergever (additive operator splitting) AOS scheme. It is applied at small regions, motivated by Adalsteinsson–Sethian level set narrow band approach, and uses Sethian's fast marching method for re-initialization. Experimental results demonstrate the power of the new method for tracking in color movies.**

Index Terms—**Additive operator splitting, color, geodesic active contours, level sets, numerical scheme, partial differential equations, segmentation, tracking.**

I. Introduction

An important problem in image analysis is object segmentation. It involves the isolation of a single object from the rest of the image that may include other objects and a background. Here, we focus on boundary detection of one or several objects by a dynamic model known as the "geodesic active contour" introduced in [4]–[7] (see also [19] and [30]).

Geodesic active contours were introduced as a geometric alternative for "snakes" [18], [32]. Snakes are deformable models that are based on minimizing an energy along a curve. The curve, or snake, deforms its shape so as to minimize an "internal" and "external" energies along its boundary. The internal part causes the boundary curve to become smooth, while the external part leads the curve toward the edges of the object in the image.

In [2] and [23], a geometric alternative for the snake model was introduced, in which an evolving curve was formulated by the Osher–Sethian level set method [24]. The method works on a fixed grid, usually the image pixels grid, and automatically handles changes in the topology of the evolving contour.

The geodesic active contour model was born latter. It is both a geometric model as well as energy functional minimization. In [4] and [5], it was shown that the geodesic active contour model is related to the classical snake model. Actually, a simplified snake model yields the same result as that of a geodesic active contour model, up to an arbitrary constant that depends on the initial parameterization. Unknown constants are an undesirable property in most automated models.

Although the geodesic active contour model has many advantages over the snake, its main drawback is its nonlinearity that results in inefficient implementations. For example, explicit Euler schemes for the geodesic active contour limit the numerical step for stability. In order to overcome these speed limitations, a multi-resolution approach was used in [34] and additional heuristic steps were applied in [25], like computationally preferring areas of high energy.

In this paper, we introduce a new method that maintains the numerical consistency and makes the geodesic active contour model computationally efficient. The efficiency is achieved by cancelling the limitation on the time step in the numerical scheme, by limiting the computations to a narrow band around the active contour and by applying an efficient re-initialization technique.

II. From Snakes to Geodesic Active Contours

Snakes were introduced in [18] and [32] as an active contour model for boundary segmentation. The model is derived by a variational principle from a nongeometric measure. The model starts from an energy functional that includes "internal" and "external" terms that are integrated along a curve.

Let the curve $\mathcal{C}(p) = \{x(p), y(p)\}$, where $p \in [0,1]$ is an arbitrary parameterization. The snake model is defined by the energy functional

$$S[\mathcal{C}] = \int_0^1 \left(|\mathcal{C}_p|^2 + \alpha |\mathcal{C}_{pp}|^2 + 2\beta g(\mathcal{C}) \right) dxdy$$

where $\mathcal{C}_p \equiv \{\partial_p x(p), \partial_p y(p)\}$ and α and β are positive constants.

The last term represents an external energy, where $g()$ is a positive edge indicator function that depends on the image $I(x,y)$, it gets small values along the edges and higher values elsewhere. For example $g(x,y) = 1/(|\nabla I|^2 + 1)$. Taking the variational derivative with respect to the curve, $\delta S[\mathcal{C}]/\delta \mathcal{C}$, we obtain the Euler–Lagrange equations

$$-\mathcal{C}_{pp} + \alpha \mathcal{C}_{pppp} + \beta \nabla g = 0.$$

One may start with a curve that is close to a significant local minimum of $S[\mathcal{C}]$ and use the Euler–Lagrange equations as a gradient descent process that leads the curve to its proper position. Formally, we add a time variable t and write the gradient descent process as $\partial_t \mathcal{C} = -\delta S[\mathcal{C}]/\delta \mathcal{C}$, or explicitly

$$\frac{d\mathcal{C}}{dt} = \mathcal{C}_{pp} - \alpha \mathcal{C}_{pppp} - \beta \nabla g.$$

The snake model is a linear model and thus an efficient and powerful tool for object segmentation and edge integration, especially when there is a rough approximation of the boundary location. There is however an undesirable property that characterizes this model. It depends on the parameterization. The model is not geometric.

Manuscript received October 25, 1999; revised June 13, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aly A. Farag.

The authors are with the Computer Science Department, The Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: romang@cs.technion.ac.il).

Publisher Item Identifier S 1057-7149(01)08199-4.

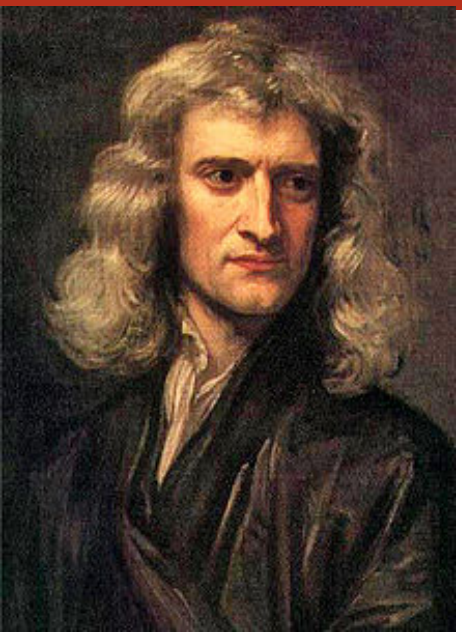
ce qu'on peut faire avec :

- ✔ lire les formules
- ✔ croire les résultats
- ✘ vérifier les résultats
- ✘ reproduire les résultats
- ✘ voir les images en détail
- ✘ voir les graphes en détail

Courtesy of Enric Meinhardt-Llopis, CANUM 2016

3. Have you ever had trouble reproducing the work of one of your student (or even your own work 😉)?

Reproducibility of experimental results is the hallmark of science

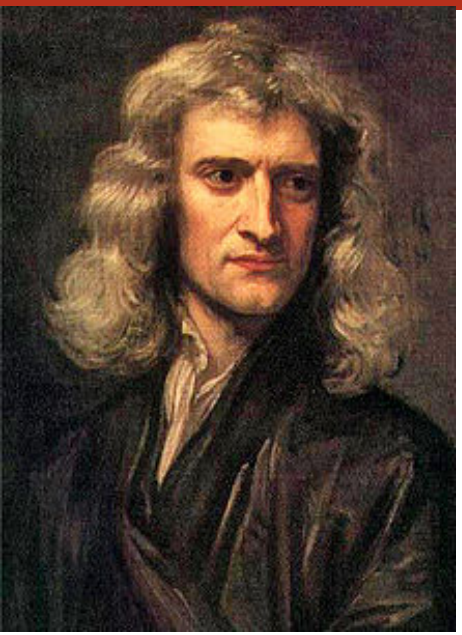


What Descartes did was a good step. You have added much several ways [..] If I have seen further it is by standing on the shoulders of Giants.

– ***Isaac Newton***, *February 1676*

Science allows to discover truth by building on previous discoveries.

Reproducibility of experimental results is the hallmark of science



What Descartes did was a good step. You have added much several ways [..] If I have seen further it is by standing on the shoulders of Giants.
*– **Isaac Newton**, February 1676*

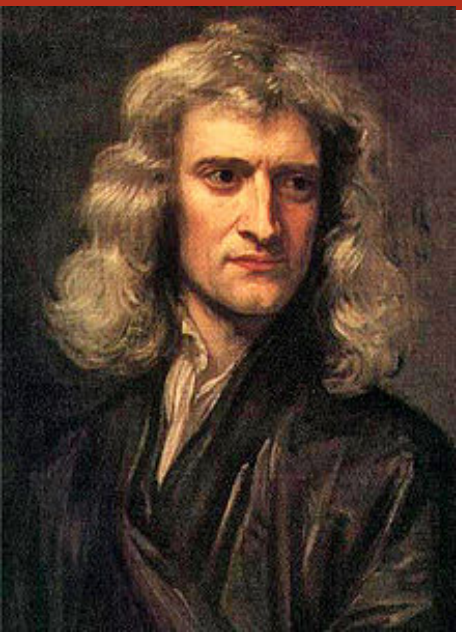
In a letter to his rival Robert Hooke

Science allows to discover truth by building on previous discoveries.

1662: **Robert Hooke**, Curator of Experiments for the Royal Society, coins the term *experimentum crucis*.

Only good experiments allow to build sound theories and refute bad ones

Reproducibility of experimental results is the hallmark of science



What Descartes did was a good step. You have added much several ways [..] If I have seen further it is by standing on the shoulders of Giants.
– ***Isaac Newton***, *February 1676*

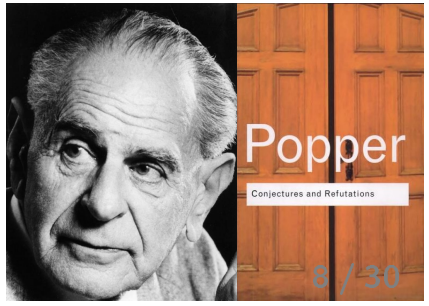
In a letter to his rival Robert Hooke

Science allows to discover truth by building on previous discoveries.

1662: **Robert Hooke**, Curator of Experiments for the Royal Society, coins the term *experimentum crucis*.

Only good experiments allow to build sound theories and refute bad ones

1934: **Karl Popper** puts the notions of falsifiability and crucial experiment as the hallmark of science



A Reproducibility Crisis?

Many scandals in social psychology, stem cells, organic transistors (The Economics of Replication) leading to retractation/resignation.

- At CNRS, we recently had the Olivier Viannet case and it appears we'll have the sequel with Susana Rivas at INRA ☹

A recent scandal In 2013, *Dong-Pyou Han*, a former assistant professor of biomedical sciences at Iowa State University was disgraced:

- Falsified blood results to make it appear as though a vaccine he was working on had exhibited anti-HIV activity
- Han and his team received $\approx$ \$19 million from NIH
- Retraction and resignation of university
- Han was sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. He was also fined US \$7.2 million!

A Reproducibility Crisis?

Many scandals in social psychology, stem cells, organic transistors (The Economics of Replication) leading to retractation/resignation.

- At CNRS, we recently had the Olivier Viannet case and it appears we'll have the sequel with Susana Rivas at INRA ☹

A recent scandal In 2013, *Dong-Pyou Han*, a former assistant professor of biomedical sciences at Iowa State University was disgraced:

- Falsified blood results to make it appear as though a vaccine he was working on had exhibited anti-HIV activity
- Han and his team received $\approx$ \$19 million from NIH
- Retraction and resignation of university
- Han was sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. He was also fined US \$7.2 million!

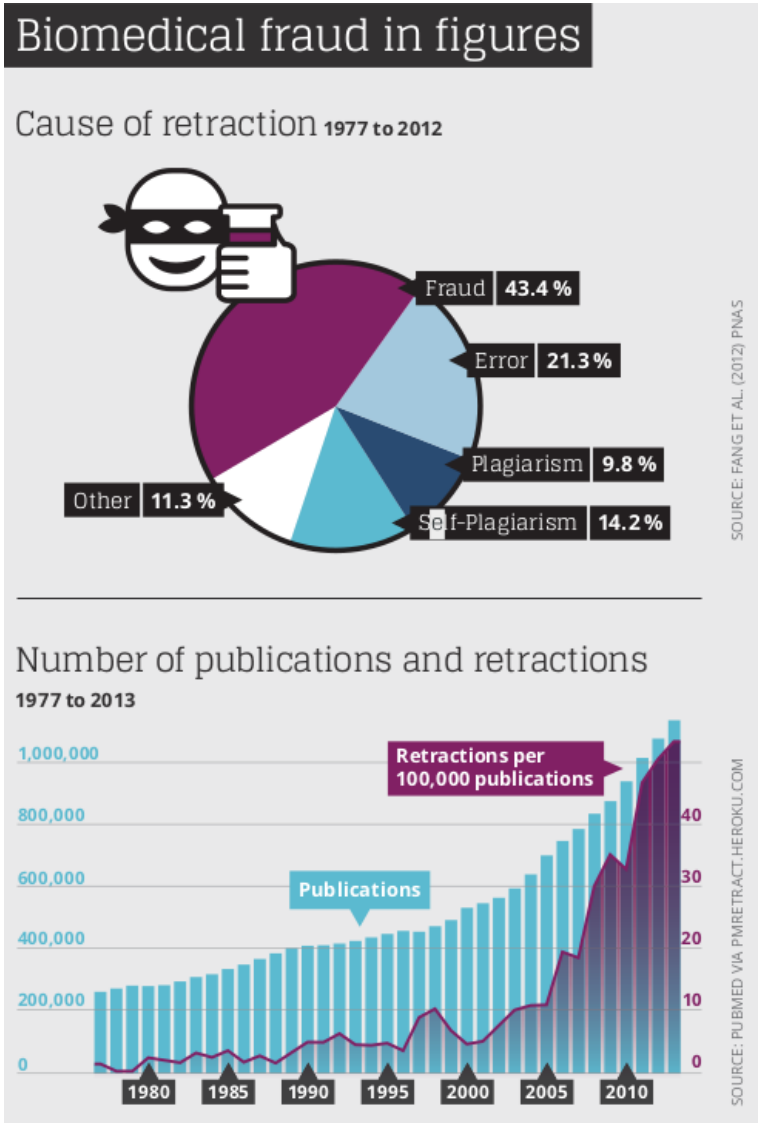
Scientific fraud is bad but we should avoid witch-hunt

- August 5, 2014, Yoshiki Sasai (stem cell, considered for Nobel Prize) hanged in his laboratory at the RIKEN (Japan). Fraud suspicion...
- More and more alerts but also false alarms...

Loosing trust What's the difference between science and charlatanism in people's mind if they get worthless/unreliable information?

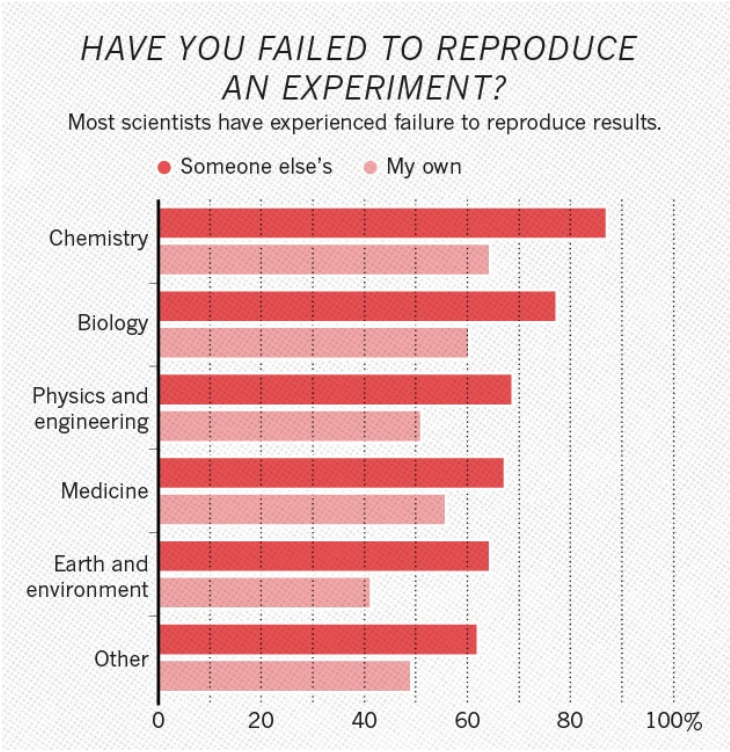
Is Fraud a new phenomenon?

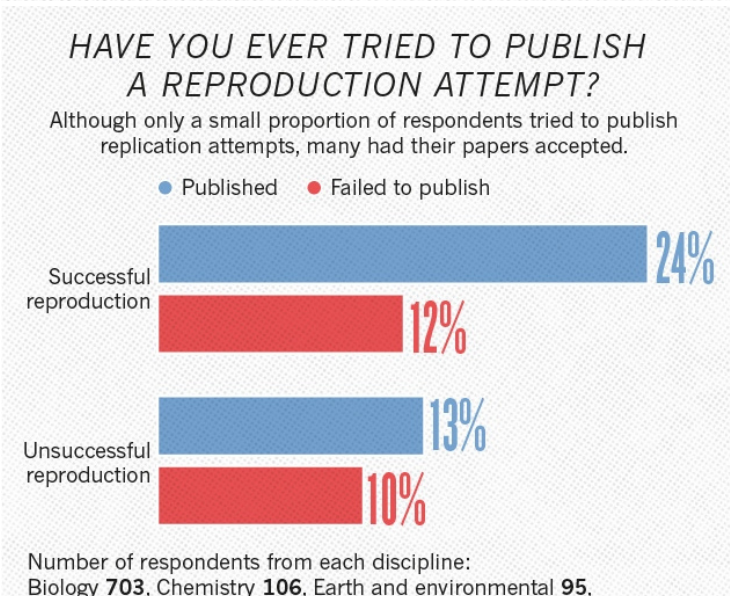
The Battle against Scientific Fraud in the CNRS International Magazine



Galileo (data fabrication), **Ptolemy** (plagiarism), **Mendel** (data enhancement), **Pasteur** (rigorous but hided failures), ...

Why are scientific studies so difficult to reproduce?





1,500 scientists lift the lid on reproducibility, Nature, May 2016

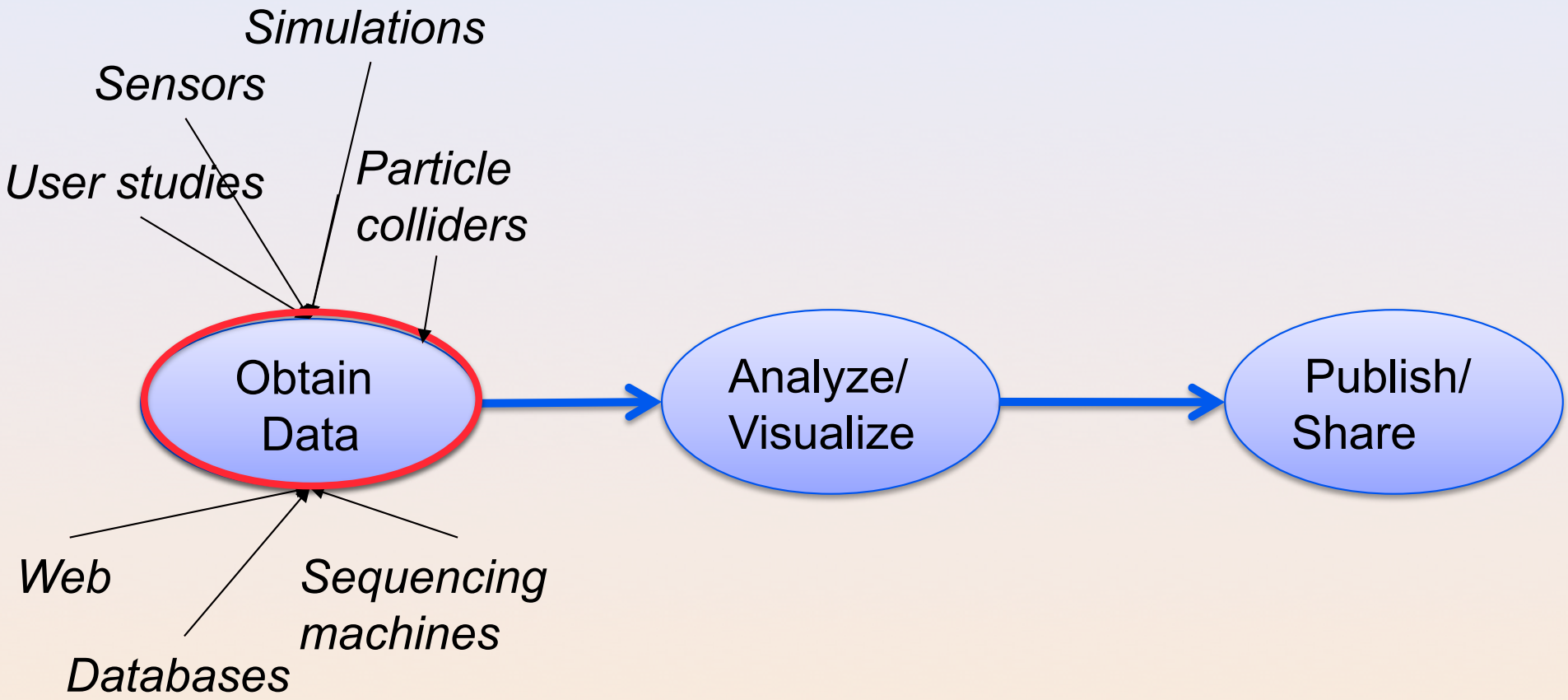
Social causes

- Fraud, conflict of interest (pharmaceutic, . . .)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1+ million articles per year!

Methodological or technical causes

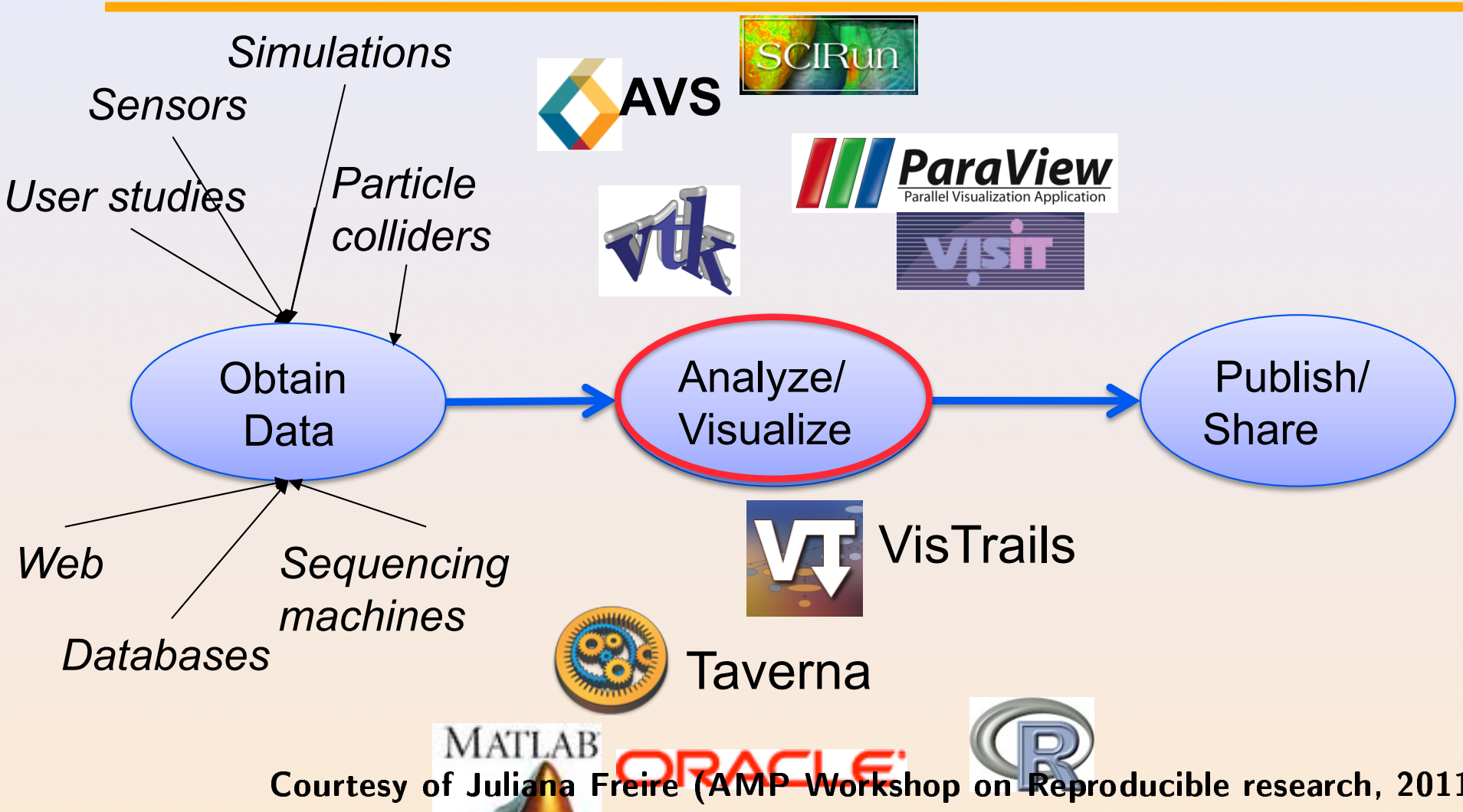
- The many biases (apophenia, confirmation, hindsight, experimenter, . . .): bad designs
- Selective reporting, weak analysis (statistics, data manipulation mistakes, computational errors)
- Lack of information, code/raw data unavailable

Science Today: Data Intensive



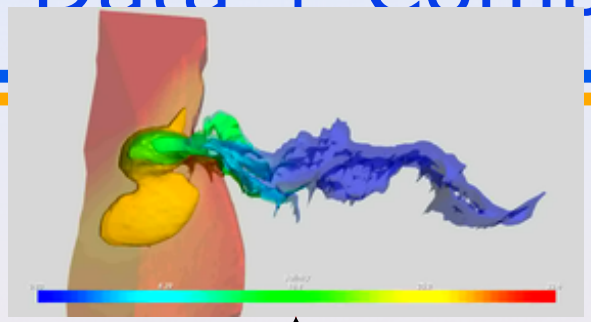
Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

Science Today: Data + Computing Intensive



Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

Science Today: Data + Computing Intensive



Simulations

Sensors

User studies

Particle colliders

Microsoft PowerPoint

Obtain Data

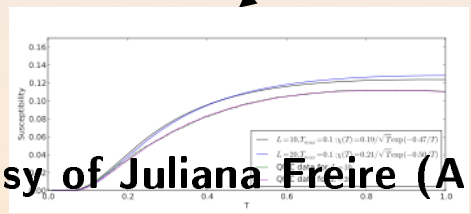
Analyze/ Visualize

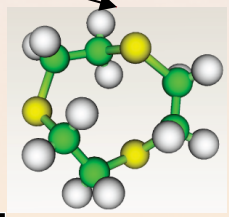
Publish/ Share

Web

Sequencing machines

Databases

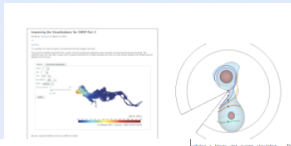




Courtesy of Juliana Freire (AM… eproducible research, 2011)

What does Science look like today ?

Science Today: Data + Computing Inte



Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

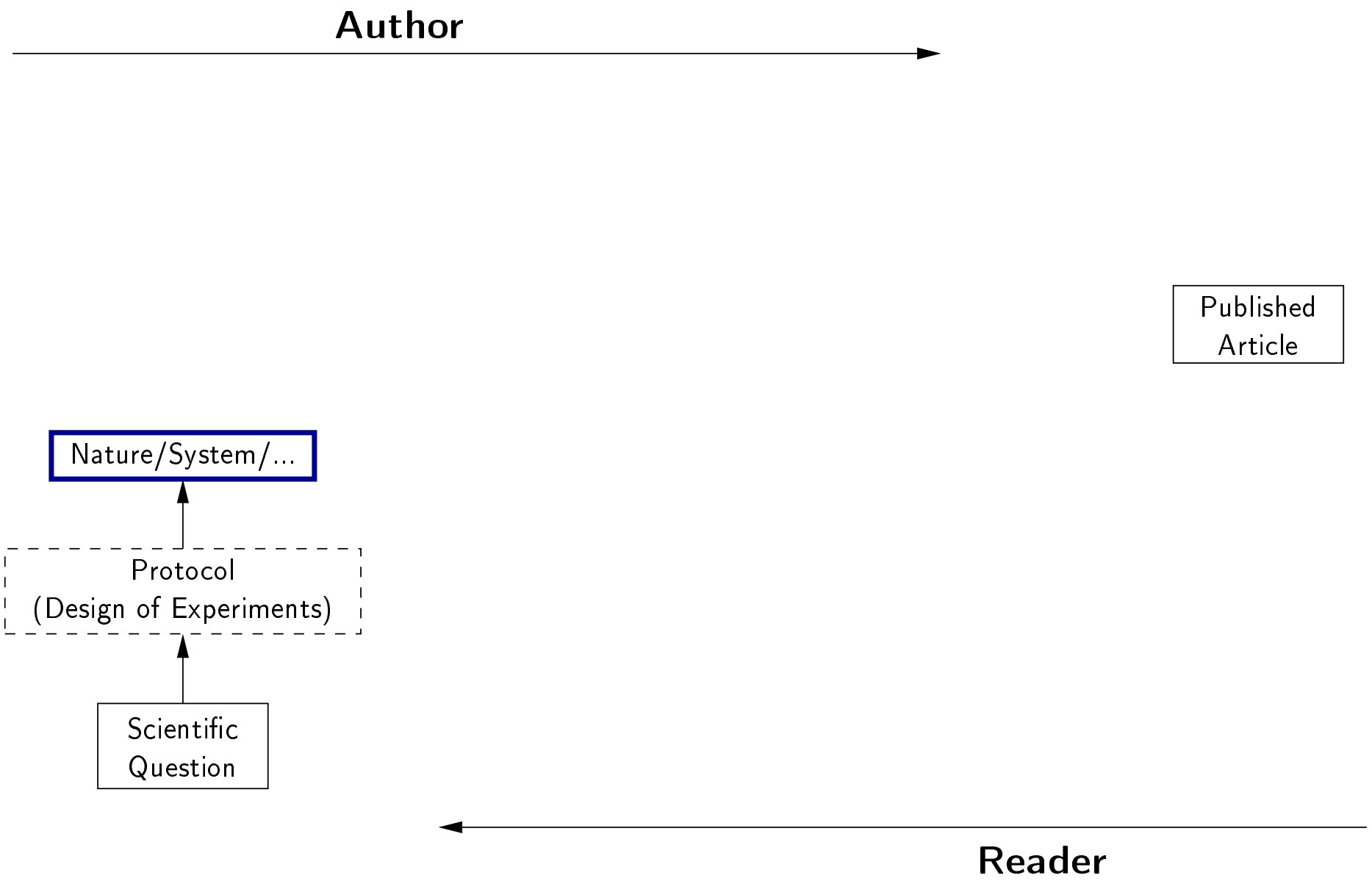
Science Today: Incomplete Publications

- Publications are just the tip of the iceberg
  - Scientific record is incomplete--- to large to fit in a paper
  - Large volumes of data
  - Complex processes
- Can't (easily) reproduce results

Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

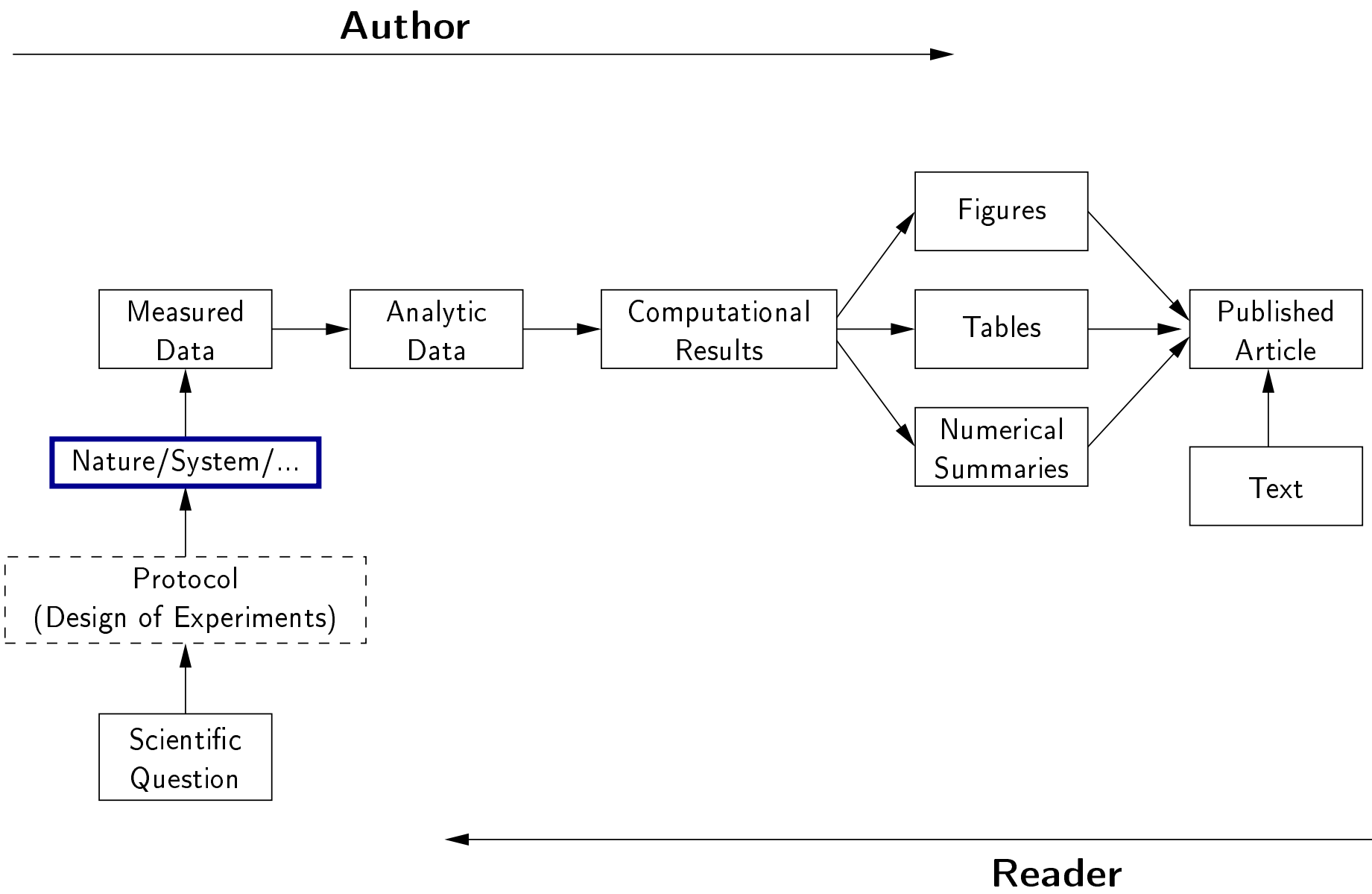
Outline

Reproducible Research: Trying to Bridge the Gap



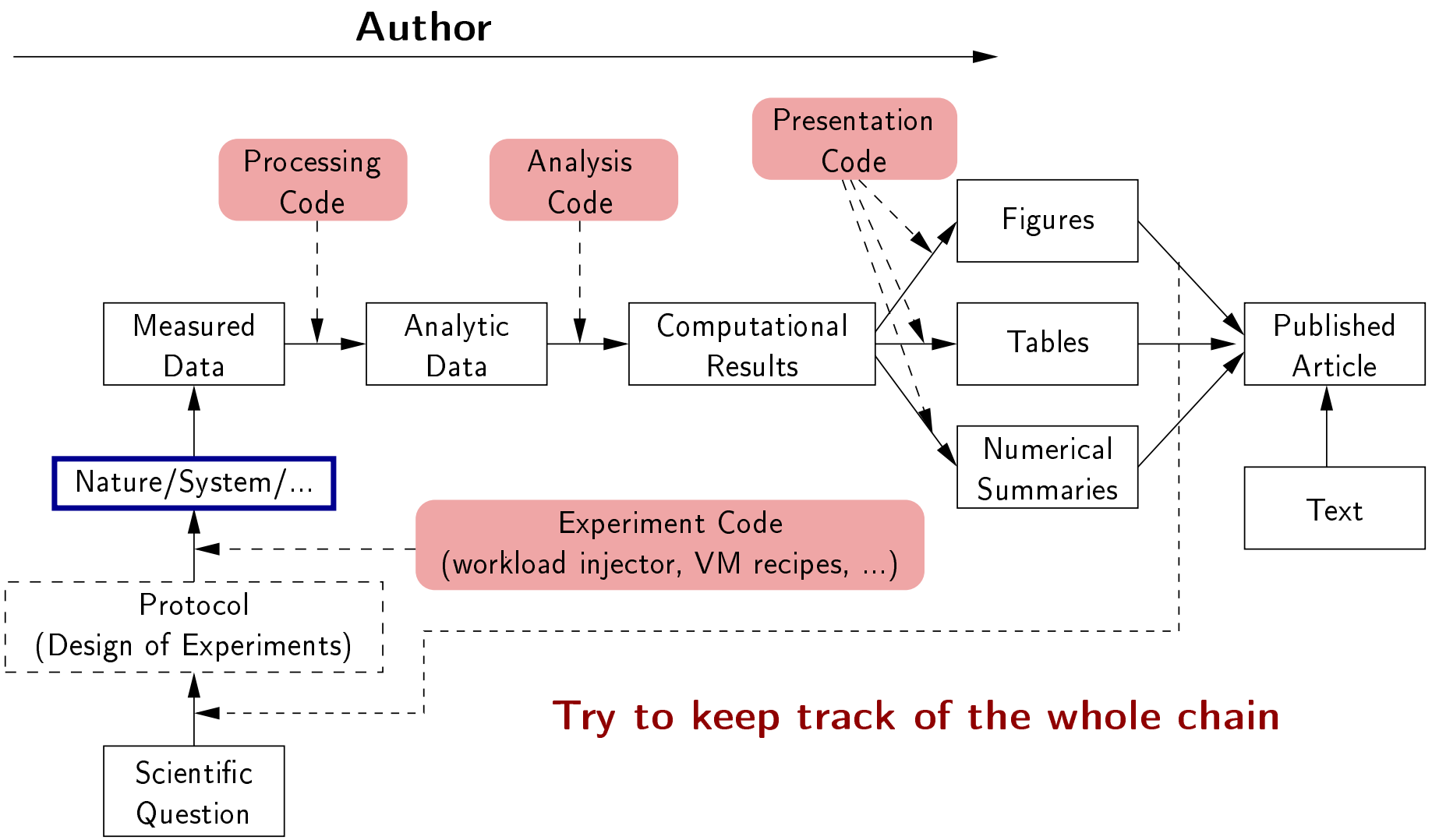
Inspired by Roger D. Peng's lecture on reproducible research, May 2014

Reproducible Research: Trying to Bridge the Gap



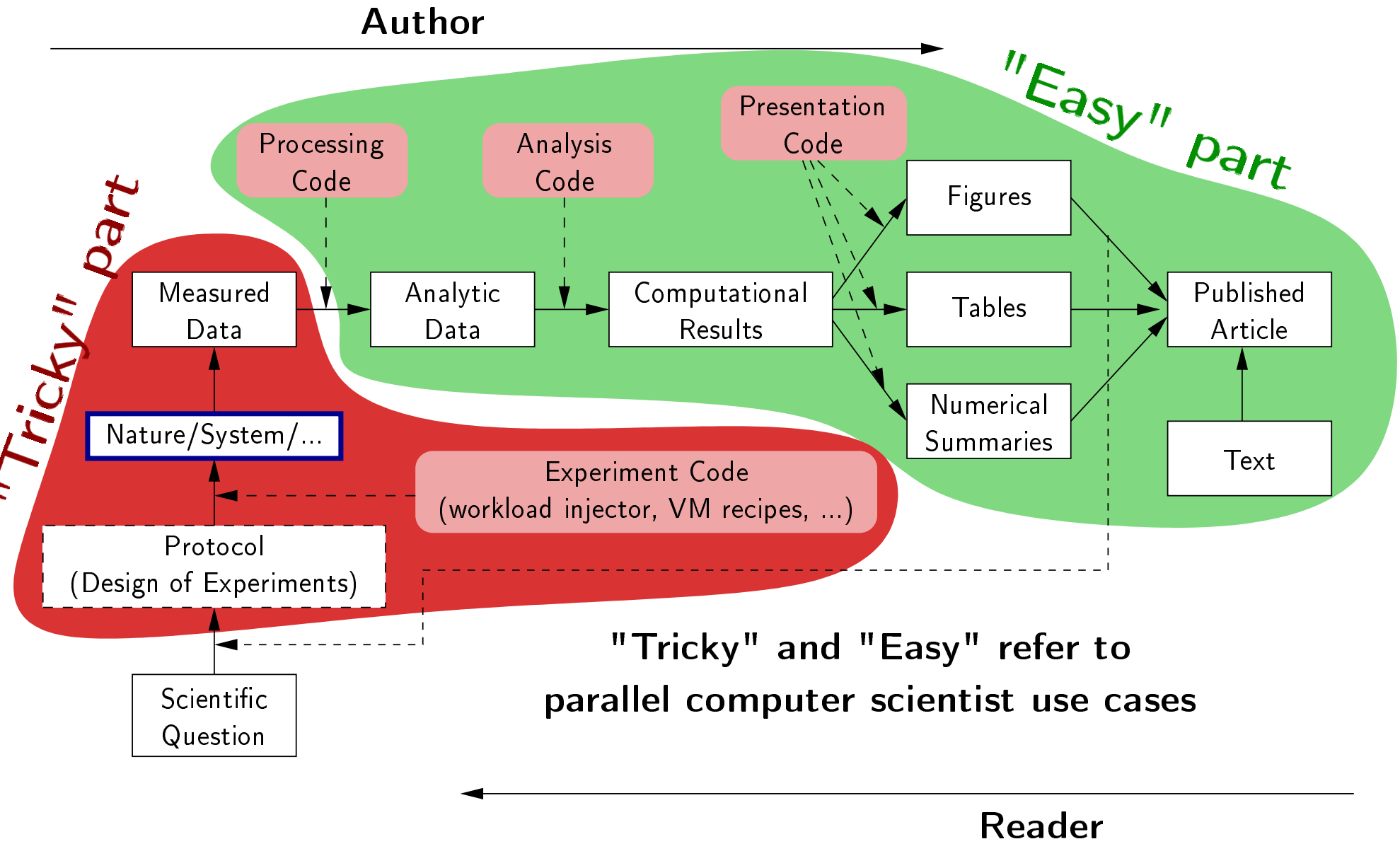
Inspired by Roger D. Peng's lecture on reproducible research, May 2014

Reproducible Research: Trying to Bridge the Gap



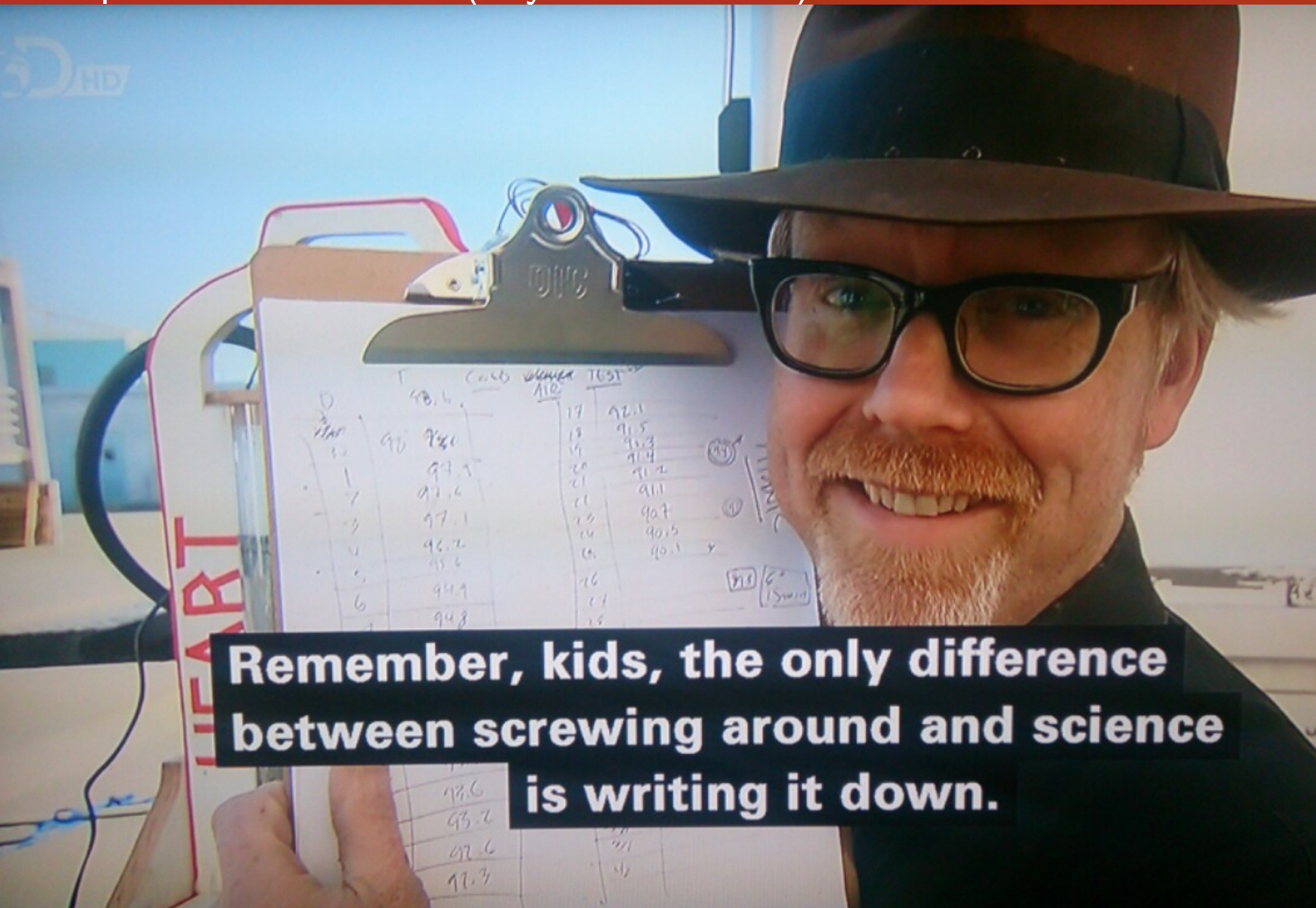
Inspired by Roger D. Peng's lecture on reproducible research, May 2014

Reproducible Research: Trying to Bridge the Gap



Inspired by Roger D. Peng's lecture on reproducible research, May 2014

Step 1: Take notes! (Mythbusters 😉)



Step 2: Control what's happening

Both the scientific process and the software environment have become extremely complex

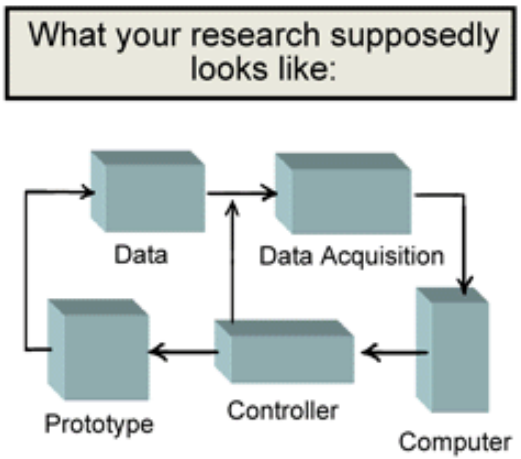


Figure 1. Experimental Diagram

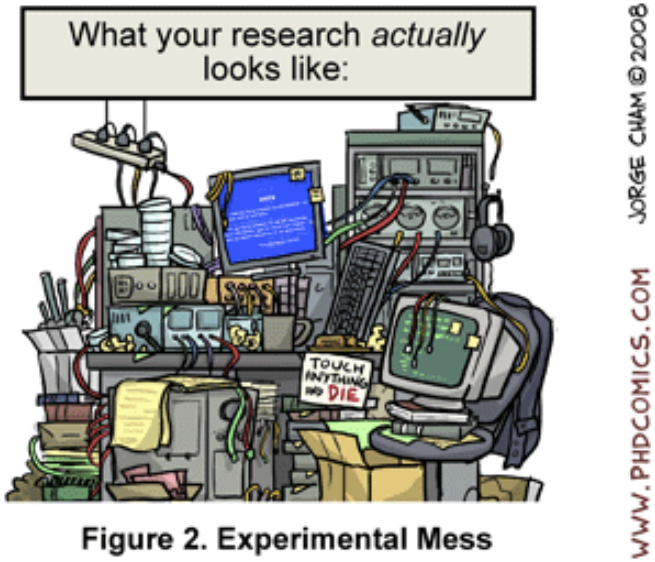


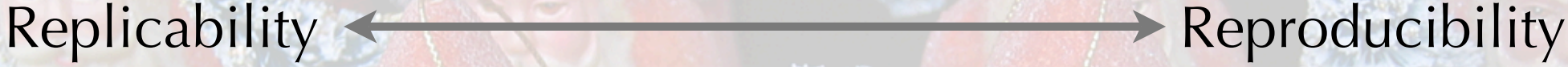
Figure 2. Experimental Mess

Computational biologists and physicist have pushed forward with the development of many tools:

- Software and environment preservation/reconstruction
- Workflow engines, data provenance
- Dissemination platforms, data/workload archive
- Laboratory notebooks, litterate programming, replicable articles

And this landscape is rapidly evolving

Step 3: Reproducibility? What Are We Talking About Exactly?



Reproduction of the original results using the same tools

- by the original author on the same machine
- by someone in the same lab/using a different machine
- by someone in a different lab

Reproduction using different software, but with access to the original code

Completely independent reproduction based only on text description, without access to the original code

Note that the **terminology varies** (repeatition, replication, reproduction, reuse, corroboration, ...)

Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

And in practice: remember the value of data

Ten Simple Rules for Reproducible Computational Research, PlosOne

1. For Every Result, Keep Track of How It Was Produced
2. Avoid Manual Data Manipulation Steps
3. Archive the Exact Versions of All External Programs Used
4. Version Control All Custom Scripts
5. Record All Intermediate Results, When Possible in Standardized Formats
6. For Analyses That Include Randomness, Note Underlying Random Seeds
7. Always Store Raw Data behind Plots
8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
9. Connect Textual Statements to Underlying Results
10. Provide Public Access to Scripts, Runs, and Results

But several ways to follow them

- **Preserve the Mess**
  Automatically keeping track of everything
- **Encourage Cleanliness**
  Ensure others can understand/adapt what was done

Outline

An *Provenance-Rich* Paper: ALPS2.0

The ALPS project release 2.0: Open source software for strongly correlated systems

B. Bauer[1] L. D. Carr[2] H.G. Evertz[3] A. Feiguin[4] J. Freire[5] S. Fuchs[6] L. Gamper[1] J. Gukelberger[1] E. Gull[7] S. Guertler[8] A. Hehn[1] R. Igarashi[9,10] S.V. Isakov[1] D. Koop[5] P.N. Ma[1] P. Mates[1,5] H. Matsuo[11] O. Parcollet[12] G. Pawłowski[13] J.D. Picon[14] L. Pollet[1,15] E. Santos[5] V.W. Scarola[16] U. Schollwöck[17] C. Silva[5] B. Surer[1] S. Todo[10,11] S. Trebst[18] M. Troyer[1]‡ M. L. Wall[2] P. Werner[1] S. Wessel[19,20]

[1]Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland
[2]Department of Physics, Colorado School of Mines, Golden, CO 80401, USA
[3]Institut für Theoretische Physik, Technische Universität Graz, A-8010 Graz, Austria
[4]Department of Physics and Astronomy, University of Wyoming, Laramie, Wyoming 82071, USA
[5]Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah 84112, USA
[6]Institut für Theoretische Physik, Georg-August-Universität Göttingen, Göttingen, Germany
[7]Columbia University, New York, NY 10027, USA
[8]Bethe Center for Theoretical Physics, Universität Bonn, Nussallee 12, 53115 Bonn, Germany

‡ Correspond

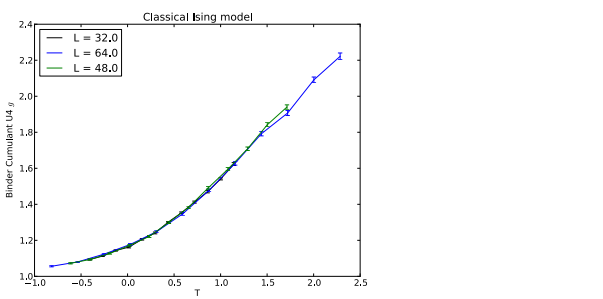
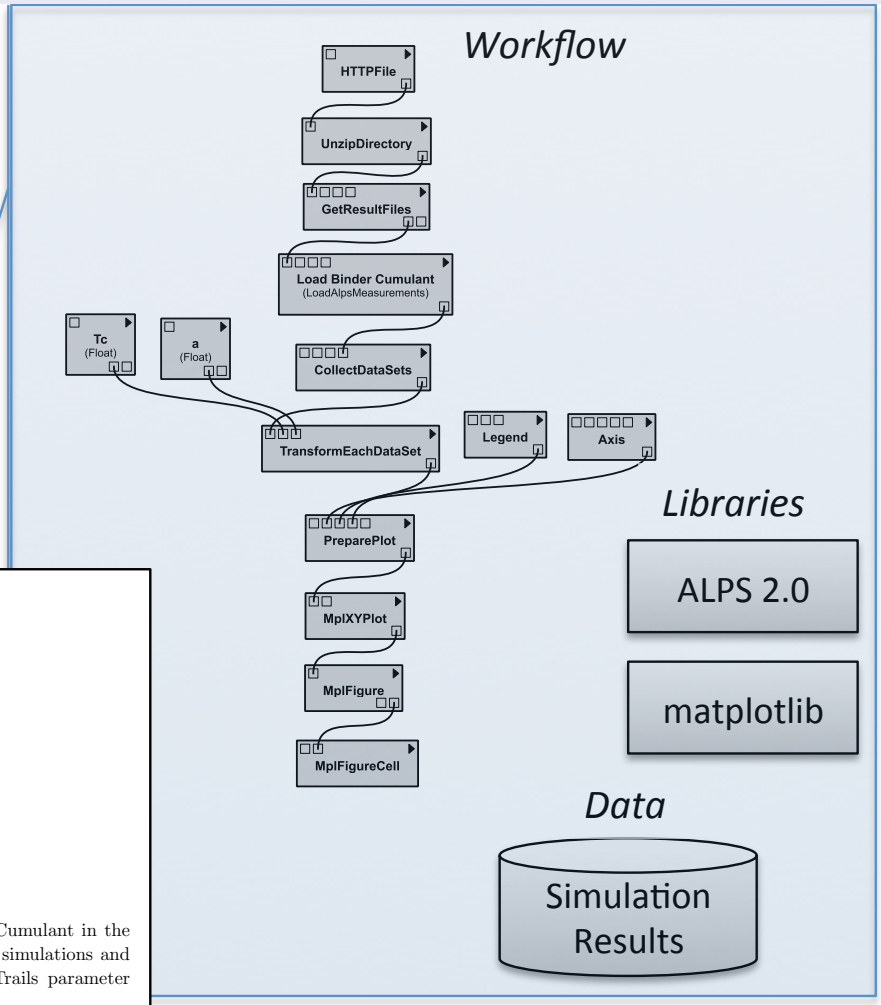


Figure 3. In this example we show a data collapse of the Binder Cumulant in the classical Ising model. The data has been produced by remotely run simulations and the critical exponent has been obtained with the help of the VisTrails parameter exploration functionality.



Courtesy of Juliana Freire (AMP Workshop on Reproducible research)

VCR: A Universal Identifier for Computational Results

Chronicing computations in real-time

VCR computation platform Plugin = Computation recorder

Regular program code

```
figure1 = plot(x)
save(figure1,'figure1.eps')
```

```
> file /home/figure1.eps saved
>
```

Courtesy of Matan Gavish and David Donoho (AMP Workshop on Reproducible research)

VCR: A Universal Identifier for Computational Results

Chronicing computations in real-time

VCR computation platform Plugin = Computation recorder

Program code with VCR plugin

```
repository vcr.nature.com
verifiable figure1 = plot(x)
```

```
> vcr.nature.com approved:
> access figure1 at https://vcr.nature.com/ffaaffb148d7
```

Courtesy of Matan Gavish and David Donoho (AMP Workshop on Reproducible research)

VCR: A Universal Identifier for Computational Results

Word-processor plugin App

LaTeX source

```
\includegraphics{figure1.eps}
```

LaTeX source with VCR package

```
\includeresult{vcr.thelancet.com/ffaaffb148d7}
```

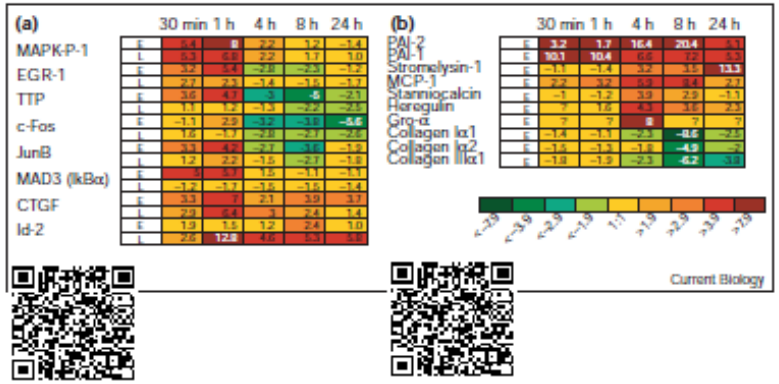
Permanently bind printed graphics to underlying result content

Courtesy of Matan Gavish and David Donoho (AMP Workshop on Reproducible research)

VCR: A Universal Identifier for Computational Results

Figure 3

Time course of serum stimulation. **(a)** Early passage (E: PD30) or late passage (L: PD89) BJ cultures were held in 0.5% serum for 2 days, then stimulated with 10% FBS. RNA levels from cultures at the indicated time points (Cy5 channel) were compared with the uninduced starting culture (Cy3 channel). Positive values indicate higher expression in induced cells; negative values indicate lower expression in induced cells. Question marks indicate that there was insufficient signal for detection. A complete listing of serum-responsive genes from this analysis is provided in Supplementary material. **(b)** The serum-responsiveness of select senescence-regulated genes in early passage (PD30) BJ fibroblasts.



senescence response appears to overlap substantially with gene expression patterns observed in activated fibroblasts during wound healing [24–26]. MCP-1, Gro-α, IL-1β and IL-15 are strong effectors of macrophage and neutrophil recruitment and activation [27,28]. The upregulation of Toll (Tlr-4) in senescent fibroblasts confirms the overall immune response behavior at senescence. Tlr-4 is an IL-1 receptor homolog and is implicated in the activation of the gene regulatory protein NF-κB, a function proposed to be part of the innate immune response [29]. The induction of IL-15 at senescence is also consistent with an innate immune response, as IL-15 can be induced by NF-κB-dependent transcription [30] and also participates in inflammatory disease processes [28].

Deficiencies in the response of senescent cells to serum stimulation have been reported, and include an inability to induce the expression of c-*fos* mRNA [31] and markers of late G1 and S phase [32]. In response to serum, expression of inflammatory chemokines, matrix-degrading proteases and their modulators is induced in early-passage dermal fibroblasts, and expression of matrix collagens is reduced. This transient burst of activity may represent the natural response of these cells in wound repair [24]. Id-2 transcripts were hyper-induced in serum-stimulated senescent

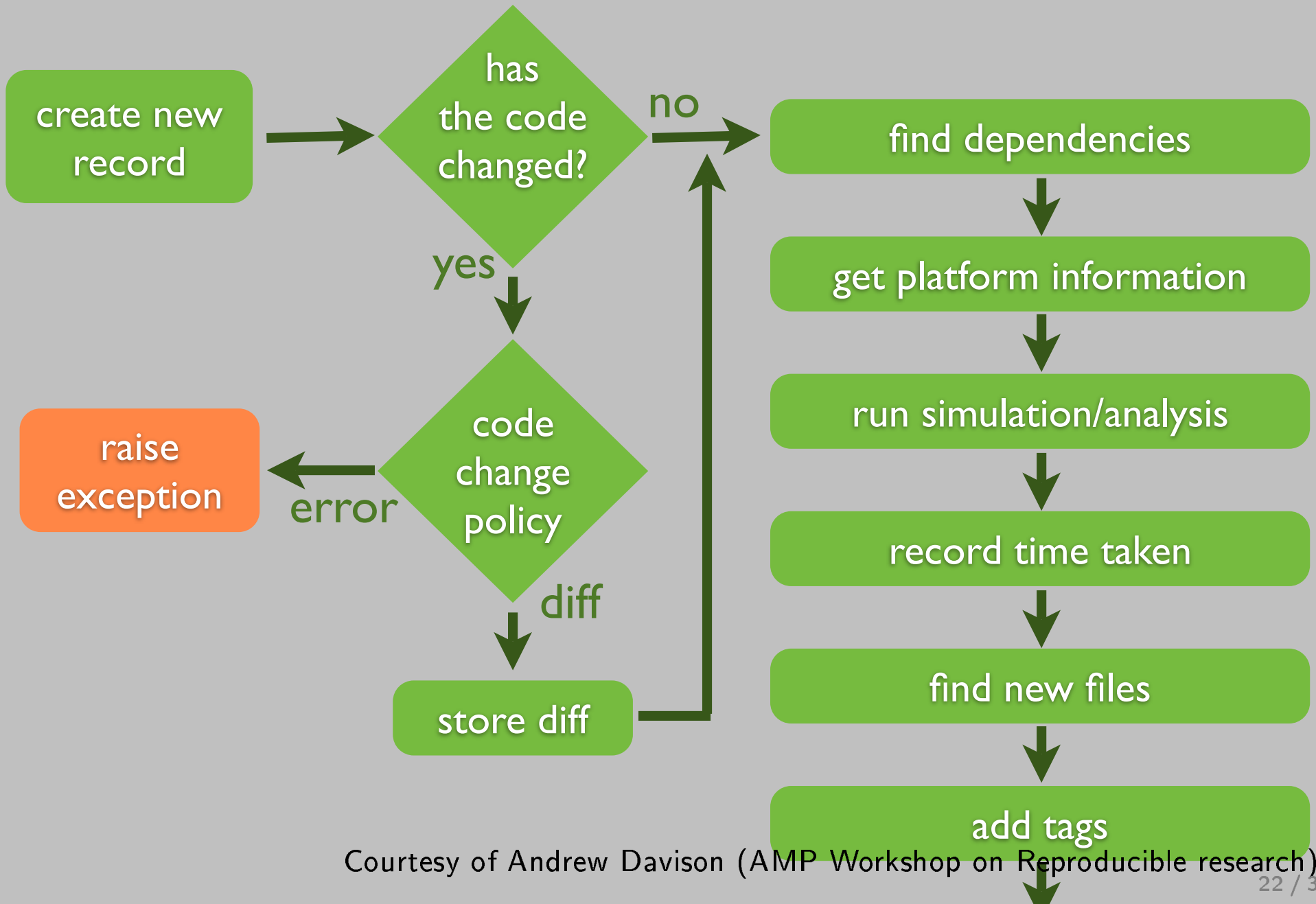
states overlap substantially with those in telomere-induced senescence (W.F., D.N.S., R. Allsopp, S. Lowe, and G. Ferbeyre, unpublished observations) and thus are likely to use many of the same activation processes.

The pattern of gene expression at senescence varies substantially in different cell types. Although the expression of matrix and structural proteins, such as the collagens, keratins and auxiliary factors, is repressed in RPE cells, inflammatory regulators are not induced, in contrast to dermal fibroblasts. Physiologically, this would make sense, as an acute inflammatory response in a tissue critical for normal vision would be likely to have deleterious consequences. However, as the RPE layer has a central role in the deposition and maintenance of extracellular matrix in the retina, decrements in the ability of senescent RPE cells to maintain appropriate expression patterns, as evidenced by decreased expression of collagens, keratins, aggrecan, transglutaminase and so on, would be predicted to have adverse affects on retinal architecture. Dysfunction of the RPE cell layer is considered to be a substantial factor in the development of age-related macular degeneration [36].

Surprisingly, early-passage HUVECs overexpress many of the markers associated with senescence in dermal fibro-

Courtesy of Matan Gavish and David Donoho (AMP Workshop on Reproducible research)

Sumatra: an "experiment engine" that helps taking notes



Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

Sumatra: an "experiment engine" that helps taking notes

```
$ smt comment 20110713-174949 "Eureka! Nobel prize
here we come."
```

Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

Sumatra: an "experiment engine" that helps taking notes

```
$ smt tag “Figure 6”
```

Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

Sumatra: an "experiment engine" that helps taking notes

Sumatra: TestProject: List of records

http://127.0.0.1:8002/

Google

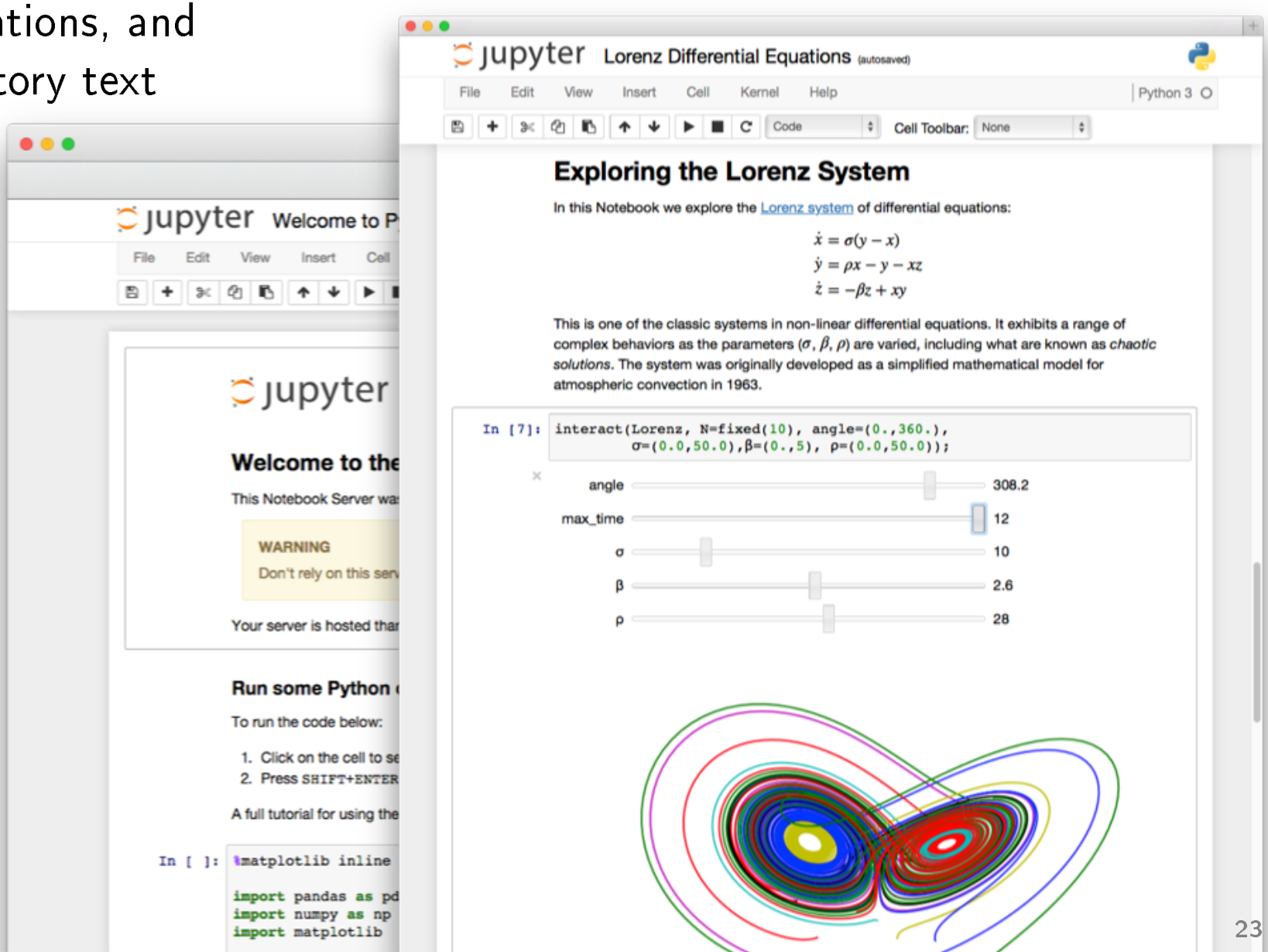
TestProject: List of records

Delete / include data	Label	Reason	Outcome	Duration	Processes	Simulator		Script			Date	Time	Tags
						Name	Version	Repository	Main file	Version			
☐	20100709-154255		'Eureka! Nobel prize here we come.'	0.59 s		Python	2.5.2	/Users/andrew/tmp/SumatraTest	main.py	396c2020ca50	09/07/2010	15:42:55	
☐	20100709-154309			0.59 s		Python	2.5.2	/Users/andrew/tmp/SumatraTest	main.py	396c2020ca50	09/07/2010	15:43:09	
☐	haggling	'determine whether the gourd is worth 3 or 4 shekels'	'apparently, it is worth NaN shekels.'	0.59 s		Python	2.5.2	/Users/andrew/tmp/SumatraTest	main.py	396c2020ca50	09/07/2010	15:43:20	foobar
☐	20100709-154338	'test effect of a smaller time constant'		0.59 s		Python	2.5.2	/Users/andrew/tmp/SumatraTest	main.py	396c2020ca50	09/07/2010	15:43:38	
☐	haggling_repeat	Repeat experiment haggling	The new record exactly matches the original.	0.58 s		Python	2.5.2	/Users/andrew/tmp/SumatraTest	main.py	396c2020ca50	09/07/2010	15:43:47	

Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

Ipython/Jupyter Notebook

Web app: create and share documents that contain live code, equations, visualizations, and explanatory text



Reprozip

Automagically pack your experiment to fight dependency hell

ON THE ORIGINAL MACHINE

```
$ pip install reprozip
$ reprozip trace ./myexperiment -my --options inputs/somefile.csv other_file_here.bin
experiment: 0%... 25%... 50%... 75%... 100%
result: 42.137
Configuration file written in .reprozip/config.yml
Edit that file then run the packer -- use 'reprozip pack -h' for help
$ reprozip pack my_experiment.rpz
[REPROZIP] 17:26:42.588 INFO: Creating pack my_experiment.rpz...
[REPROZIP] 17:26:42.589 INFO: Adding files from package coreutils...
[REPROZIP] 17:26:42.601 INFO: Adding files from package libc6...
[REPROZIP] 17:26:42.906 INFO: Adding other files...
[REPROZIP] 17:26:43.450 INFO: Adding metadata...
```


ON ANOTHER MACHINE

```
$ pip install reprounzip[all]
$ reprounzip vagrant setup my_experiment.rpz mydirectory
Bringing machine 'default' up with 'virtualbox' provider...
==> default: Importing base box 'remram/debian-7-amd64'...
==> default: Booting VM...
==> default: Machine booted and ready!
==> default: Running provisioner: shell...
$ reprounzip vagrant run mydirectory
experiment: 0%... 25%... 50%... 75%... 100%
result: 42.137
$ reprounzip vagrant upload /tmp/new_config:global-config
$ reprounzip vagrant run mydirectory --cmdline ./myexperiment --other --options
inputs/somefile.csv
experiment: 0%... 25%... 50%... 75%... 100%
result: -17.814
```

New Tools for Computational Reproducibility

- Dissemination Platforms:
  - ResearchCompendia.org
  - IPOL
  - Madagascar
  - MLOSS.org
  - thedatahub.org
  - nanoHUB.org
  - Open Science Framework
  - The DataVerse Network
  - RunMyCode.org
- Workflow Tracking and Research Environments:
  - VisTrails
  - Kepler
  - CDE
  - Galaxy
  - GenePattern
  - Synapse
  - Sumatra
  - Taverna
  - Pegasus
- Embedded Publishing:
  - Verifiable Computational Research
  - Sweave
  - knitR
  - Collage Authoring Environment
  - SHARE

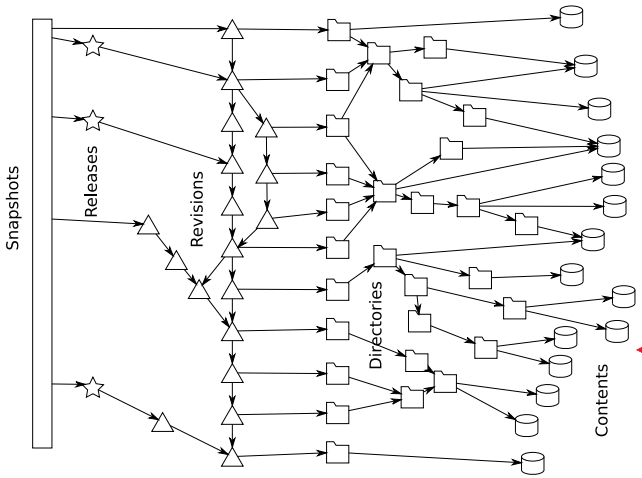
Courtesy of Victoria Stodden (UC Davis, Feb 13, 2014)

And also: **Org-Mode** ☺, **Figshare**, **Zenodo**, **ActivePapers** ☺, **Elsevier executable paper** ☹, ...

Outline

On the "technical" side (1/2)

- Better documenting what we do: Laboratory notebooks
  - Literate programming is great for analysis, and reproducible articles but does not go well yet with conducting experiments and workflows
  - A real adoption of such practice requires more storage and the ability to navigate in such information
- Better software engineering practice: Public releases, devops approach
  - Moving/evolving technology. Preservation ? Adoption ?
  - Should not slow down research
- Fighting against software/data degradation: Software Heritage, zenodo
  - Challenges: multiple! exploitation, navigation, storage, ...

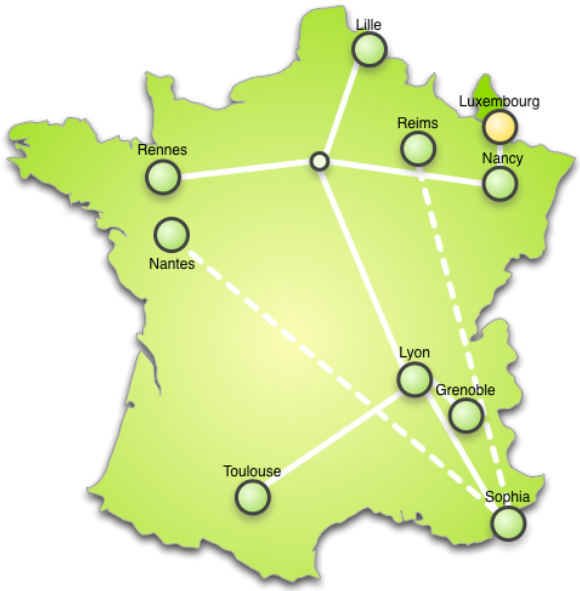


Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

On the "technical" side (2/2)

- Better experimental practice and platforms: FIT IoT-lab, G5K are world leading experimental infrastructures; rely on standard simulators (SimGrid, NS3)
  - Maintenance cost, keeping in pace with technology, practices for prototype platforms, control, sharing of experimental conditions with others, experimental engines



- Workload sharing, repositories.
  - Storage, evolving workload, cleaning/curating data, meta data to know how to use it, anonymization

On the "social" side

Slight cultural changes in our relation to publication and daily practice

- Changing our social model to favor adoption of better practice
  - Artifact evaluation, open reviews, ... (e.g., IPOL, ReScience)
  - Promote a different model
- Learning is the essence of our work. $\rightsquigarrow$ Train our researchers and students
  - Better teaching/understanding of statistics, experimental practice, design of experiments

It's up to us. **We** should **care** and **take the lead**

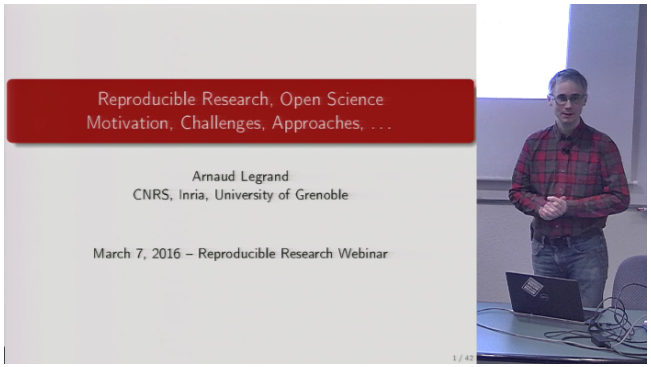
Main benefit:

- Higher confidence in our work $\rightsquigarrow$ definite competitive advantage
- Our research becomes sound, deeper, auditable, more visible, reusable, ...

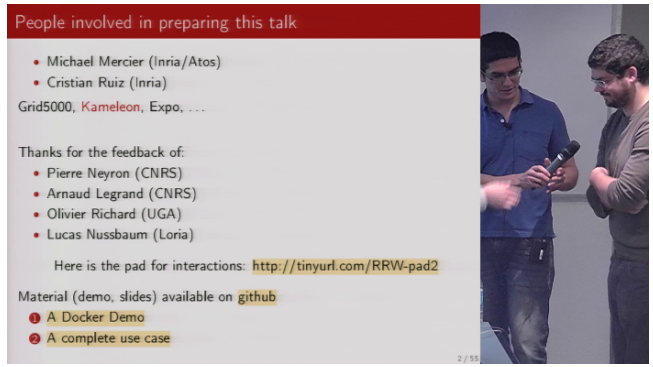
Webinars: Learning by Doing

Many different tools/approaches developed in various communities

1. Replicable article
2. Logging your activity
3. Logging and backing up your data
4. Organizing your data
5. Mastering your environment
6. Controlling your experiments
7. Making your data/code/article available
8. Publication modes
9. Artifact Evaluation



Litterate programming



Controling your environment

Numerical reproducibility

Logging and backing up

https://github.com/alegrand/RR_webinars