Reproducible Research: Where Do We Stand?

Arnaud Legrand
CNRS, Inria/POLARIS, University of Grenoble

February 28, 2017 – ENS Rennes

A little background information about myself

1997 ENS Lyon in Computer Science

2000-2003 Ph.D. Computer Science, École Normale Supérieure de Lyon. *Heterogeneous parallel algorithms and scheduling : static and dynamic approaches* (O. Beaumont and Y. Robert).

2003-2004: Post-Doct. École Normale Supérieure de Lyon

2004-2005: Post-Doct. Univ. California San Diego

Oct. 2004-. . . : Tenured CNRS Researcher at Laboratoire d'Informatique de Grenoble (Inria MESCAL team).

Main research topics:

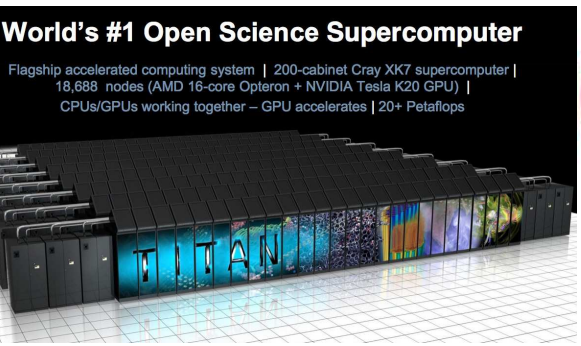
1. **Scheduling for Distributed Platforms**: parallel algorithms, scheduling, combinatorial optimization, game theory, . . .
2. **Simulation and Performance Evaluation**: modeling, simulation and validation, probabilities and statistics, data visualisation, . . .

Jan. 2015 Leader of the Inria POLARIS team

POLARIS in a nutshell

Large distributed infrastructures

- **HPC/cloud/...**
- Wireless networks
- Smart grids
- Transportation systems




Common questions scalability, resilience, adaptability, capacity planning, energy consumption, ...

Common characteristics ever growing size, distributed, heterogeneous, user-centric

$\rightsquigarrow$ stochastic nature

- Experiment design: observation tools, experimental methodology (design, control)
- Modeling and Simulation: discrete event simulation, emulation, Markov chains, perfect sampling, mean field limits, Monte Carlo methods, ...
- Analysis: workload characterization, visualization and anal. of parallel applications
- Optimization: stochastic approximations, game theory, mean field games, primal dual optimization, learning, information theory

Reproducible Research: Where Do We Stand?

Arnaud Legrand
CNRS, Inria/POLARIS, University of Grenoble

February 28, 2017 – ENS Rennes

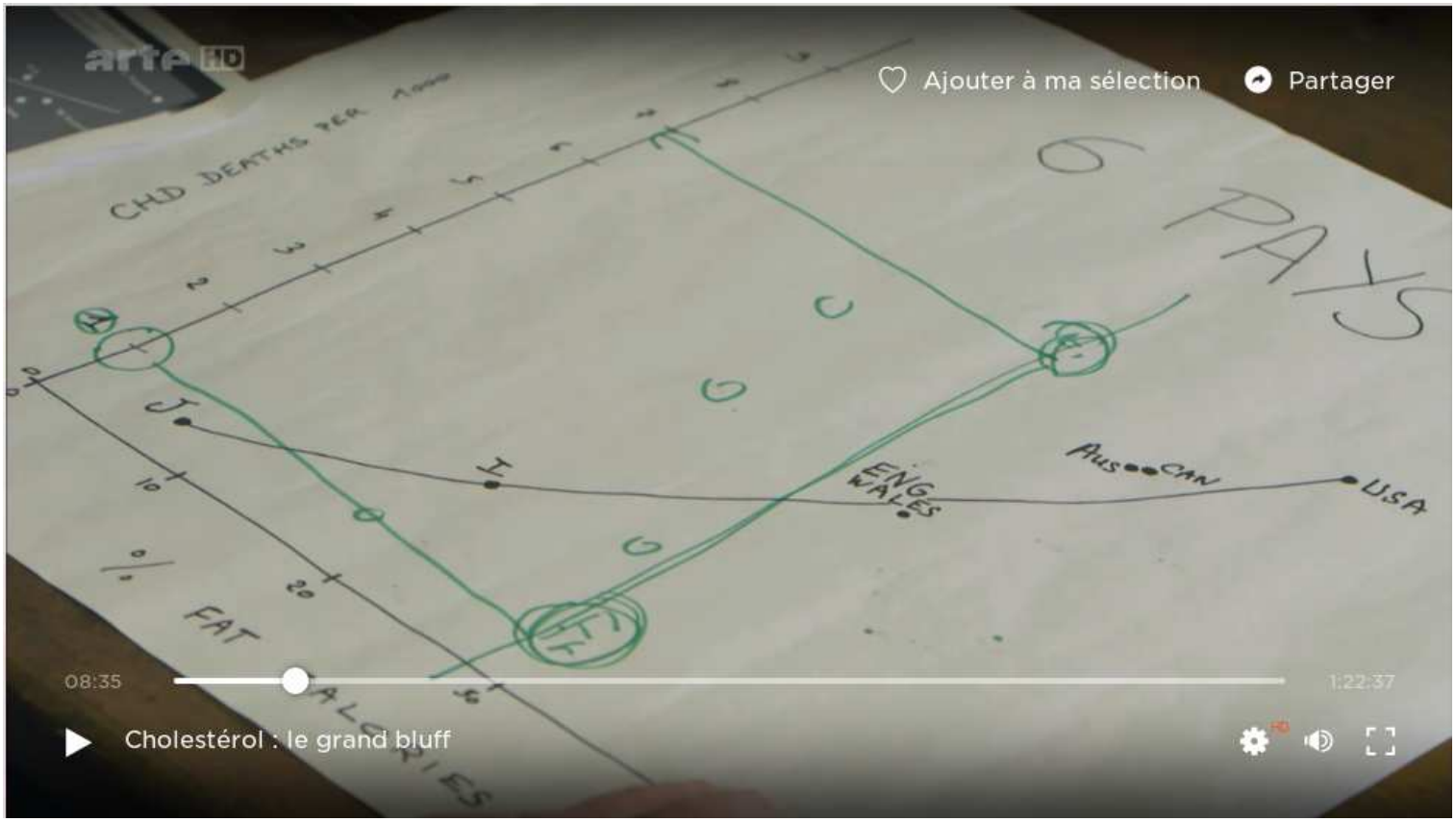
Outline

A vivid debate: Cholesterol and Statins

Cholesterol: le grand bluff (Arte, 18/10/2016 @ 20h50)

A vivid debate: Cholesterol and Statins

Cholesterol: le grand bluff (Arte, 18/10/2016 @ 20h50)



"Careful" selection of data and influence from the industry ☹

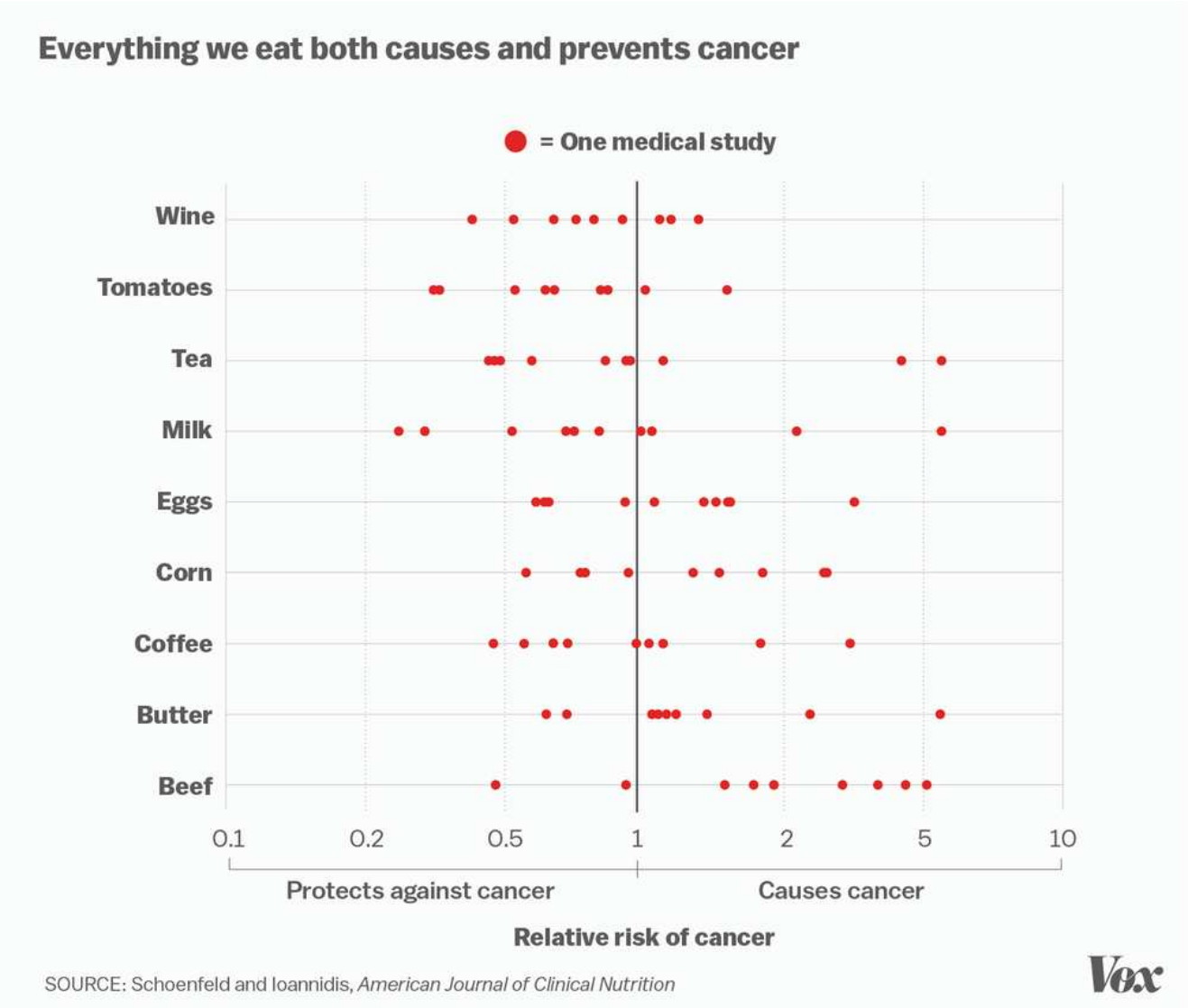
Inconsistencies

What should we eat then?

Is everything we eat associated with cancer? A systematic cookbook review, Schoenfeld and Ioannidis, *Amer. Jour. of Clinical Nutrition*, 2013.

Inconsistencies

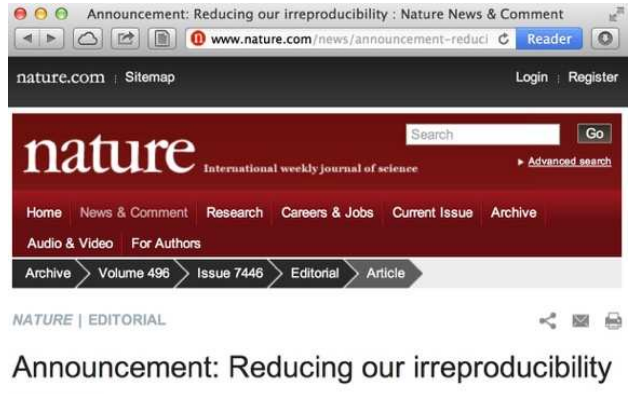
What shoul



Is everything we eat associated with cancer? A systematic cookbook review, Schoenfeld and Ioannidis, *Amer. Jour. of Clinical Nutrition*, 2013.

Public evidence for a Lack of Reproducibility

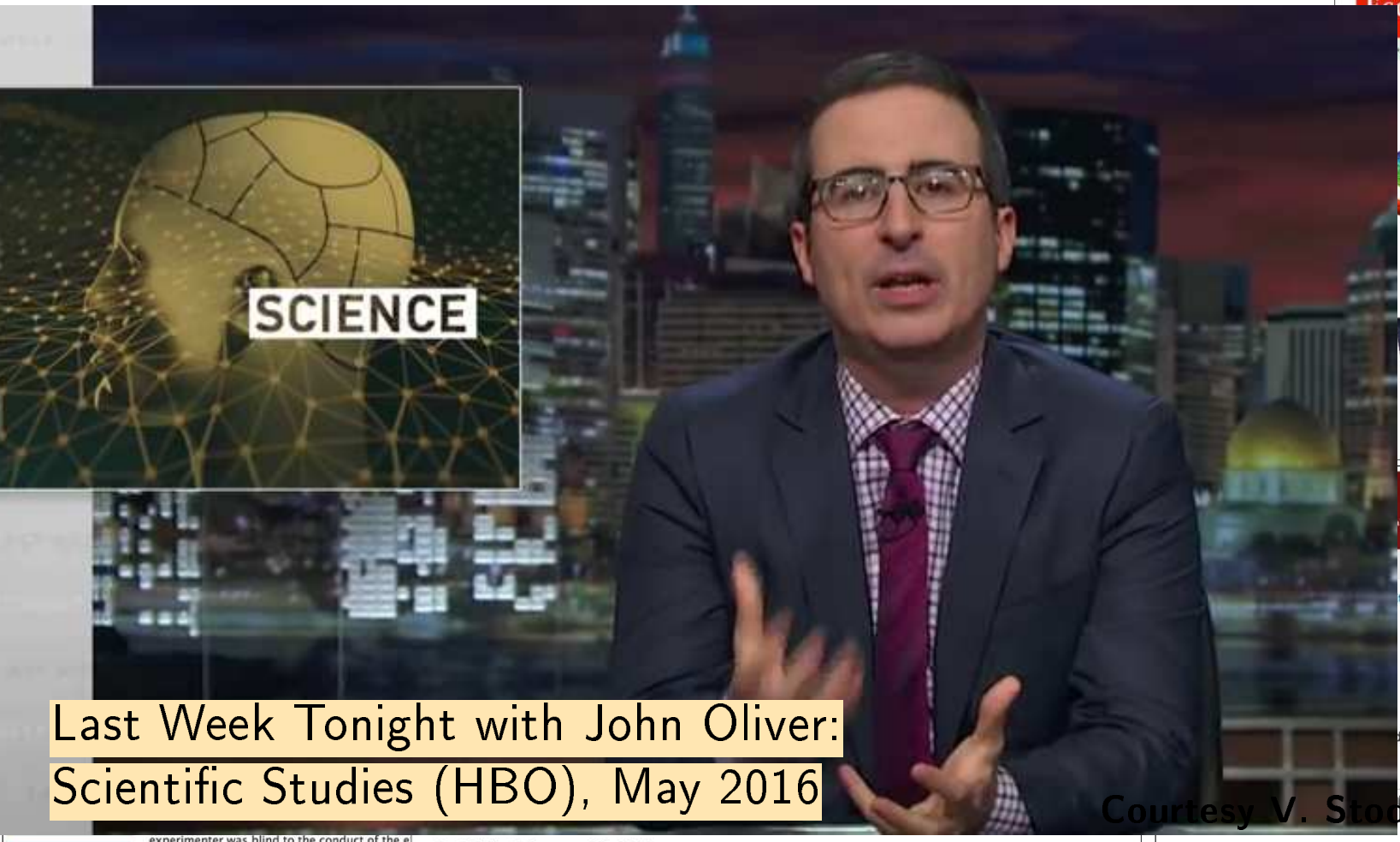
- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005.
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010



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



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

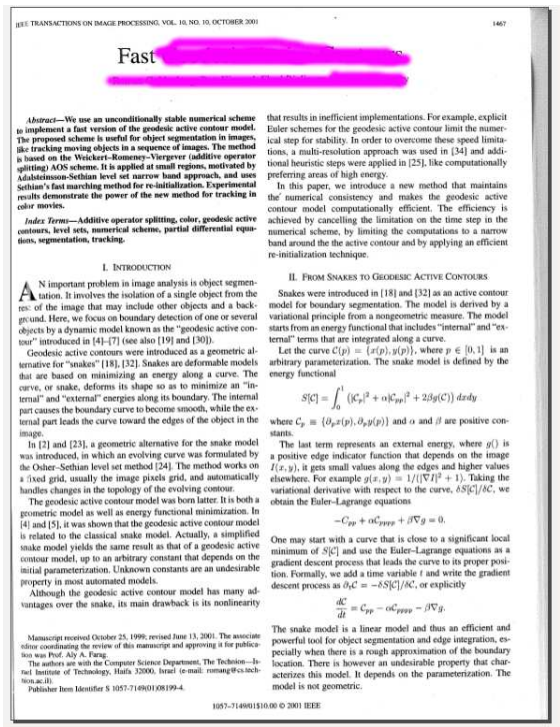
1. Have you ever tried to reproduce some research results ?
2. Have you ever failed ?

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 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.

1057–7149/01$10.00 © 2001 IEEE

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

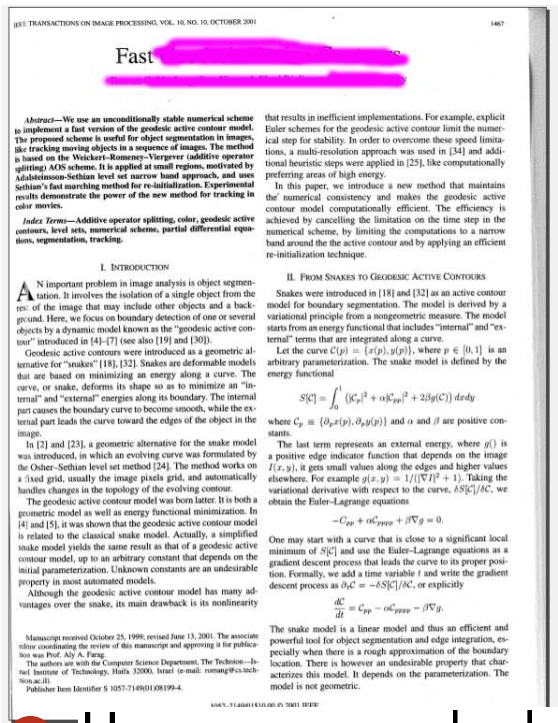
Courtesy of Enric Meinhardt-Llopis, CANUM 2016

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 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*

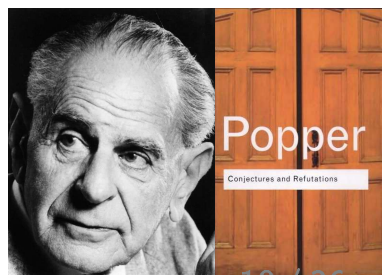
In a letter to his rival Robert Hooke

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?

The Duke University scandal with scientific misconduct on lung cancer

- *Nature Medicine* - 12, 1294 - 1300 (2006) Genomic signatures to guide the use of chemotherapeutics, by Anil Potti and 16 other researchers from Duke University and University of South Florida
- Major commercial labs licensed it and were about to start using it before two statisticians discovered and publicized its faults

> Dr. Baggerly and Dr. Coombes found errors almost immediately. Some seemed careless — moving a row or a column over by one in a giant spreadsheet — while others seemed inexplicable. The Duke team shrugged them off as "clerical errors."

> The Duke researchers continued to publish papers on their genomic signatures in prestigious journals. Meanwhile, they started 3 trials using the work to decide which drugs to give patients.

- Retractions: January 2011. Ten papers that Potti coauthored in prestigious journals were retracted for varying reasons **Courtesy of Adam J. Richards**

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

Definitely

Similar scandals in social psychology, stem cells, organic transistors (The Economics of Replication).

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!

Definitely

Similar scandals in social psychology, stem cells, organic transistors (The Economics of Replication).

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!

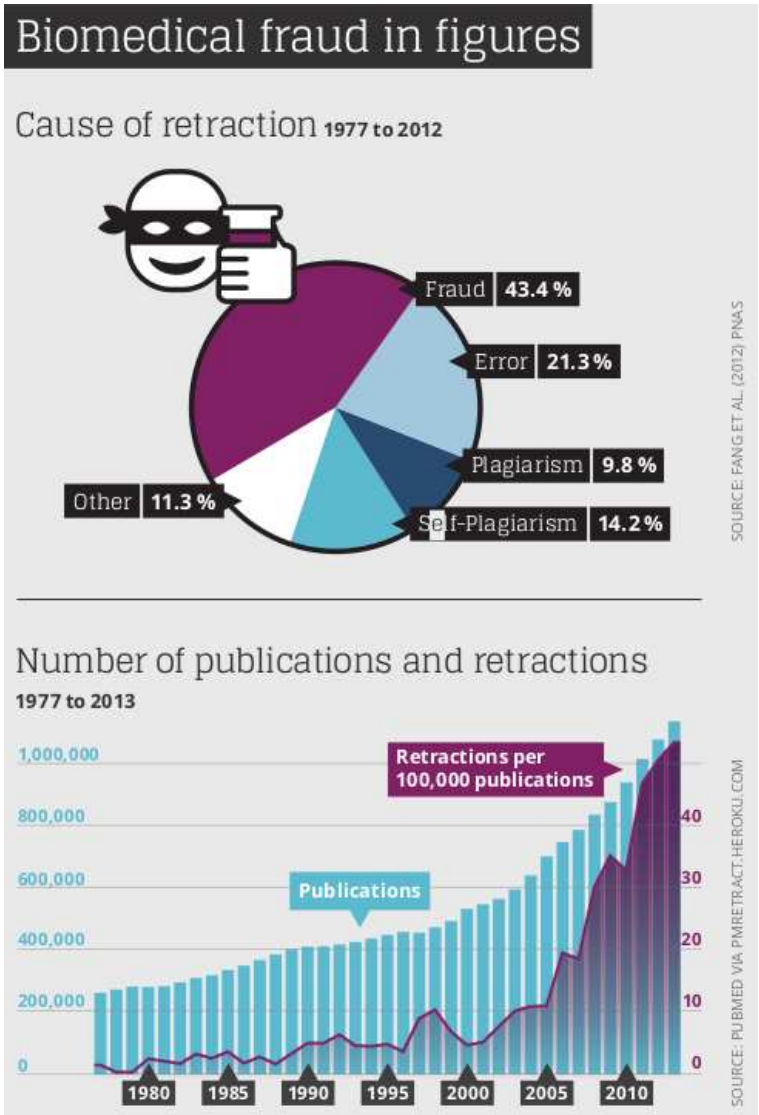
We should avoid witch-hunt

- August 5, 2014, Yoshiki Sasai (stem cell, considered for Nobel Prize) hanged in his laboratory at the RIKEN (Japan). Fraud suspicion...
- In 1986, a young postdoctoral fellow at MIT accused her director, Thereza Imanishi-Kari, of falsifying the results of a study published in Cell and co-signed by the Nobel laureate David Baltimore. [..] Declared guilty, Univ. presidency resignation, and finally cleared. On hold carreer for ten years...

Scientific fraud is bad but let's be careful Have a look at the wikipedia *list of academic scandals*. On a totally different aspect, do not forget to also have a look at the *plagiarism* and *paper generation* entries at *having fun with h-index*

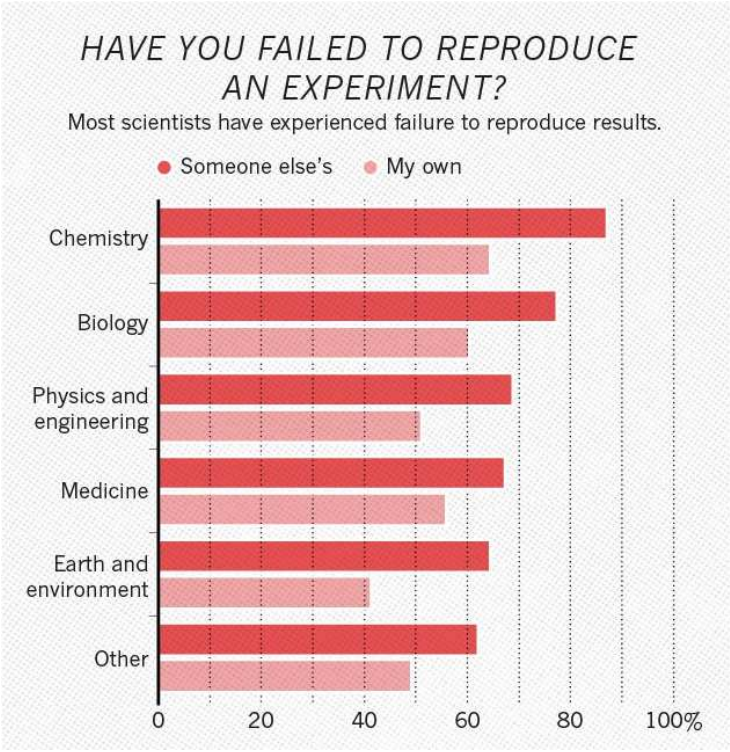
Is Fraud a new phenomenon?

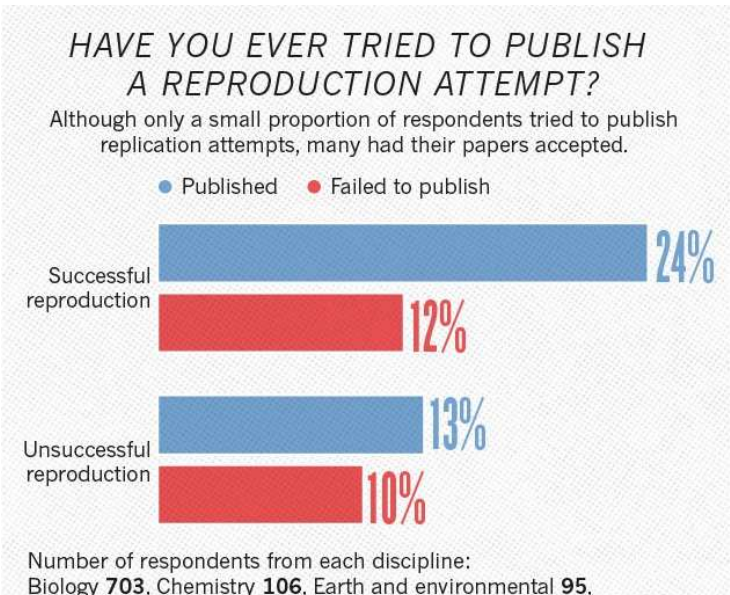
The Battle against Scientific Fraud in the CNRS International Magazine



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

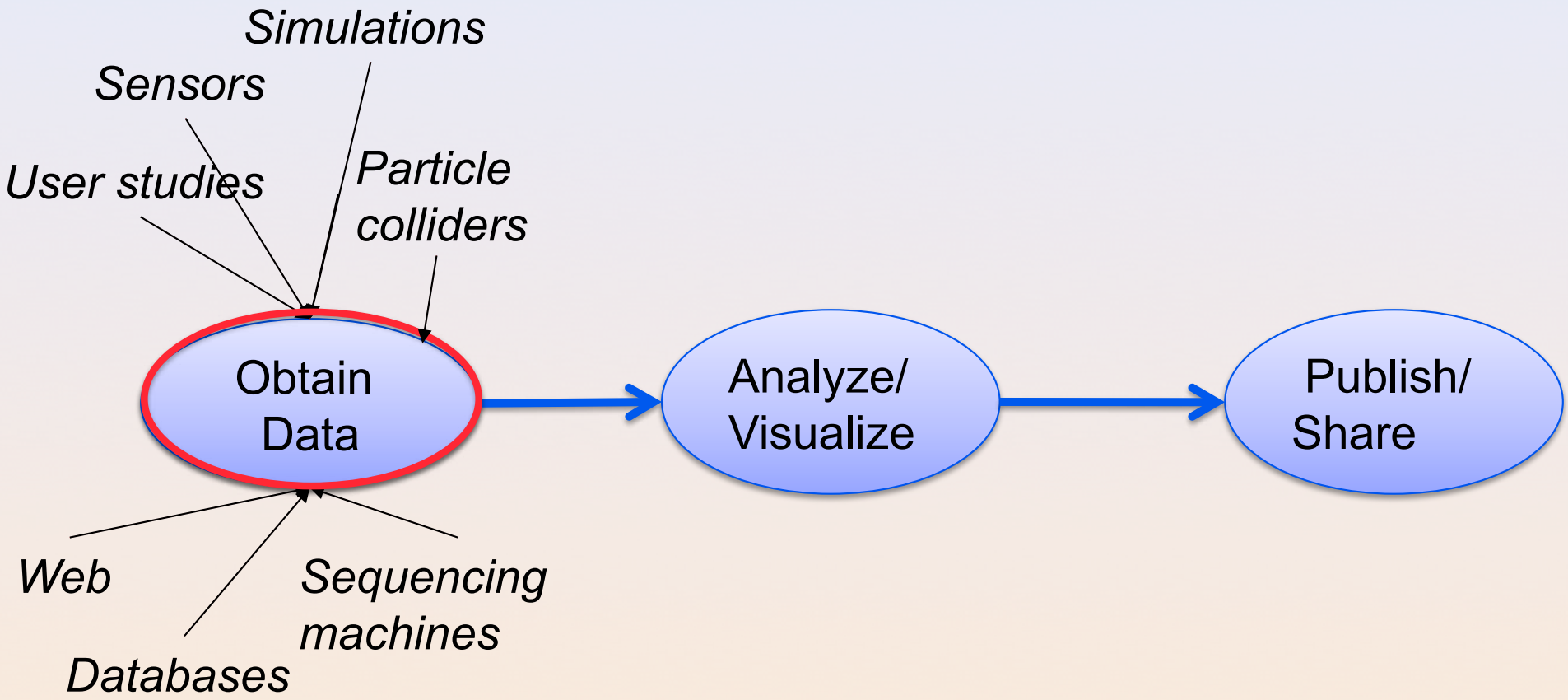
Methodological or technical causes

- 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

Social causes

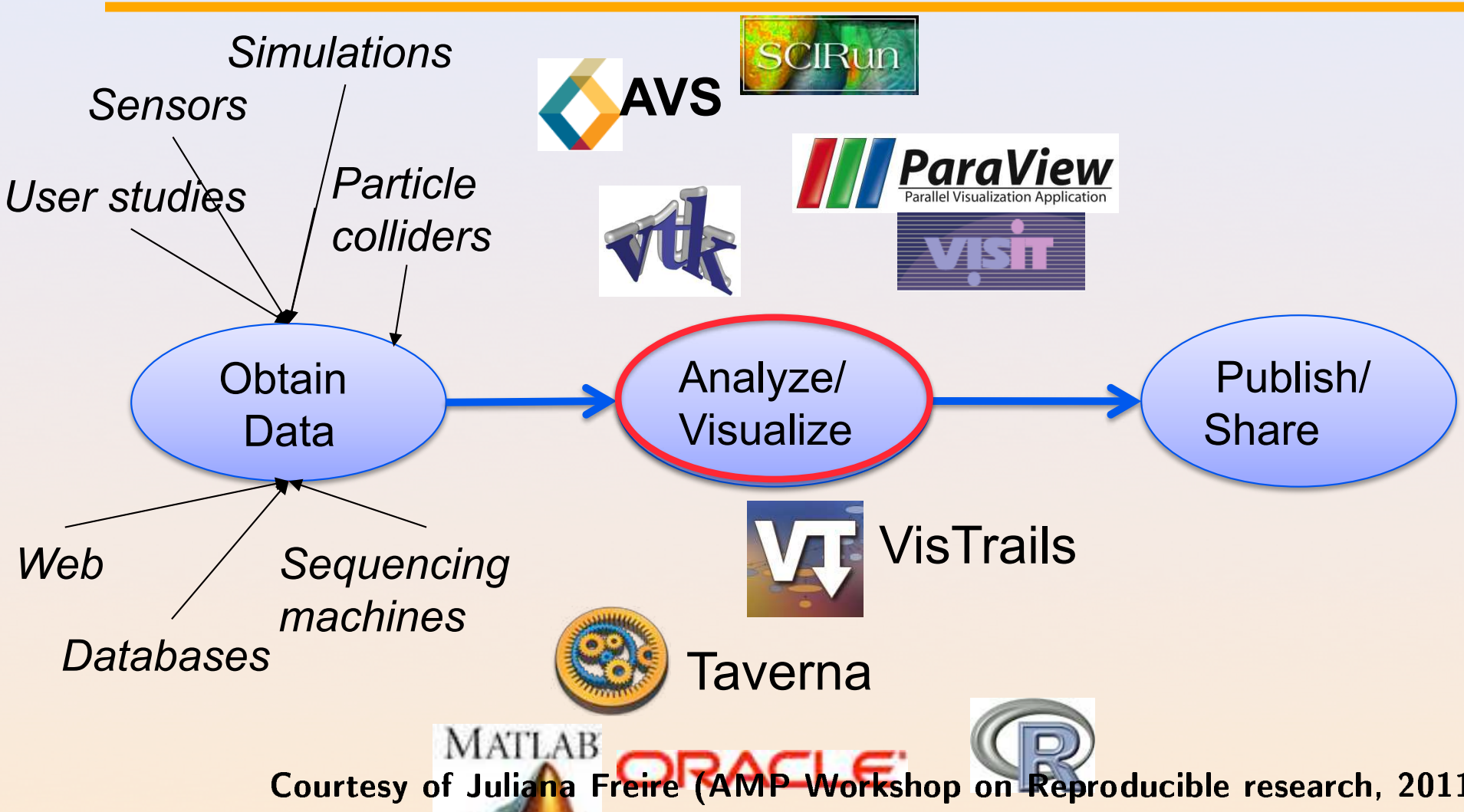
- Fraud, conflict of interest (pharmaceutic)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1+ million articles per year!

Science Today: Data Intensive



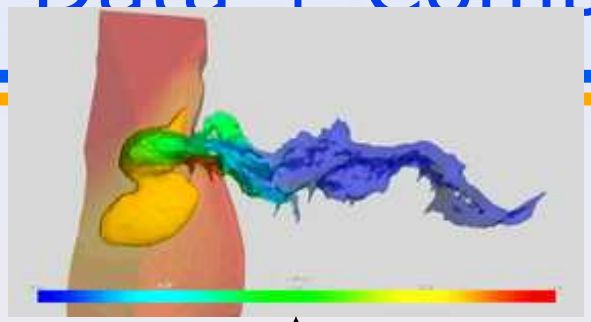
Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

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

Obtain Data

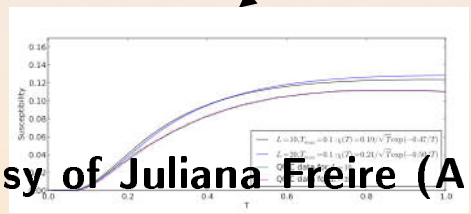
Analyze/ Visualize

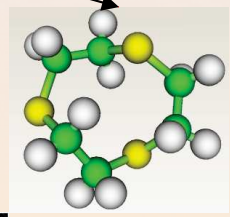
Publish/ Share

Web

Sequencing machines

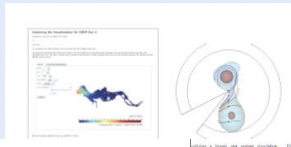
Databases

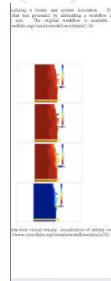




Courtesy of Juliana Freire (AM eproducible research, 2011)

Science Today: Data + Computing Inte

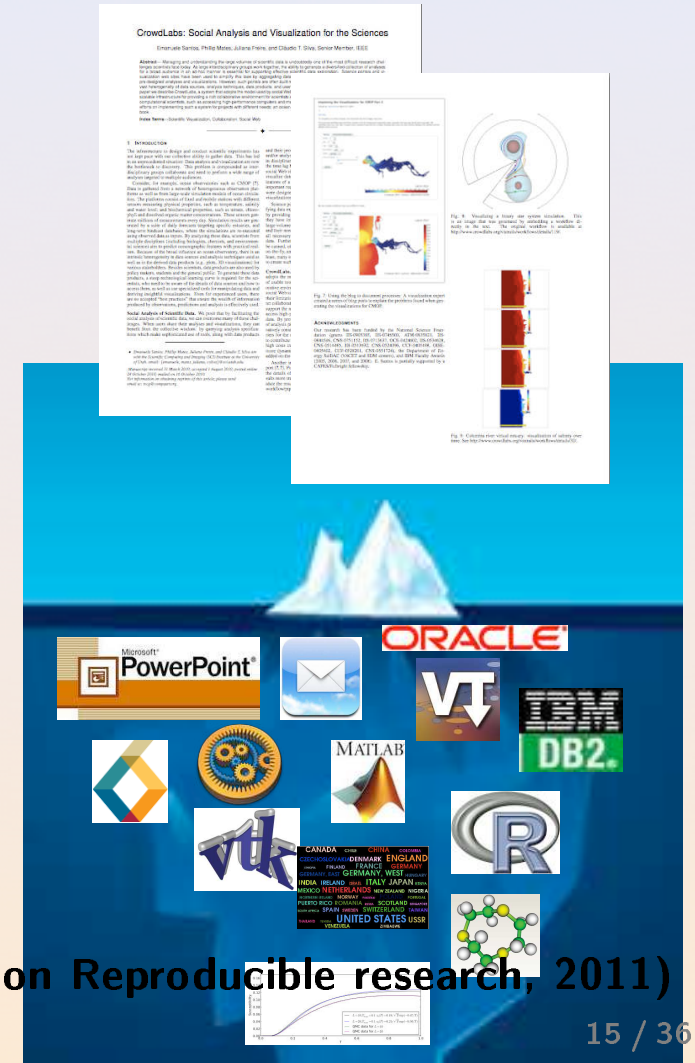




Courtesy of Juliana Freire (AMP Workshop on Reproducible research, 2011)

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

Defining Computer Science (Dowek and others)

Computing

- A science: Science of artificial... but not only
- A technology, an industry (Hardware, software, network, services,...)
- Applications: increasing area
  - Scientific impact: computational science (simulation), *big data*
  - Social impact: numerical/Information society

Defining Computer Science (Dowek and others)

Computing

- A science: Science of artificial... but not only
- A technology, an industry (Hardware, software, network, services,...)
- Applications: increasing area
  - Scientific impact: computational science (simulation), *big data*
  - Social impact: numerical/Information society

Computer Science

Concepts

- **Information**: Representation, communication, compression,...
- **Algorithm**: Operative process
- **Programming Language**: link between levels of abstraction
- **Architecture** (Computing Engine): abstraction of the physical world
- **Human in the loop**

Methods: Back and forth between theory and experimentation. Automatic abstraction transform. Self-generated tools.

All this is about Natural Sciences. Should we care ?

Computer Science is young and inherits from Mathematics, Engineering, Nat. Sciences, Linguistic, ...

Purely theoretical scientists whose practice is close to mathematics may not be concerned (can't publish a math article without releasing the proofs).

> *Computer science is not more related to computers than Astronomy to telescopes*
>
> – *Dijkstra*

Right, why should we care about computers? They are deterministic machines after all, right? 😉

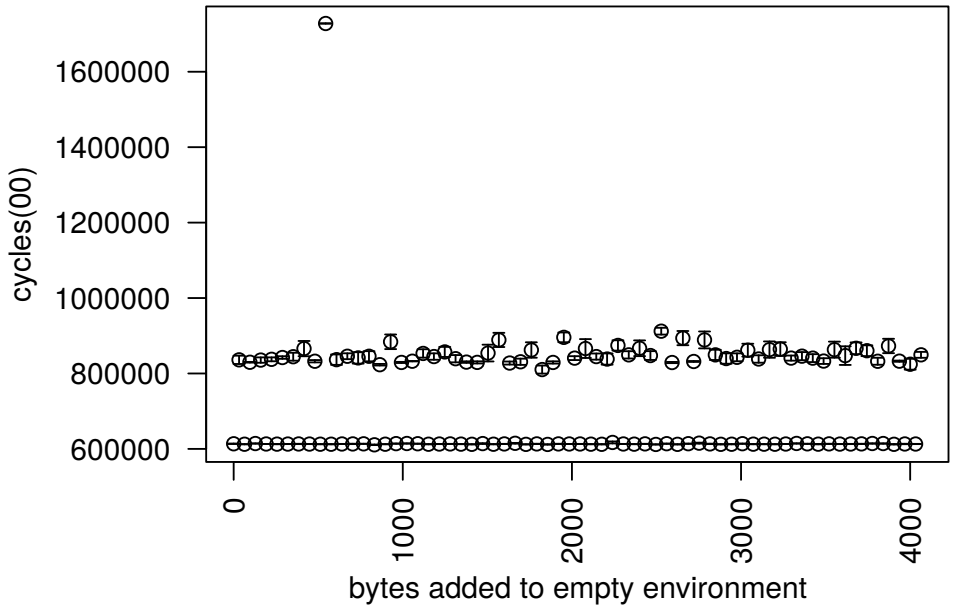
Model $\neq$ **Reality**. Although designed and built by human beings, computer systems are so complex that mistakes easily slip in...

Experimenting with computers

Machines are real!

Brendan Gregg: Shouting in the data center

Machines are complicated



Mytkowicz et al. Producing wrong data without doing anything obviously wrong!
ACM SIGPLAN Not. 44(3), March 2009

Our reality evolves!!! The hardware keeps evolving so most results on old platforms quickly become obsolete (although, we keep building on such results 😉).

- We need to regularly revisit and allow others to build on our work!

Computer performance ? Well, I design algorithms!

- "Real" problems are all NP-hard, Log-APX, etc.
- Real workload = ~~NP-completeness proof widgets~~, regularities and properties (difficult to formally state but that should be exploited)

Algorithms are evaluated on particular workloads that impact
both their running time and the quality of the solutions

Computer performance ? Well, I design algorithms!

- "Real" problems are all NP-hard, Log-APX, etc.
- Real workload = ~~NP-completeness proof widgets~~, regularities and properties (difficult to formally state but that should be exploited)

Algorithms are evaluated on particular workloads that impact both their running time and the quality of the solutions
Machine Learning: Trouble at the lab, The Economist 2013

> *According to some estimates, three-quarters of published scientific papers in the field of machine learning are bunk because of this "overfitting".* – *Sandy Pentland (MIT)*

Computer performance ? Well, I design algorithms!

- "Real" problems are all NP-hard, Log-APX, etc.
- Real workload = ~~NP-completeness proof widgets~~, regularities and properties (difficult to formally state but that should be exploited)

Algorithms are evaluated on particular workloads that impact both their running time and the quality of the solutions

Machine Learning: Trouble at the lab, The Economist 2013

> *According to some estimates, three-quarters of published scientific papers in the field of machine learning are bunk because of this "overfitting".* – *Sandy Pentland (MIT)*

Image Processing: True horror stories, E. Meinhardt-Llopis, CANUM 2016

- *The proposed multigrid algorithm converges to the solution of the problem in O(N)* using biharmonic functions
- Surprisingly, our naive multi-scale Gauss-Seidel converges much faster

Computer performance ? Well, I design algorithms!

- "Real" problems are all NP-hard, Log-APX, etc.
- Real workload = ~~NP-completeness proof widgets~~, regularities and properties (difficult to formally state but that should be exploited)

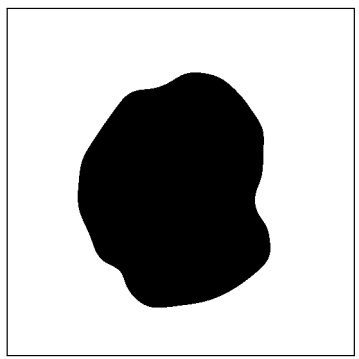
Algorithms are evaluated on particular workloads that impact both their running time and the quality of the solutions

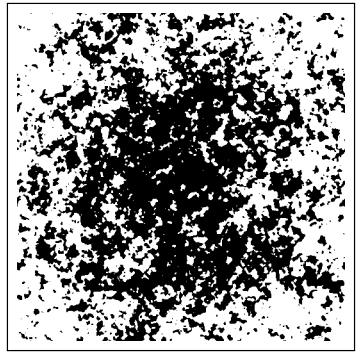
Machine Learning: Trouble at the lab, The Economist 2013

> *According to some estimates, three-quarters of published scientific papers in the field of machine learning are bunk because of this "overfitting".* – *Sandy Pentland (MIT)*

Image Processing: True horror stories, E. Meinhardt-Llopis, CANUM 2016

- *The proposed multigrid algorithm converges to the solution of the problem in O(N)* using biharmonic functions
- Surprisingly, our naive multi-scale Gauss-Seidel converges much faster





All I care about is the algorithm output

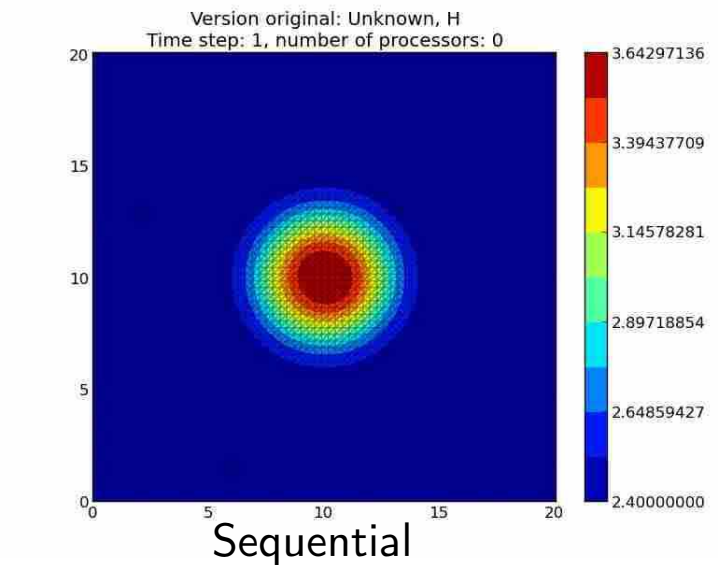
Did I mention we have parallel machines nowadays? 😉

Telemac2D: the simplest gouttedo simulation

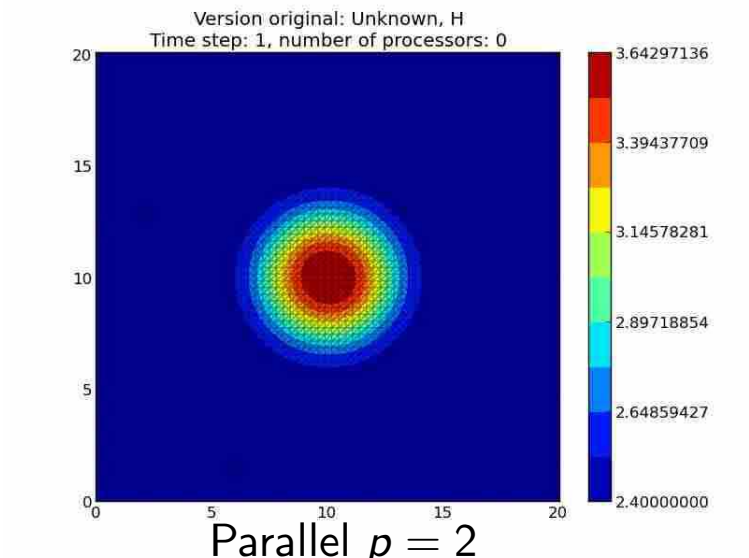
The gouttedo test case

- 2D-simulation of a water drop fall in a square bassin
- Unknown: water depth for a 0.2 sec time step
- Triangular mesh: 8978 elements and 4624 nodes

Expected numerical reproducibility (time step = 1, 2, . . .)



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

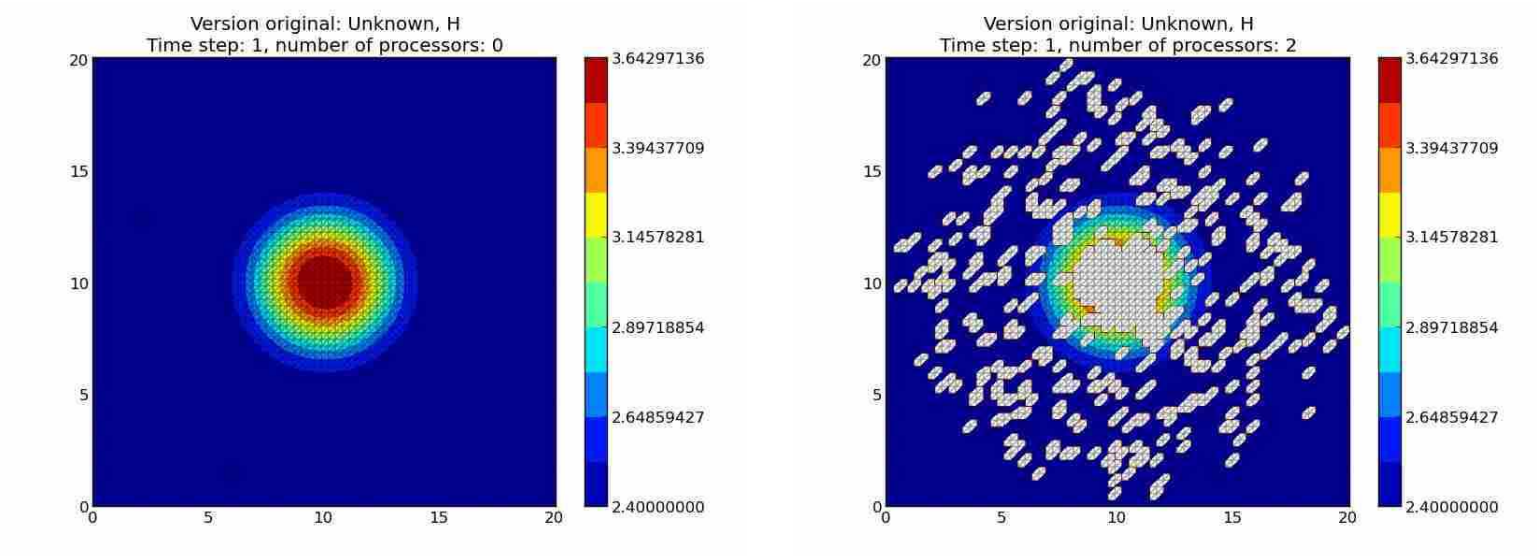
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 1



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

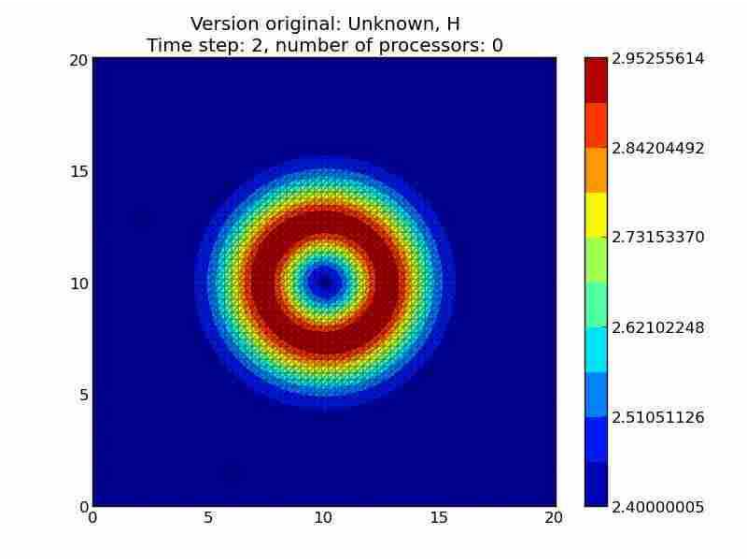
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

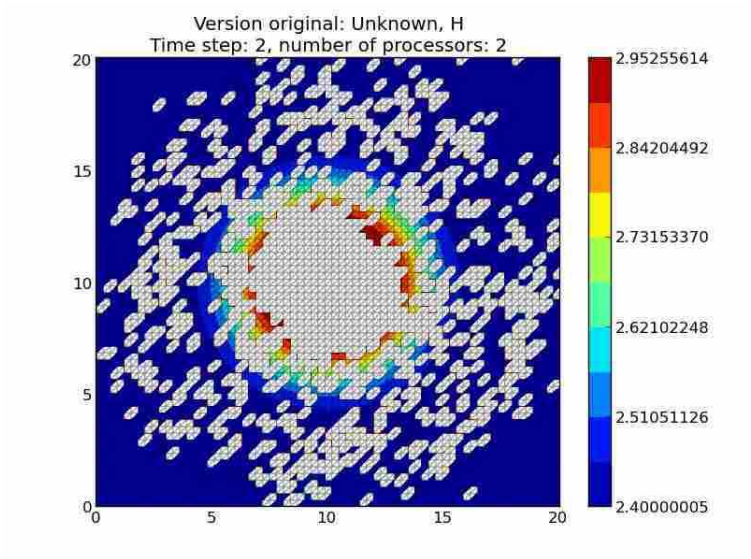
A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 2



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

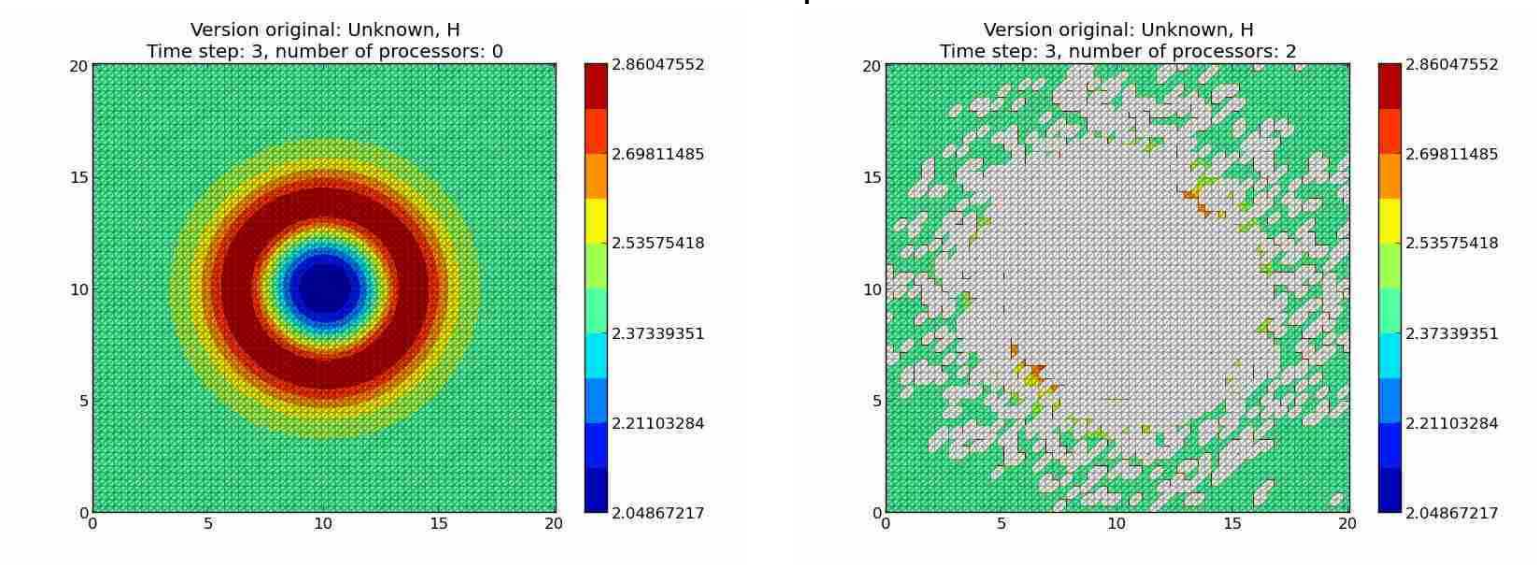
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 3



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

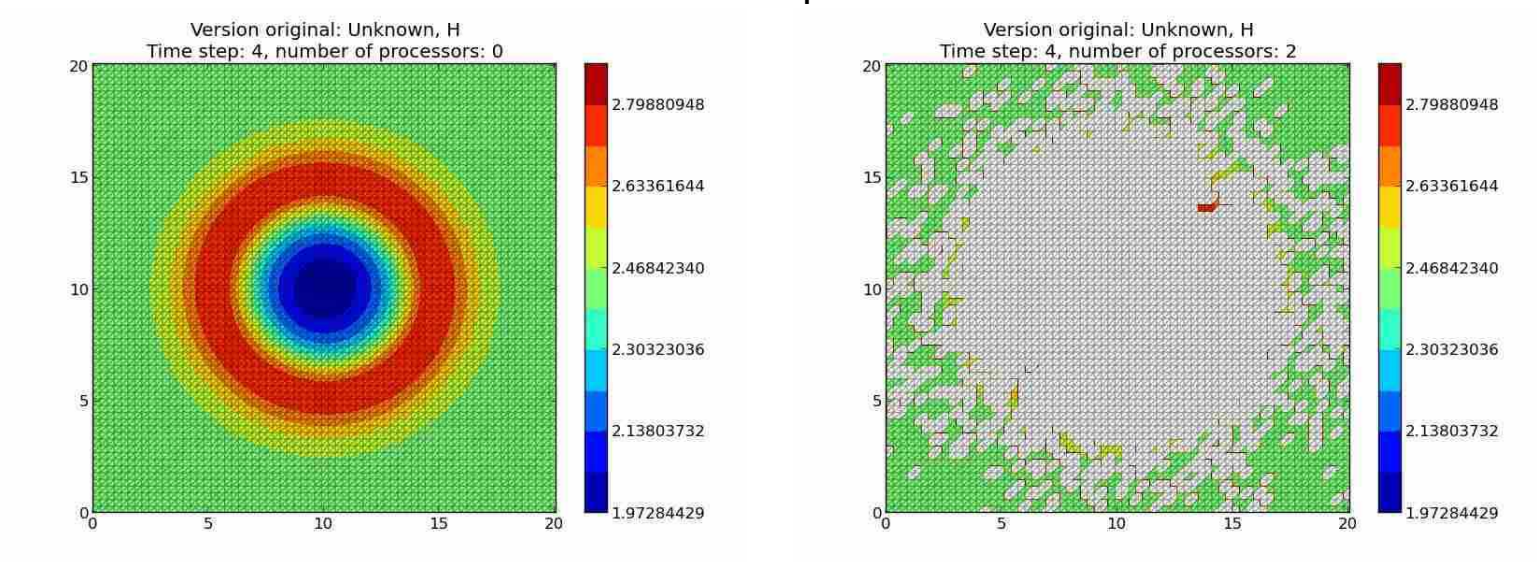
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 4



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

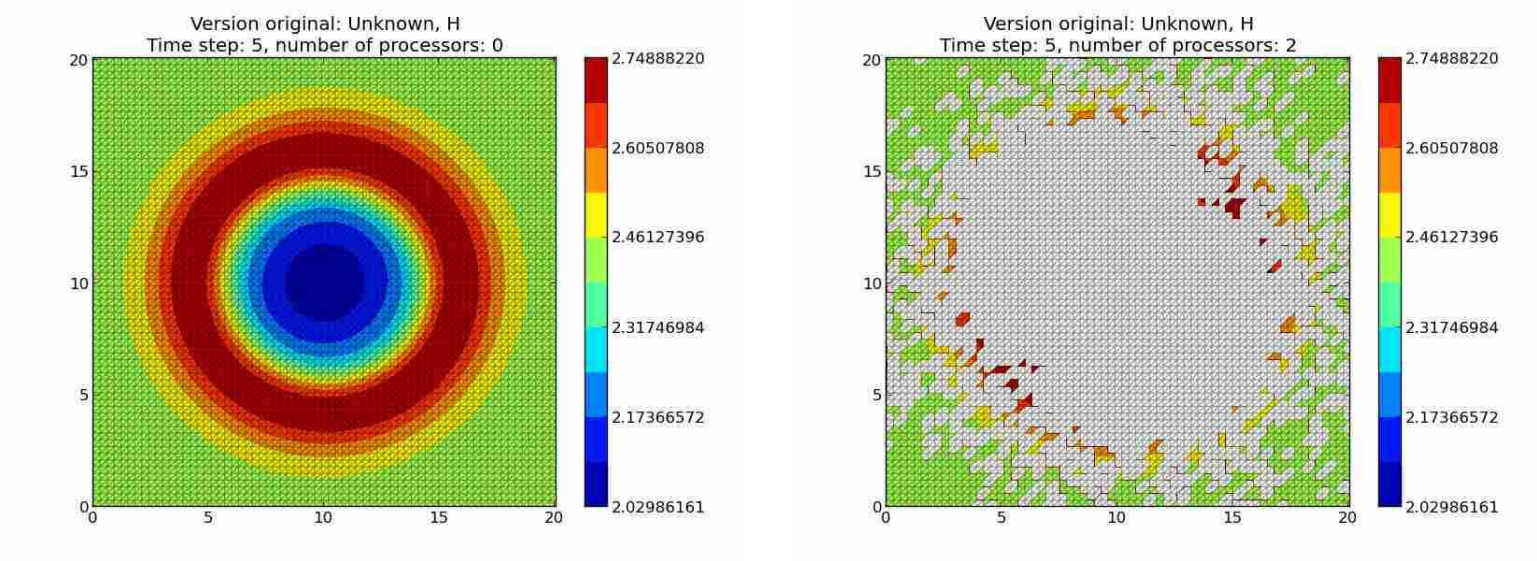
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 5



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

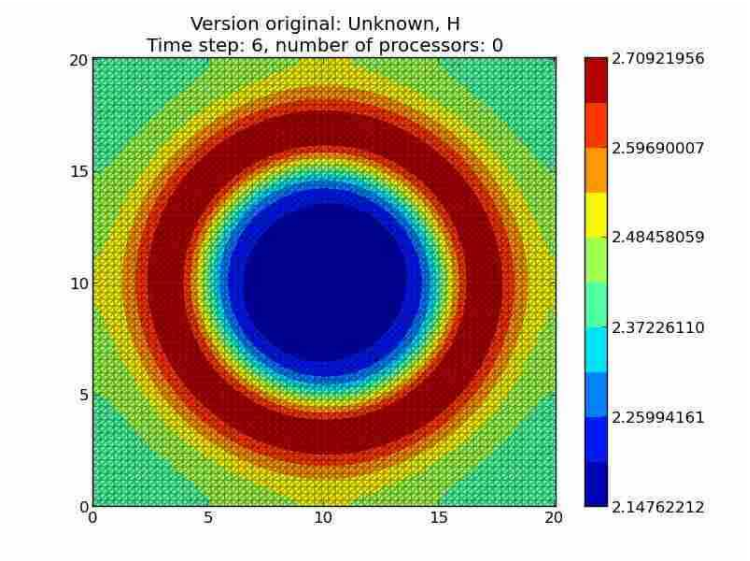
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

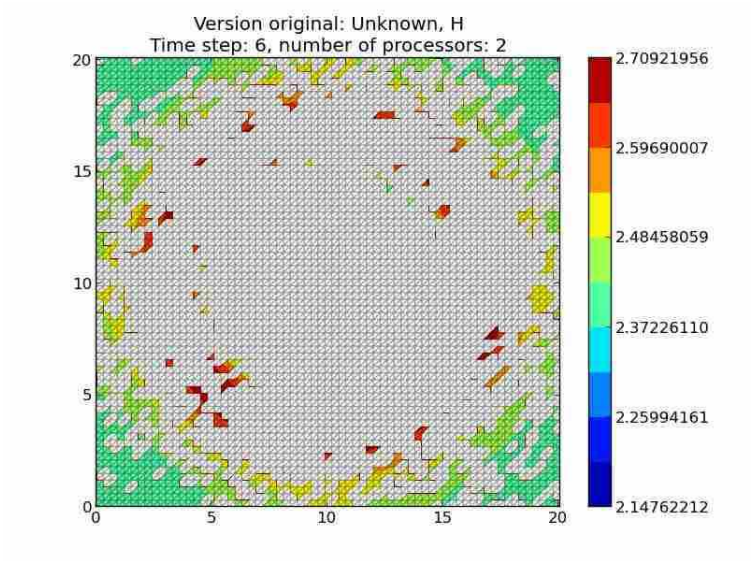
A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 6



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

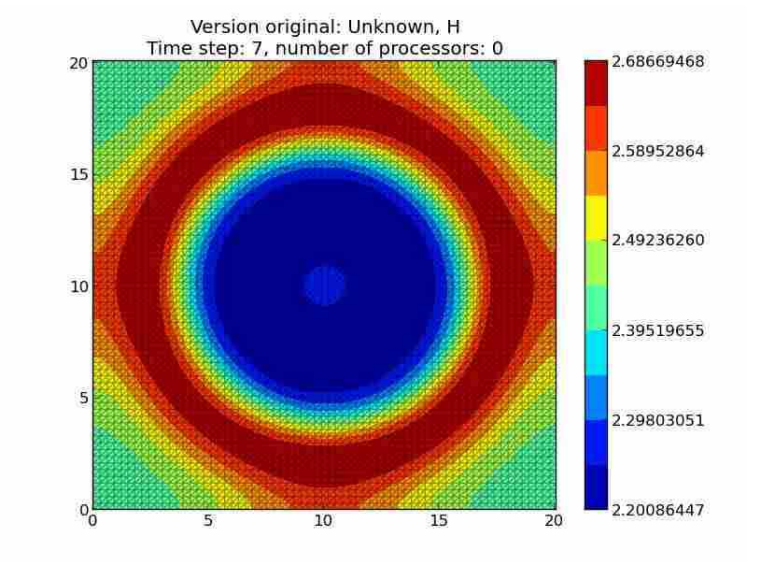
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

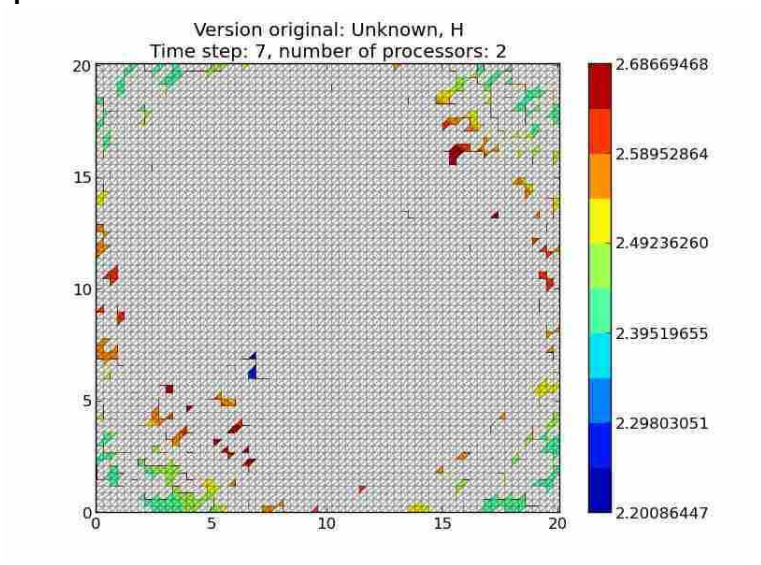
A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 7



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

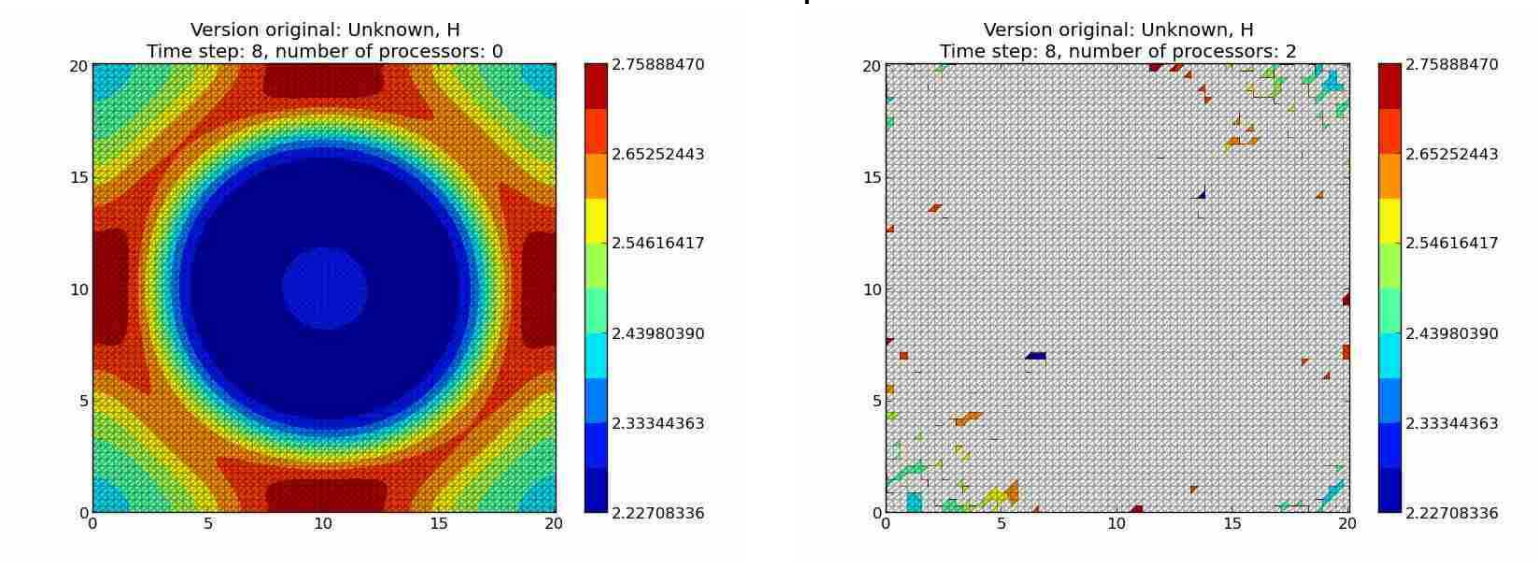
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 8



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

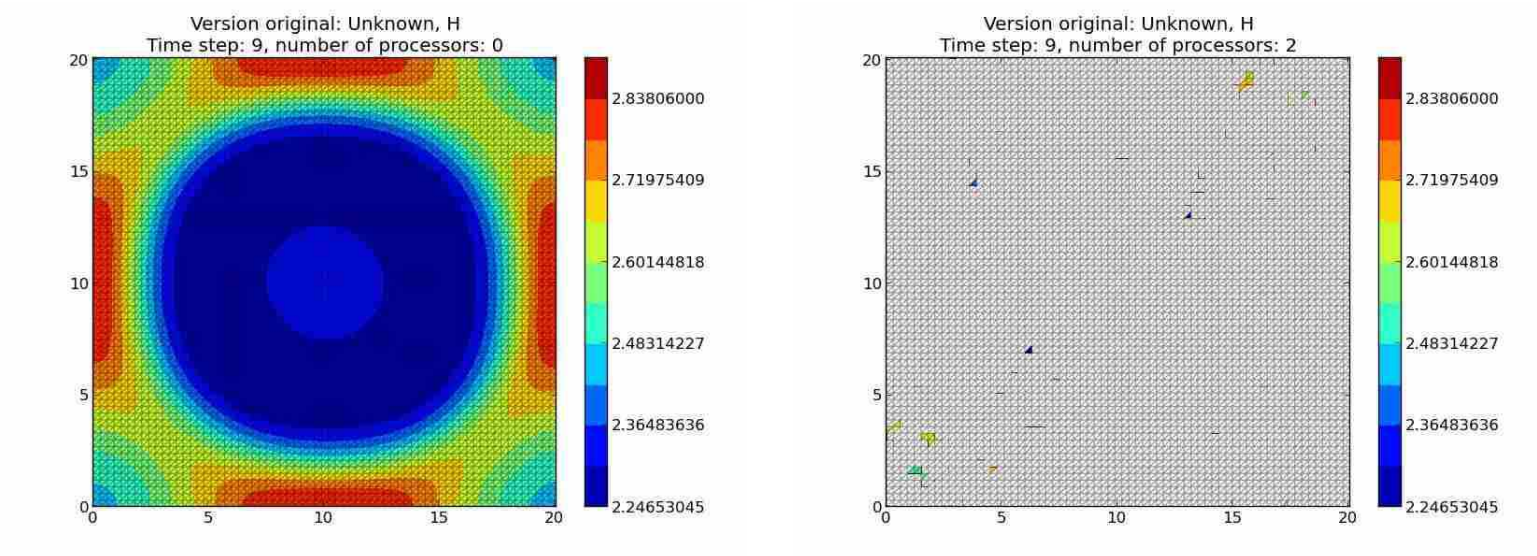
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 9



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

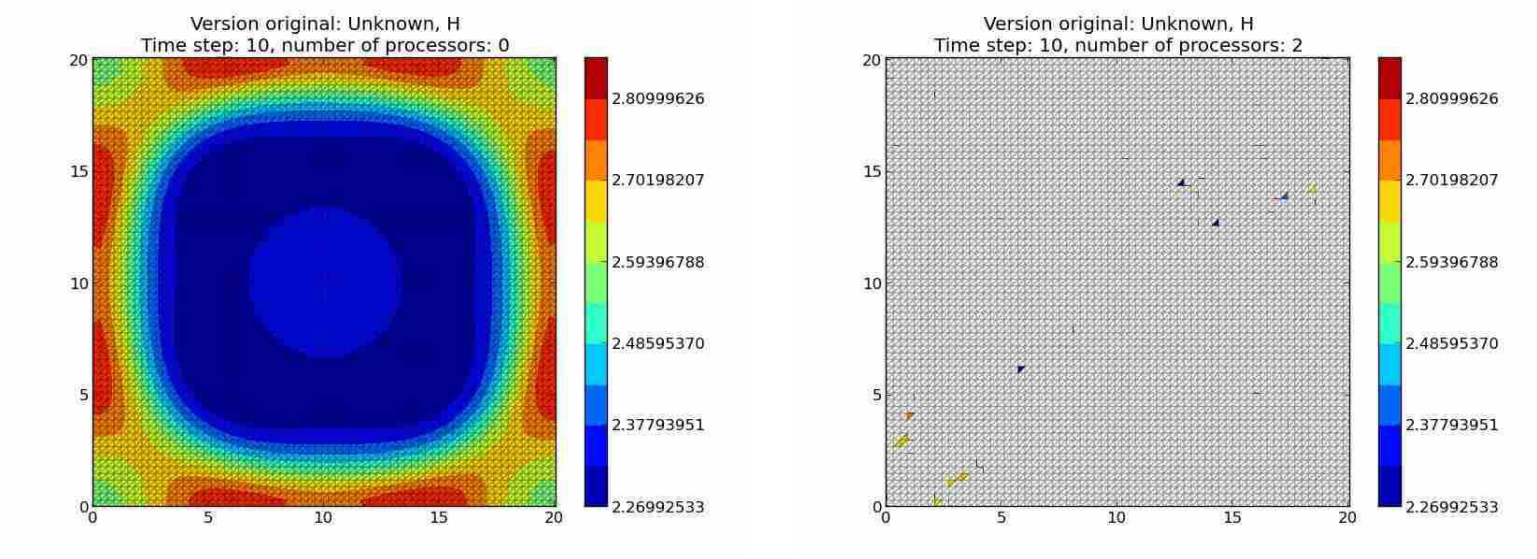
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 10



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

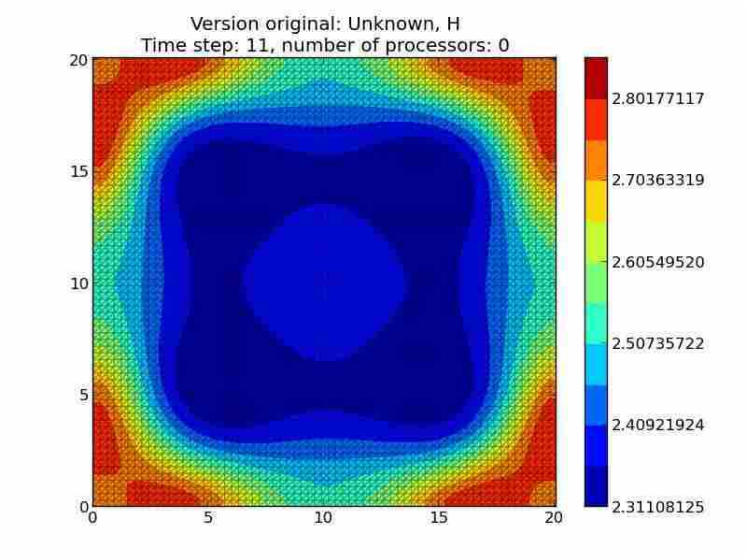
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

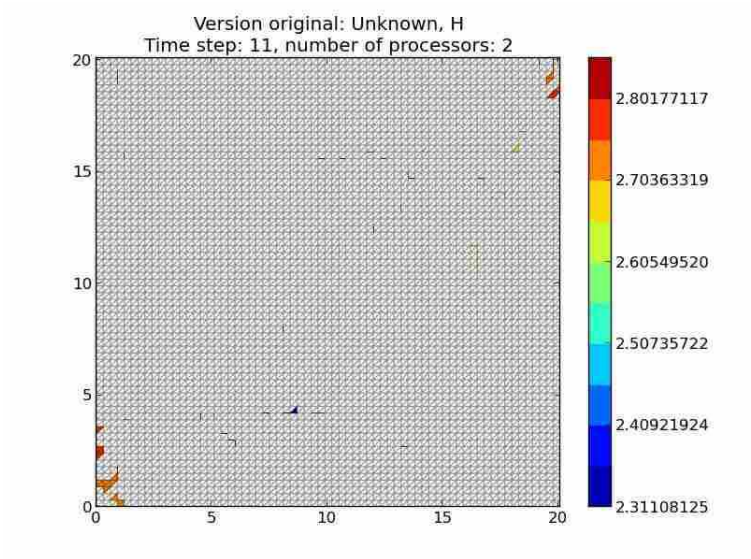
A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 11



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

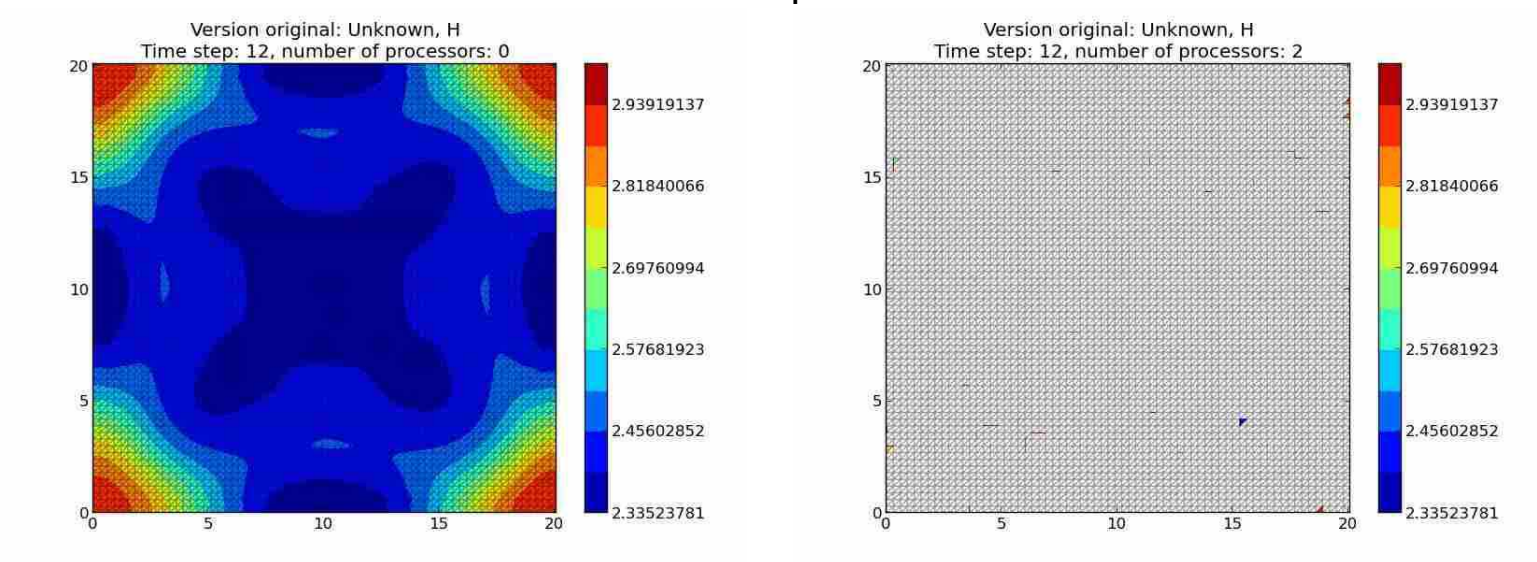
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 12



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

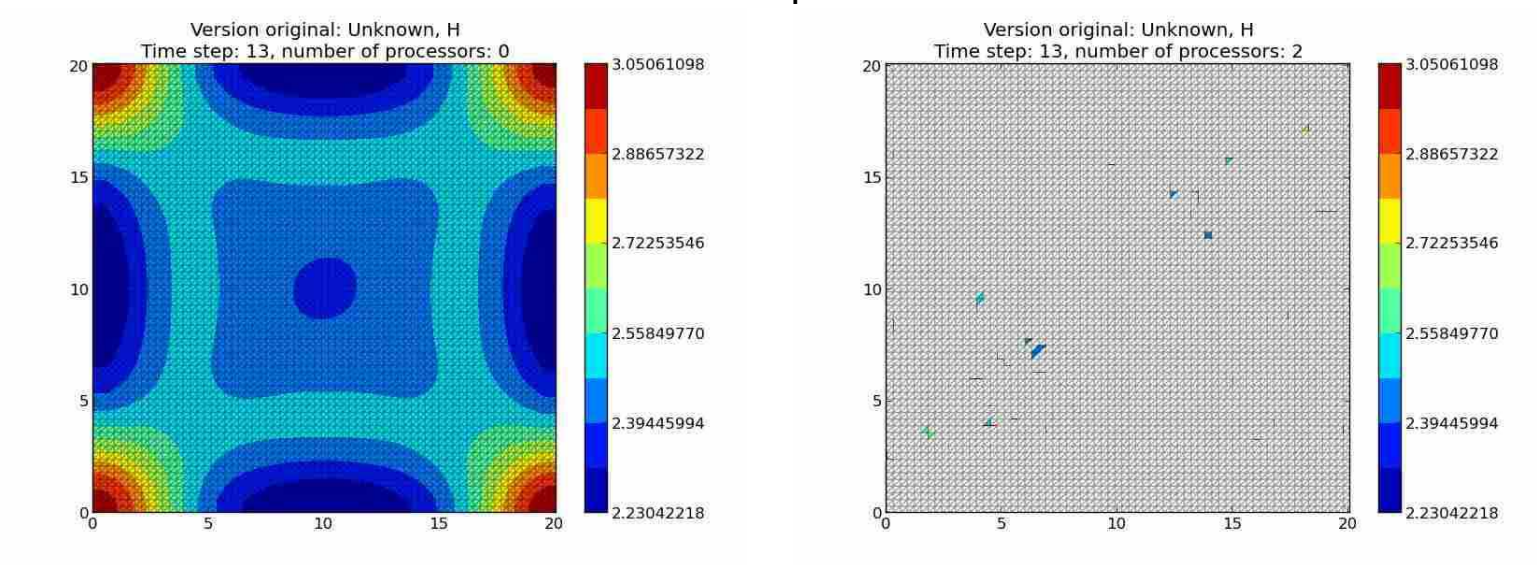
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 13



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

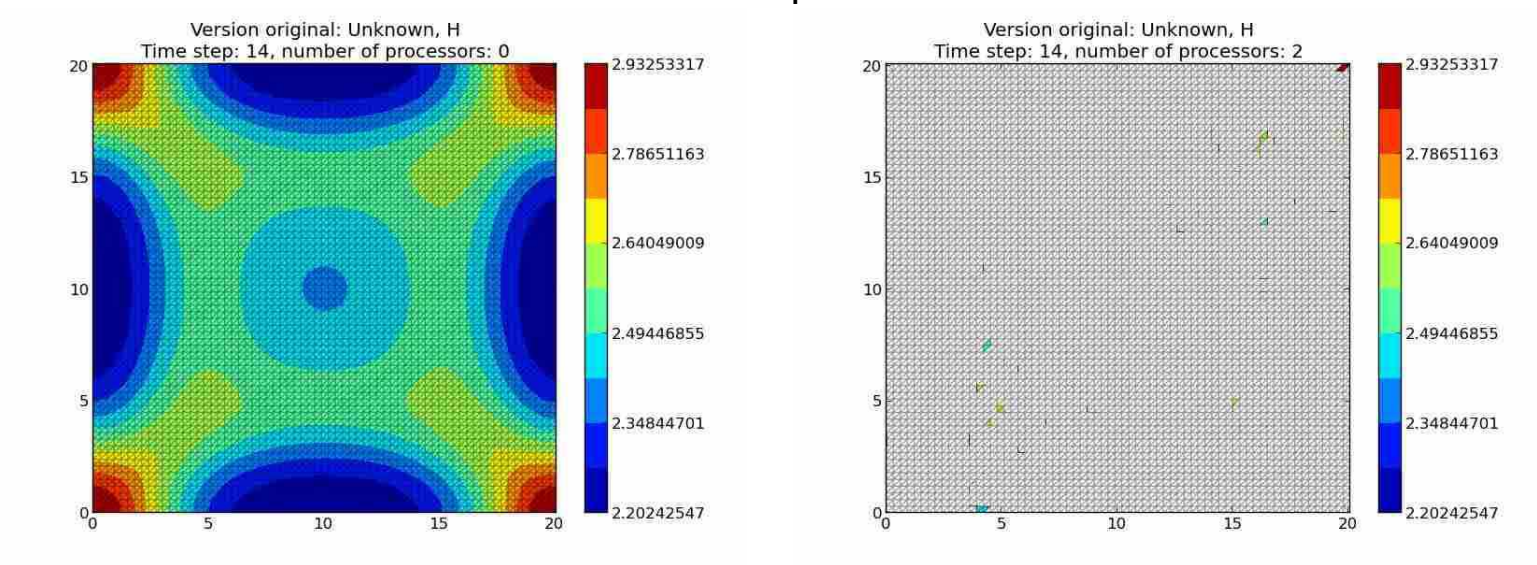
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

A white plot displays a non-reproducible value

Numerical reproducibility?

time step = 14



Sequential

Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

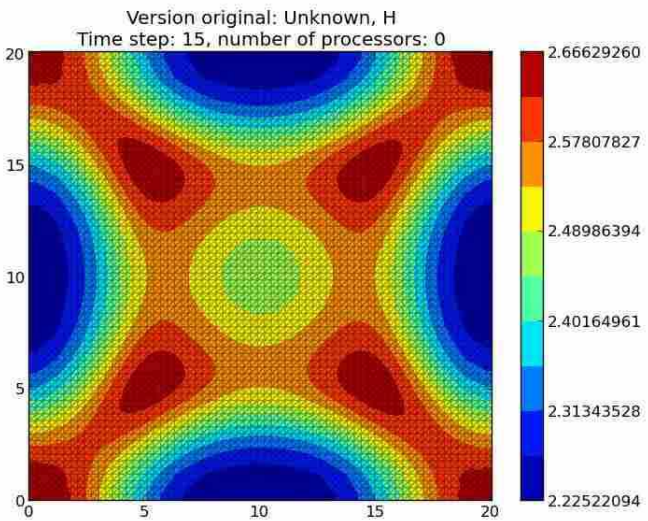
All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

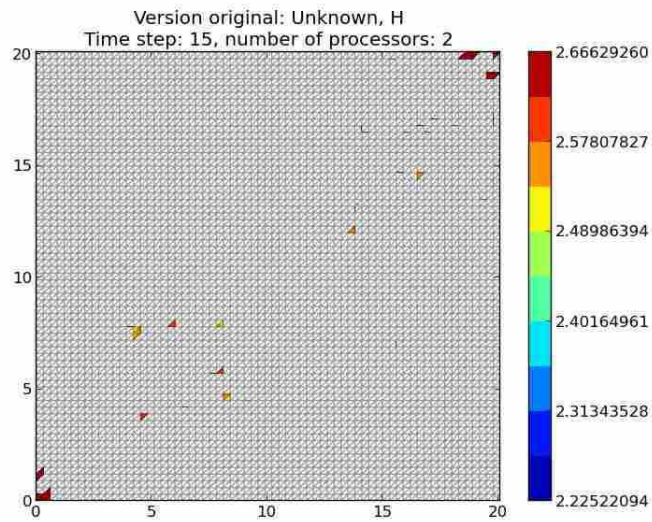
A white plot displays a non-reproducible value

NO numerical reproducibility!

time step = 15



Sequential



Parallel $p = 2$

Courtesy of P. Langlois and R. Nheili

All I care about is the algorithm output

Did I mention we have parallel machines nowadays? 😉

These numerical issues can become quite harmful in real use cases.

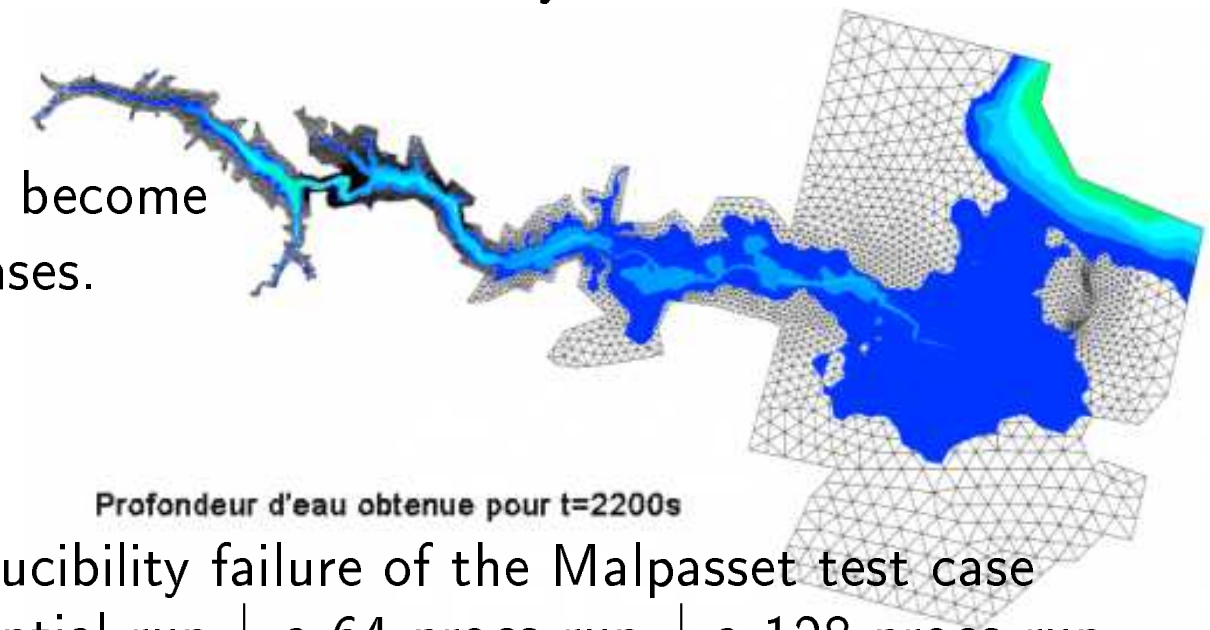


TABLE 1.1: Reproducibility failure of the Malpasset test case

	The sequential run	a 64 procs run	a 128 procs run
depth H	0.3500122E-01	0.2748817E-01	0.1327634E-01
velocity U	0.4029747E-02	0.4935279E-02	0.4512116E-02
velocity V	0.7570773E-02	0.3422730E-02	0.7545233E-02

Numerical reproducibility?: Approximations in the model, in in the algorithm, in its implementation, in its execution.

The whole chain needs to be revisited.

Courtesy of P. Langlois and R. Nheili

Controling what's happening

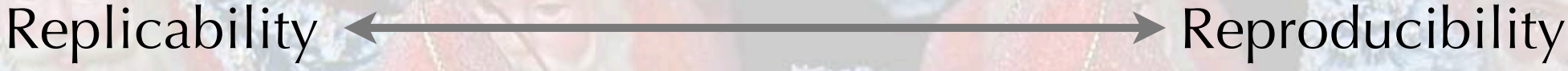
Both the scientific process and the software environment has become extremely complex
Computational biologists and physicist have pushed forward with the development of many tools:

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

And this landscape is rapidly evolving

Outline

Reproducibility: What Are We Talking About?



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

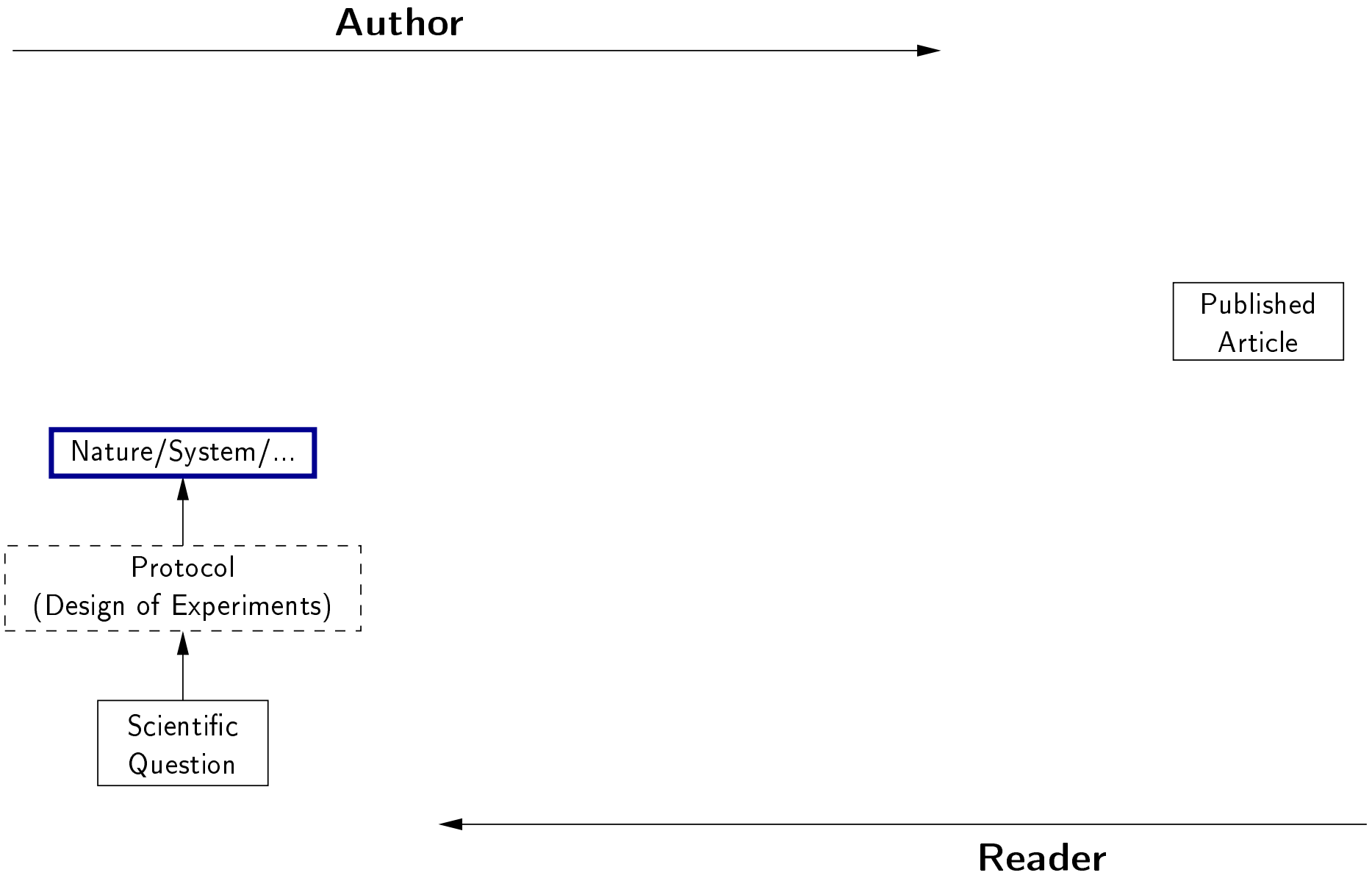
Reproduction using different software, but with access to the original code

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

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

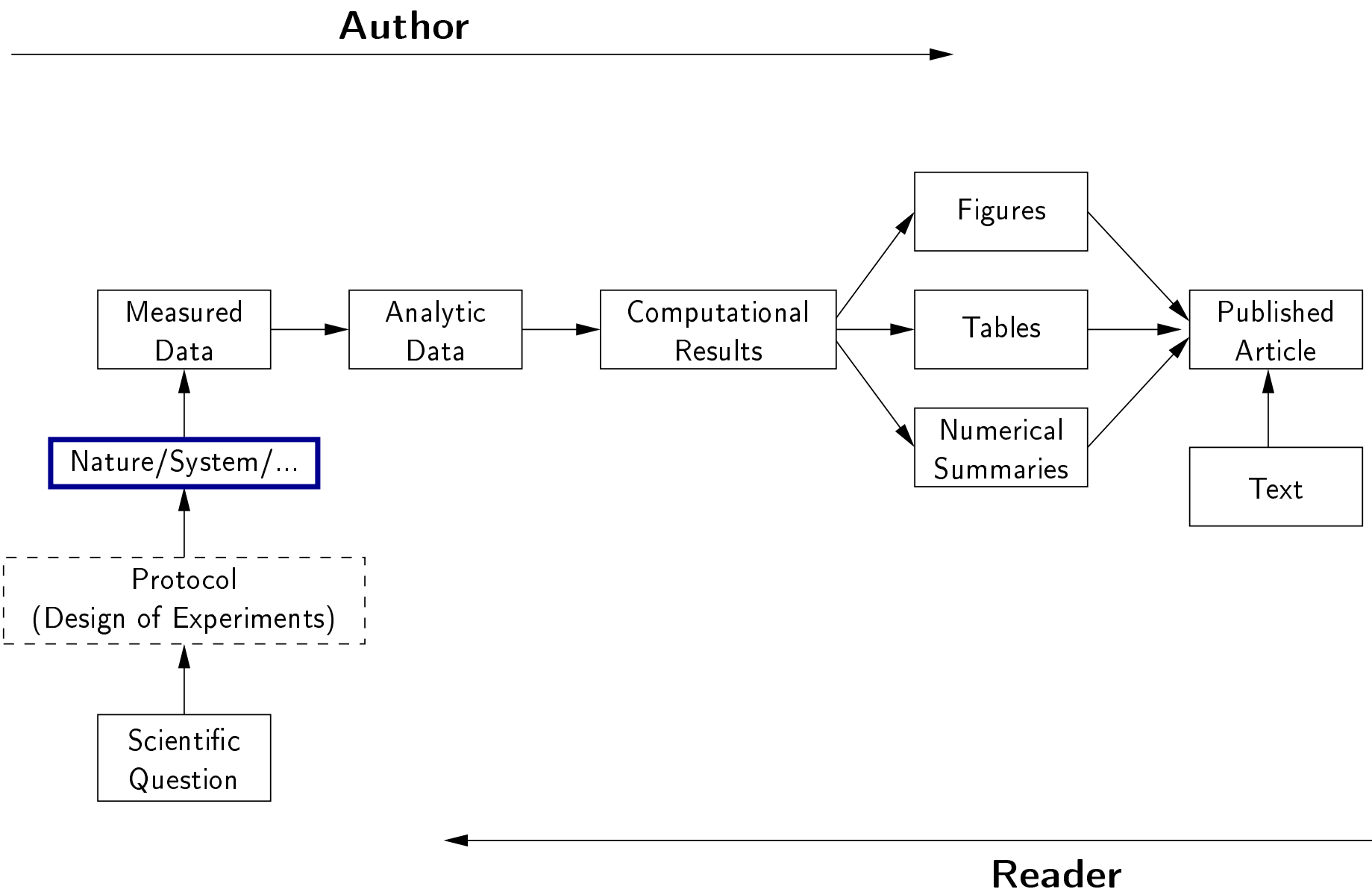
Courtesy of Andrew Davison (AMP Workshop on Reproducible research)

Reproducible Research: Trying to Bridge the Gap



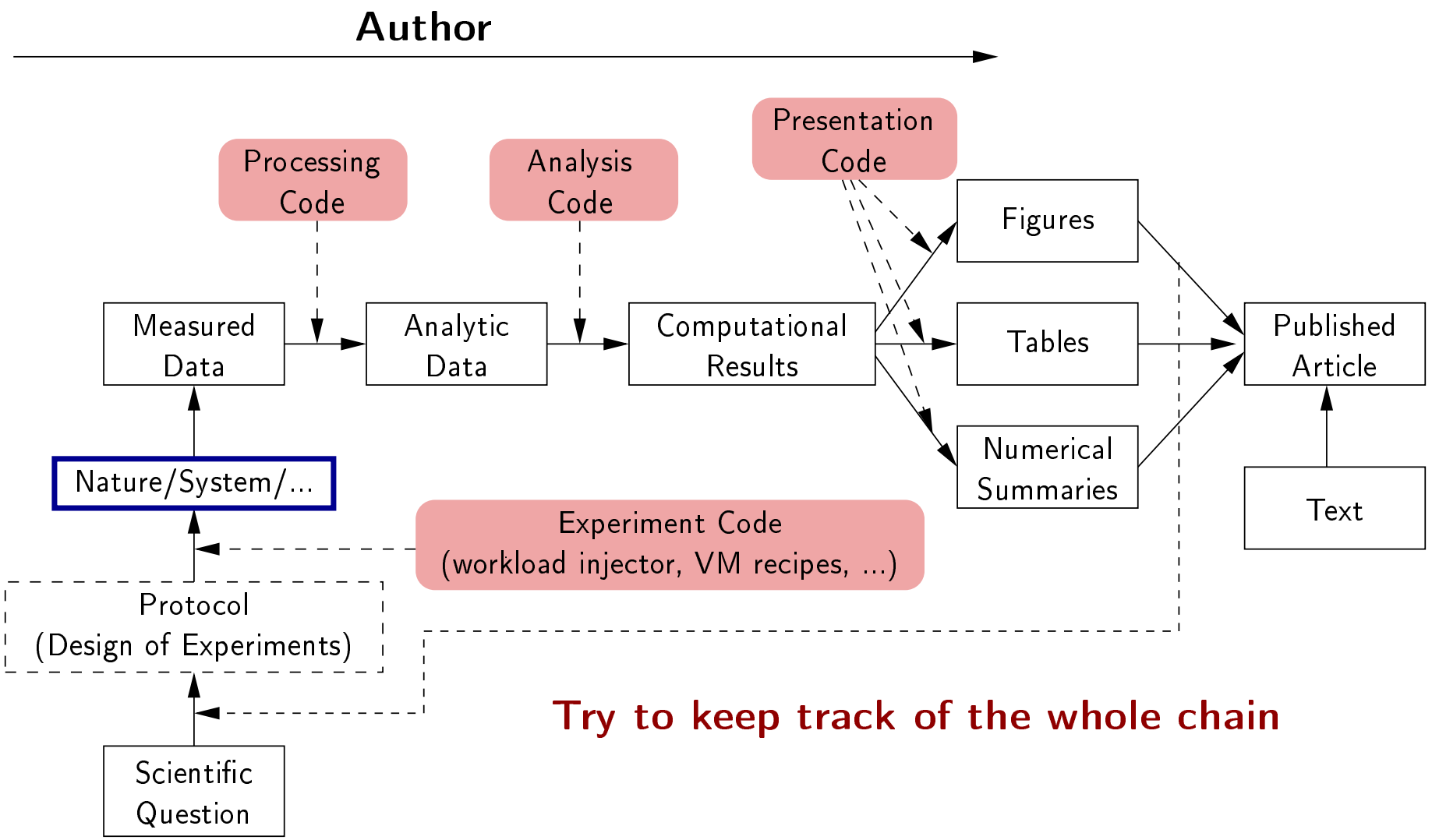
Inspired by Roger D. Peng's lecture on reproducible research, May 2014

Reproducible Research: Trying to Bridge the Gap



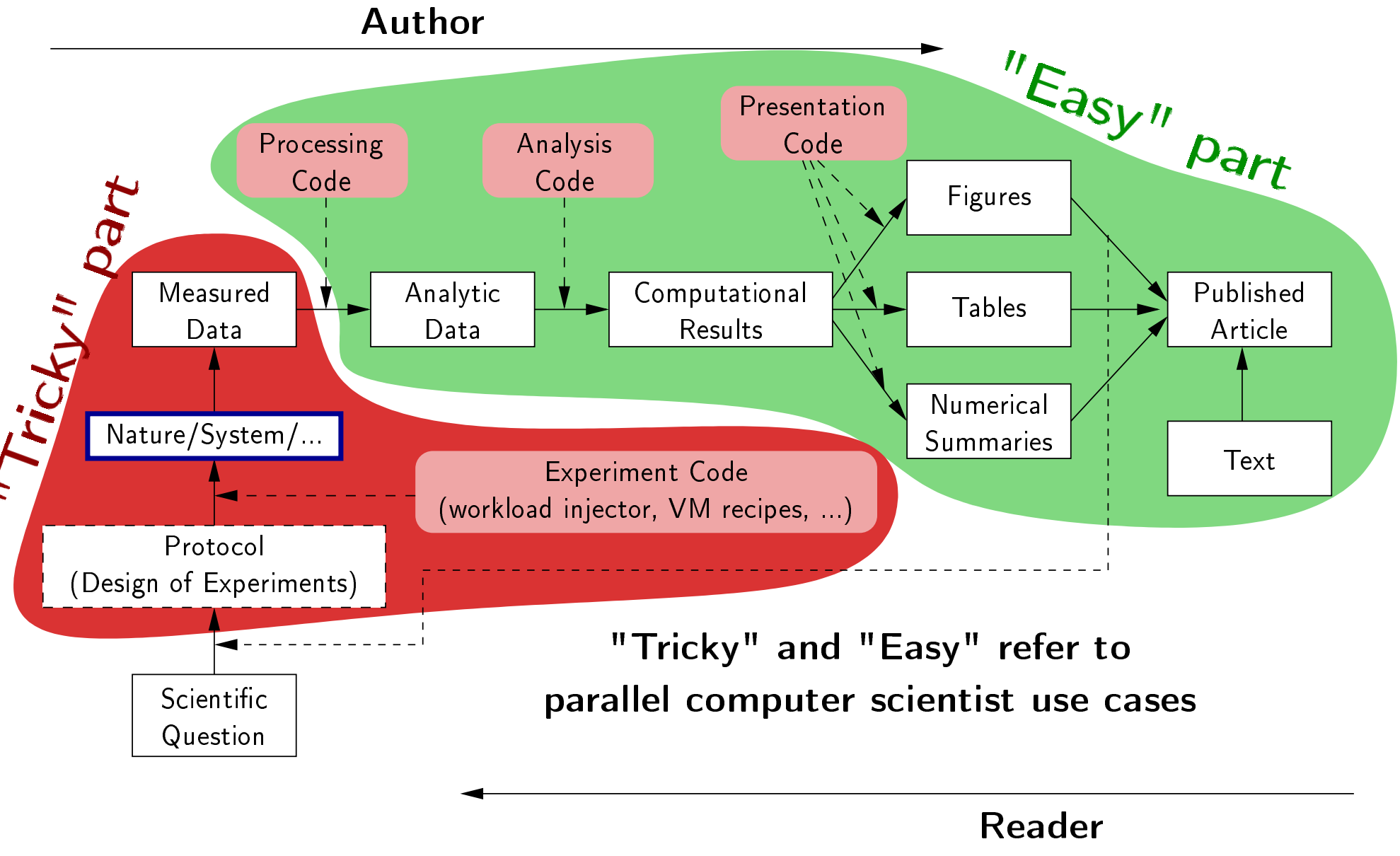
Inspired by Roger D. Peng's lecture on reproducible research, May 2014

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

Our Approach: An Infrastructure to Support Provenance-Rich Papers [Koop et al., ICCS 2011]

- Tools for *authors* to create reproducible papers
  - Specifications that encode the computational processes
  - Package the results
  - Link from publications

Support different approaches

- Tools for testers to repeat and validate results
  - Explore different parameters, data sets, algorithms
- Interfaces for searching, comparing and analyzing experiments and results
  - Can we discover better approaches to a given problem?
  - Or discover relationships among workflows and the problems?
  - How to describe experiments?

Courtesy of Juliana Freire (AMP Workshop on Reproducible research)

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

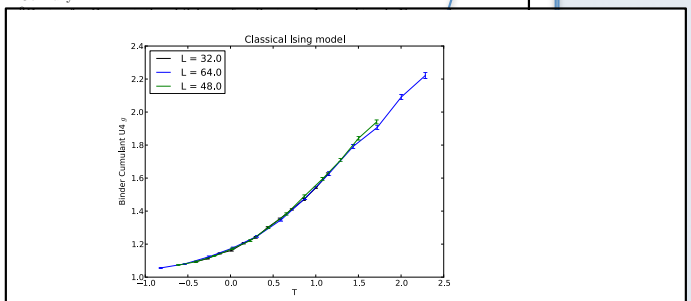
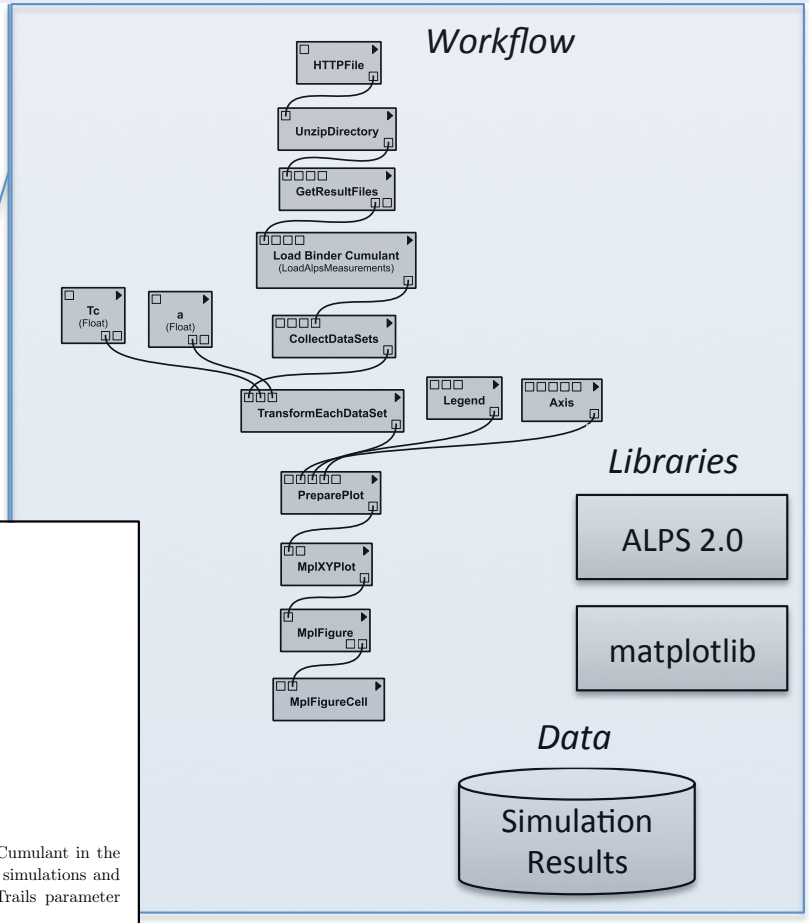


Figure 3. In this example we show a data collapse of the Binder Cumulant in the classical Ising model. The data has been produced by remotely run simulations and the critical exponent has been obtained with the help of the VisTrails parameter exploration functionality.



[Bauer et al., JSTAT 2011]

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}
```

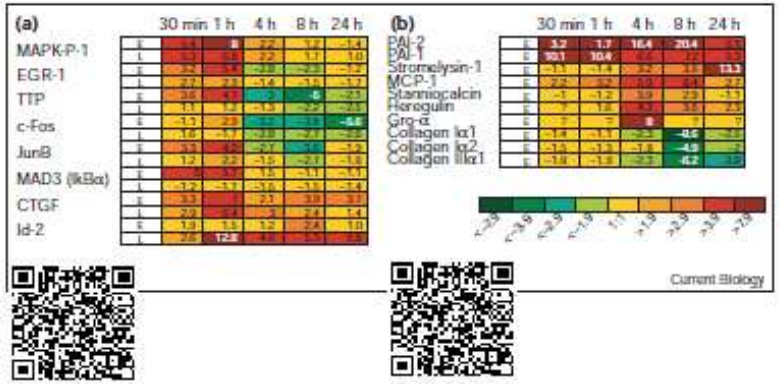
Permanently bind printed graphics to underlying result content

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

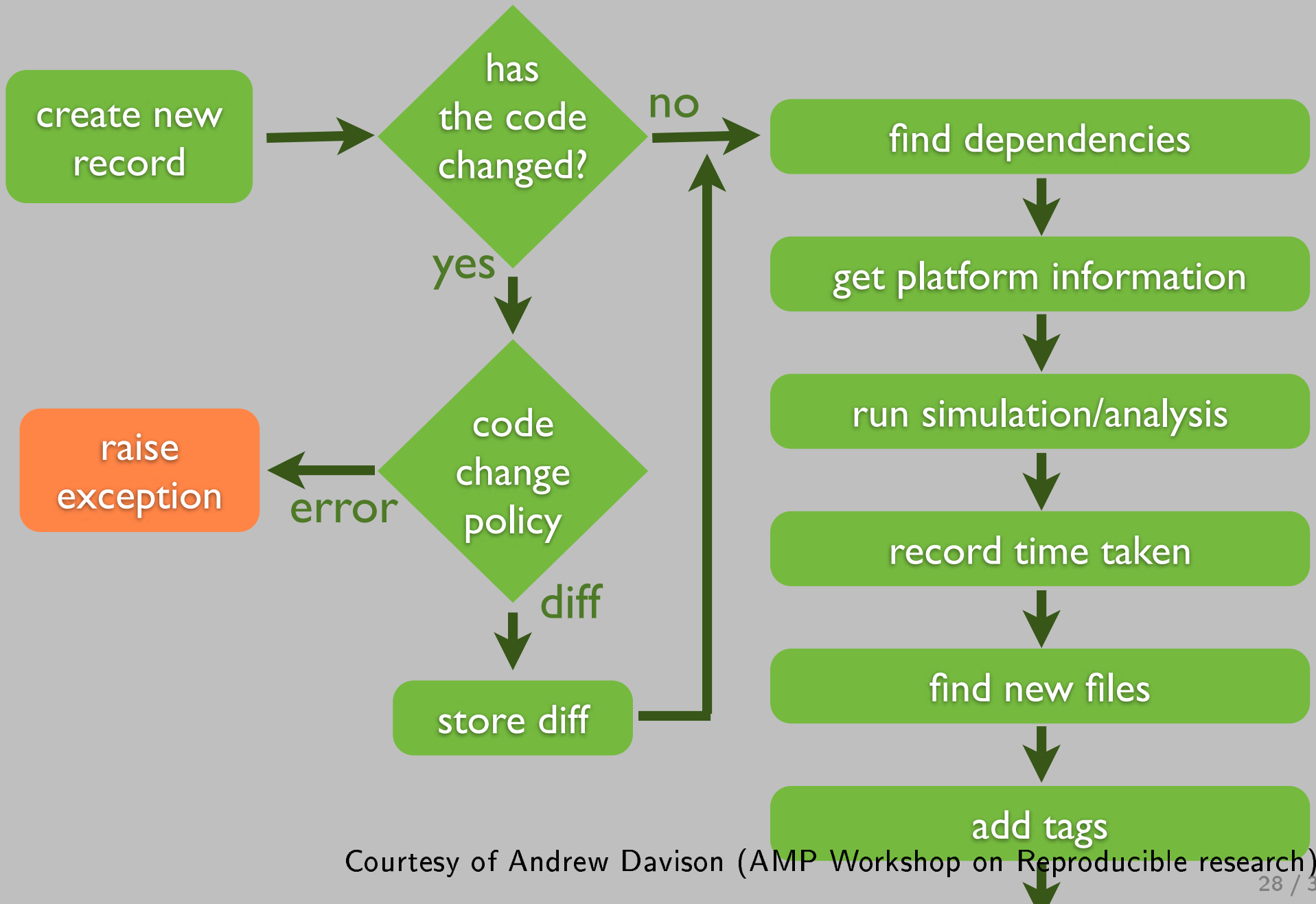
states overlap substantially with those in telomere-induced senescence (W.F., D.N.S., R. Allsopp, S. Lowe, and G. Ferbeyre, unpublished observations) and thus are likely to use many of the same activation processes.

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

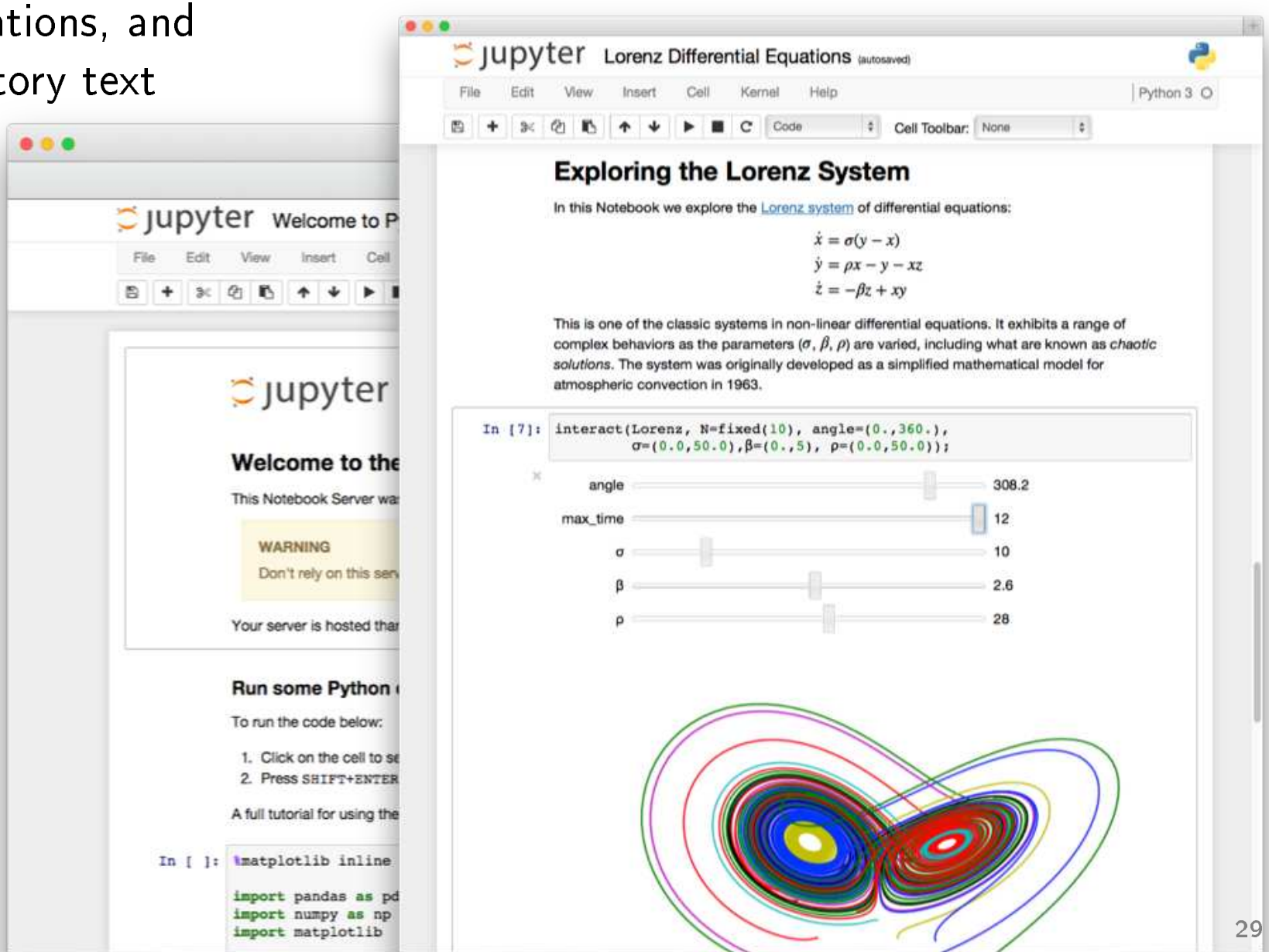
TestProject: List of records

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

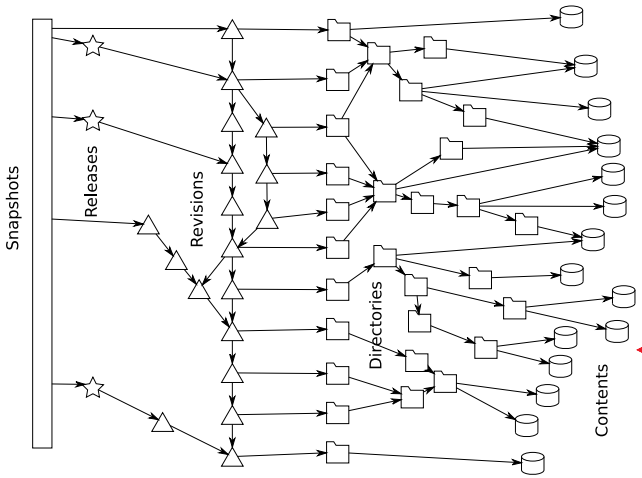
Courtesy of Victoria Stodden (UC Davis, Feb 13, 2014)

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, ...

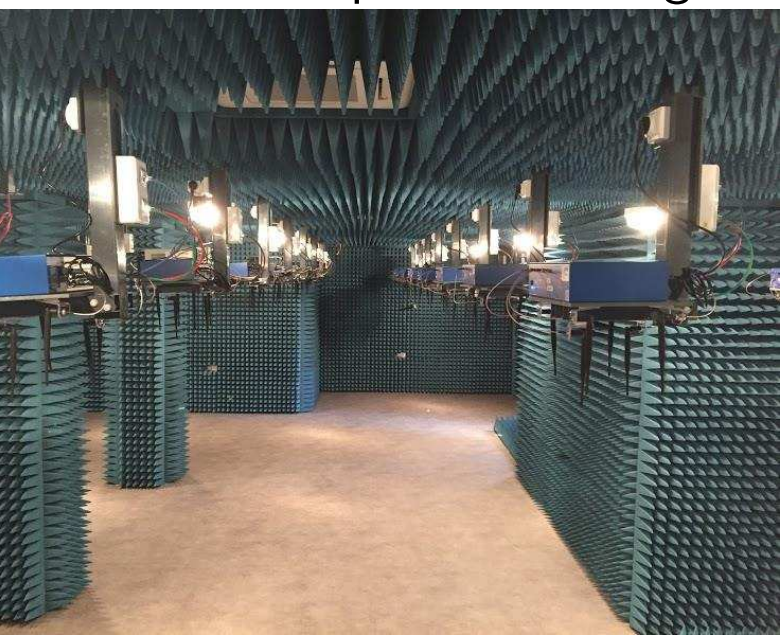


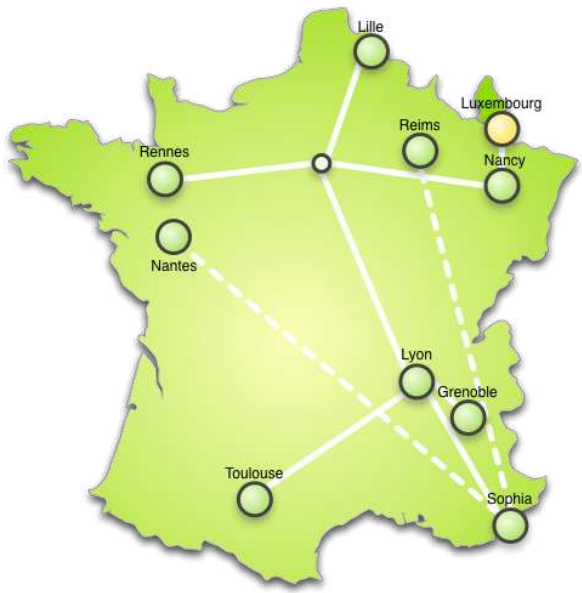
Software Heritage

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

- 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

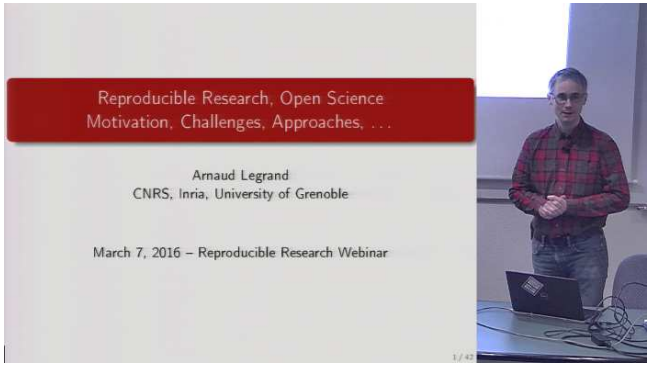
It's up to us. **We** should **care** and **take the lead**

- Slight cultural changes in our relation to publication and daily practice
- Higher confidence in our work $\rightsquigarrow$ definite competitive advantage
- Our research will become sound, deeper, auditable, more visible, reusable, ...

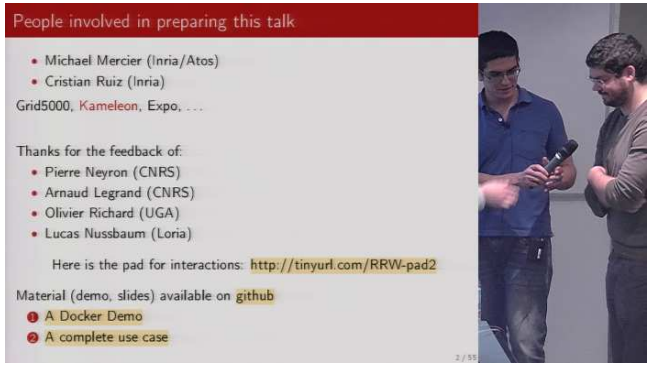
Webinars: Learning by Doing

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



Litterate programming



Controling your environment

Numerical reproducibility

Logging and backing up

Next webinars: in two weeks on Artifact Evaluation!

https://github.com/alegrand/RR_webinars