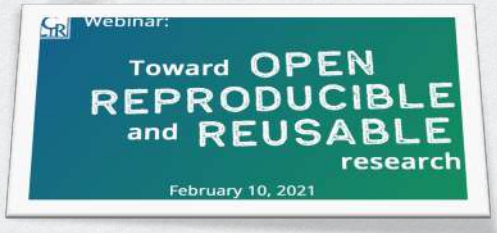
Computational Reproducibility vs Transparency: *Is it FAIR enough?*

Bertram Ludäscher
Director, Center for Informatics Research in Science & Scholarship (**CIRSS**)
School of Information Sciences (**iSchool@Illinois**)
& National Center for Supercomputing Applications (**NCSA**)
& Department of Computer Science (**CS@Illinois**)

with special thanks (and apologies) to
Timothy McPhillips
& ***Workshop on Research Objects (RO),*** *eScience, San Diego, 2019*
& ***Reproducibility of Data-Oriented Experiments in e-Science****, Dagstuhl Seminar, 2016*



ILLINOIS
School of Information Sciences

Whole Tale

NCSA

Overview

- **FAIR** data, code, and **reproducibility**
- The **Reproducibility Crisis** ...
- ... and an **R-Words** (terminology) crisis?
- Reproducibility and **Information Gain** (PRIMAD)
- => shift from R-Words to **T-Words**: **Transparency** ...
- Capturing and querying **Provenance**
- Reproducibility & Transparency in **Whole Tale**

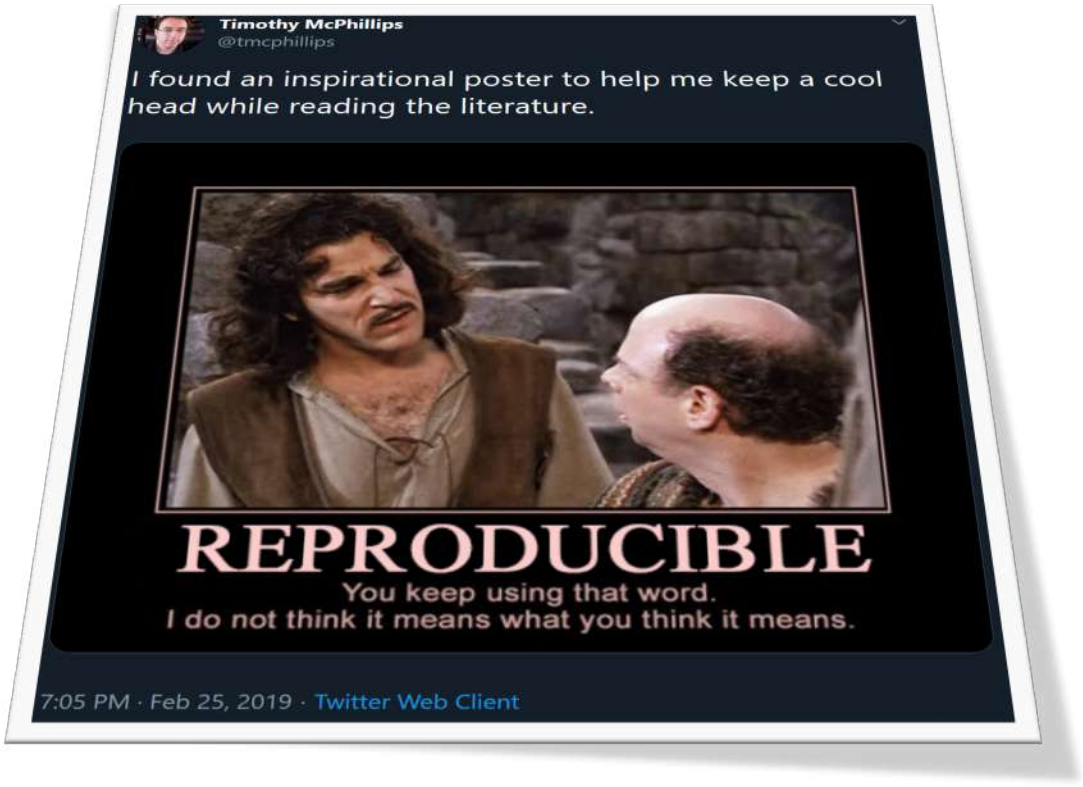
FAIR data, code, … Reproducibility

- FAIR data principles: data should be *findable*, *accessible*, *interoperable*, *reusable*
- Metadata (duh!) is key!
- .. the principles are now being adapted (*mutatis mutandis*) for ***code***, ***scientific workflows***, …

- Can we do something about "**the reproducibility crisis**"?
  - e.g. by focusing on ***computational reproducibility*** *…!?*

Is Reproducibility really so complicated?

- ***Reproducibility*** crisis?
- ***Terminology*** crisis?
- Or *gullibility* crisis?



- What **is reproducibility** anyway?
- And who is **responsible** for it?

Pop QUIZ: What is the **single most effective** way to make **your research** more **reproducible**?

a) Employ the *interoperability standards* for scientific data, metadata, software, and *Research Objects*

b) Carefully *record and report* your work

c) Use *open source software* and make any new or modified code freely available.

d) Apply *FAIR (findable, accessible, interoperable, reusable)* principles

e) Do all of your work in *software containers*

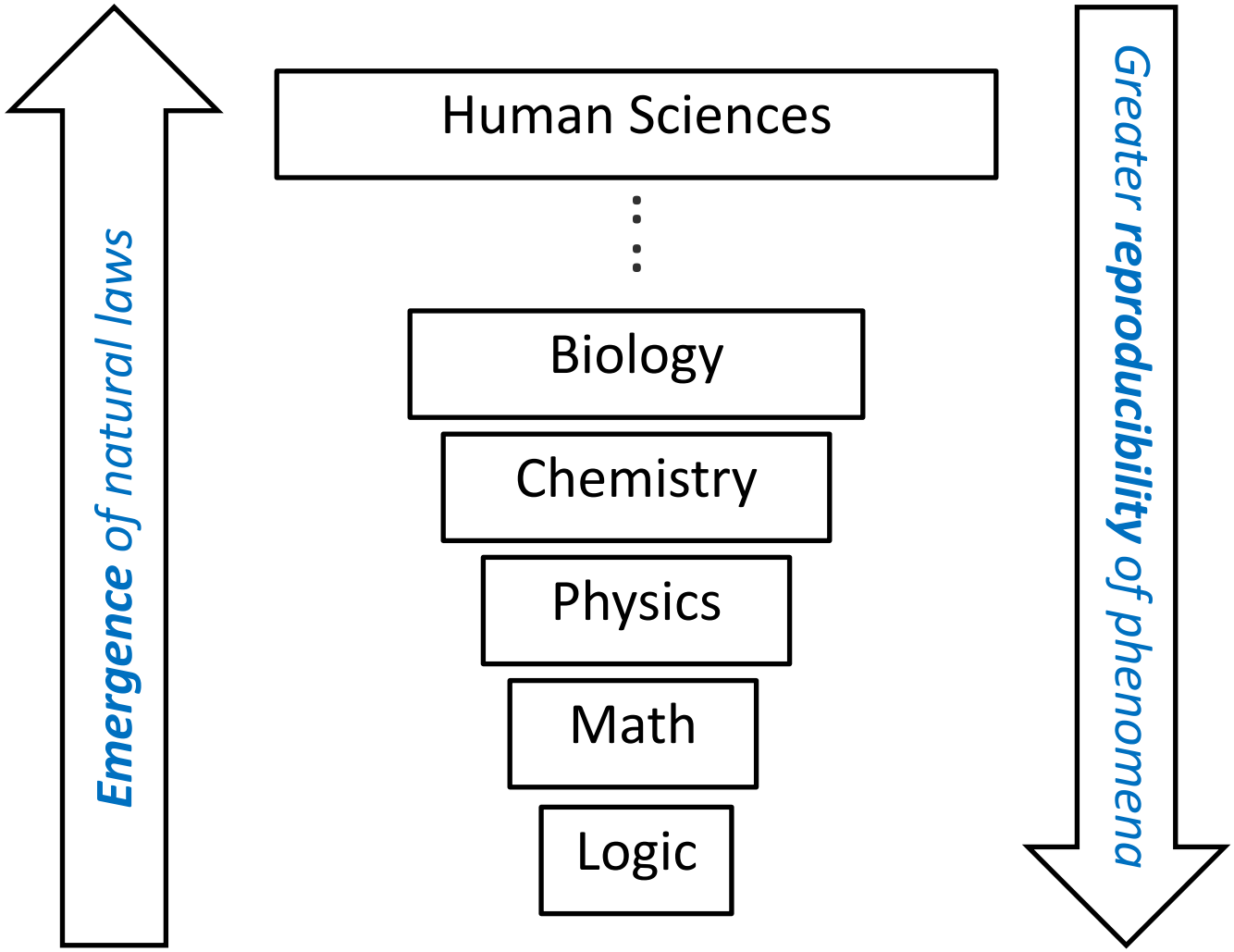
f) Focus your research on *intrinsically reproducible* phenomena

Basic Assumptions made by researchers in the Natural Sciences …

- We are discovering things that are **the way they are** whether we go and look for them or not.
- We are discovering things that **conceivably could be different** than they happen to be. To find out how things actually are **we must go look**.
- It **does not matter who does the looking**. Everyone with the same opportunity to look will find the same things to be true.

… nature as the ultimate reproducibility arbiter …

Is there a hierarchy of intrinsic reproducibility?



It's not so simple…

- *… but things tend to get "messier" further up …*

Limits on reproducibility in the natural sciences

- Nature is **not** a digital computer. It's more of an **entropy generator** built on **chaos** and (true) **randomness** with natural laws, math, and logic serving as **constraints**.
- Good experiments are hard to design and to perform **even once**.
- Instruments can be **costly** and limited in supply.
- Many phenomena **cannot be studied via experiment** at all.
- **Past events** are crucial to many theories.
- Some things happen only once.
- *… so let's hold back the horses (for now) on extensive and expensive computational reproducibility studies?? …*

But what is always possible? Transparency!

FASEB* definition of transparency

Transparency: *The reporting of experimental materials and methods in a manner that provides enough information for others to* ***independently assess and/or reproduce*** *experimental findings.*

- Transparency is what allows an experiment to be **reviewed** and **assessed *independently*** by others.
- Transparency ***facilitates reproduction*** of results but ***does not require reproduction*** to support review and assessment.
- ***It is considered a problem*** if **exact repetition** of the steps in reported research is **required** either to evaluate the work or to reproduce results.

** The **Federation of American Societies for Experimental Biology** comprises 30 scientific societies and over 130,000 researchers.*

Quantifying **Repeatability**

- Experiments on natural phenomena generally are **not** exactly repeatable.
- **Materials**, **conditions**, **equipment**, and instruments all **vary**.
- **Uncertainty** is **intrinsic** to most measurements.
- Experimental biologists perform ***replicate* experiments** to assess end-to-end **repeatability**.

  Technical replicates: Measurements and data analyses performed on the **same sample** using the **same equipment** multiple times.

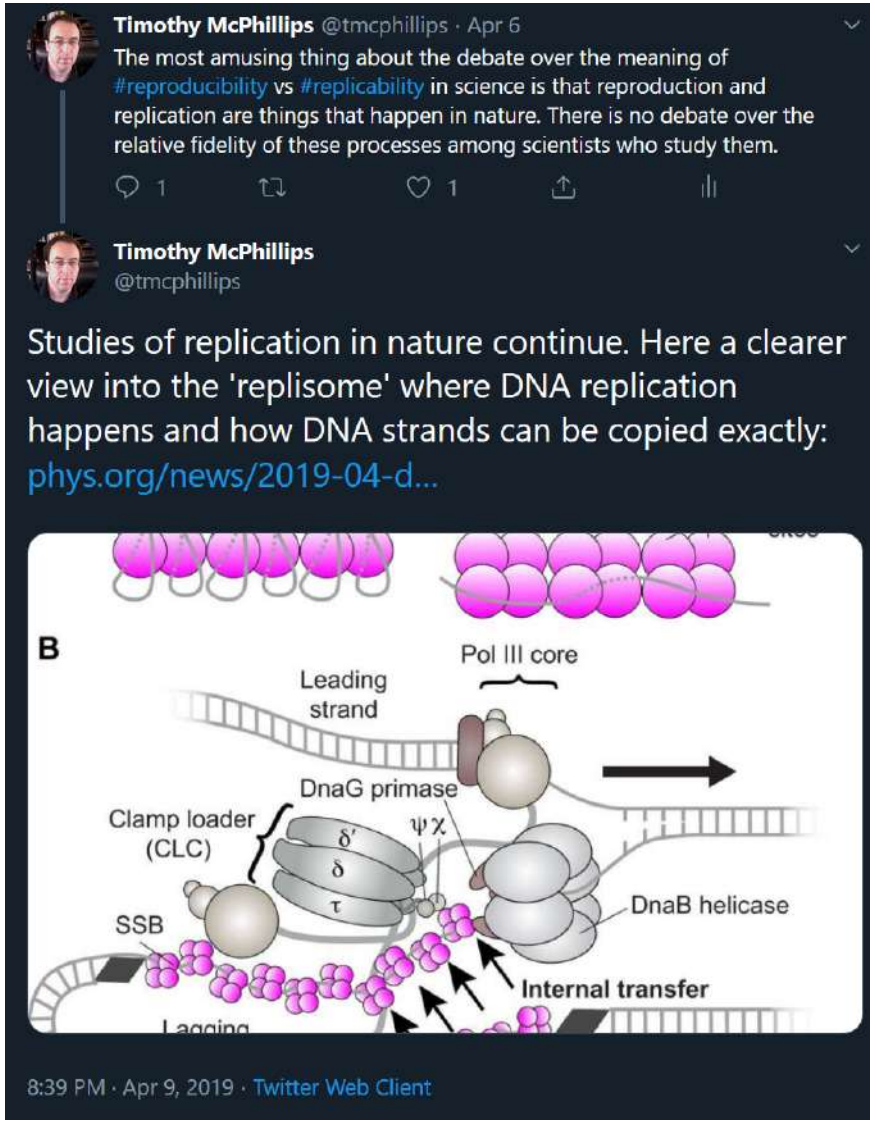
  Biological replicates: Measurements and data analyses performed on different but **biologically equivalent samples** on the **same equipment**.

A mystery?? Why are these "replicates", not "reproductions"?

Replication and **Reproduction** are natural processes that biologists study *(.. a lot!)*

- Amazing aspect of life is the incredible **fidelity** with which genetic material—DNA—is ***replicated*** within cells.
- DNA **replication** is carried out by the ***replisome***—which even detects and corrects errors on the fly!
- Organisms ***reproduce*** and have ***reproductive*** systems.
- Biological **reproduction** is ***much lower fidelity*** than DNA **replication**. In fact, the process of reproduction often ***encourages variation*** in the children.

Experimental replicates assess the **highest possible fidelity** at which an experiment can be **repeated**—by the **same** researcher, using the **same** equipment, on the **same** or equivalent samples, **immediately one after the other** in time.



Theorists talk about **replication**

- Dawkins' ***selfish genes*** are **replicators**.
- Debate in **origins of life** research:

  *Did **replication** or **metabolism** come first?*
- Could life have started ***before*** high-fidelity replication of genetic material was achieved?
- For these theorists and philosophers **high-fidelity** is the **defining** characteristic of **replication**.

Stanford Encyclopedia of Philosophy

Replication and Reproduction

First published Wed Dec 5, 2001; substantive revision Tue Sep 25, 2018

The problem of replication and reproduction arises out of the history of genetics [see the entry gene for a historical review]. It is tied to the concept of the gene and its generalization in an evolutionary context [see the entry evolution]. Richard Dawkins introduced the notion of replicators—things that self-replicate—as a universalization of evolutionary understandings of genes. Dawkins argued that replicators are the *sine qua non* of evolution by natural selection [see the entry natural selection], while other accounts only require *reproduction* as one of its defining features. What exactly is a replicator? How are replicators different from genes? Can evolution by natural selection occur without the existence of replicators? Besides the biological domain, are there any other domains in which replicators have been postulated? To answer these questions, we will first provide some background for Dawkins' notion of replicator and its ties with the concepts of the gene and information. We will then introduce the distinction between *Replicators* and *Vehicles* in the context of biological evolution and followed by the extension of this to other domains. Finally, we will discuss some of the challenges to the idea that replicators are necessary conditions for evolution by natural selection.

FASEB* definitions of *reproducibility* and *replicability*

Maximal fidelity to original **experiment**, greater fidelity to original result.

Replicability: *The ability to* ***duplicate*** *(i.e., repeat) a* ***prior result*** *using the* ***same source materials*** *and* ***methodologies****. This term should only be used when referring to repeating the results of a* ***specific experiment*** *rather than an entire study.*

Reproducibility: *The ability to achieve* ***similar*** *or nearly identical* ***results*** *using* ***comparable materials*** *and* ***methodologies****. This term may be used when specific* ***findings from a study*** *are obtained by an independent group of researchers.*

Less fidelity to original **study**, lower fidelity result expected.

** The Federation of American Societies for Experimental Biology comprises 30 scientific societies and over 130,000 researchers.*

Beyond Reproduction and Replication: Exact Repeatability

- Digital computers use logic gates to achieve replication of information at such a low error rate we can call it ***exact***.
- Computers pull the exactness of logic and discrete mathematics up to the level of macroscale phenomena– quite a feat.
- Exactness is (effectively) achievable for computer hardware, compiled software, program executions, and computing environments.
- Researchers employing digital computers have access to **a new kind of reproducibility** never before seen in science: ***exact repeatability***.

ACM Initiative …

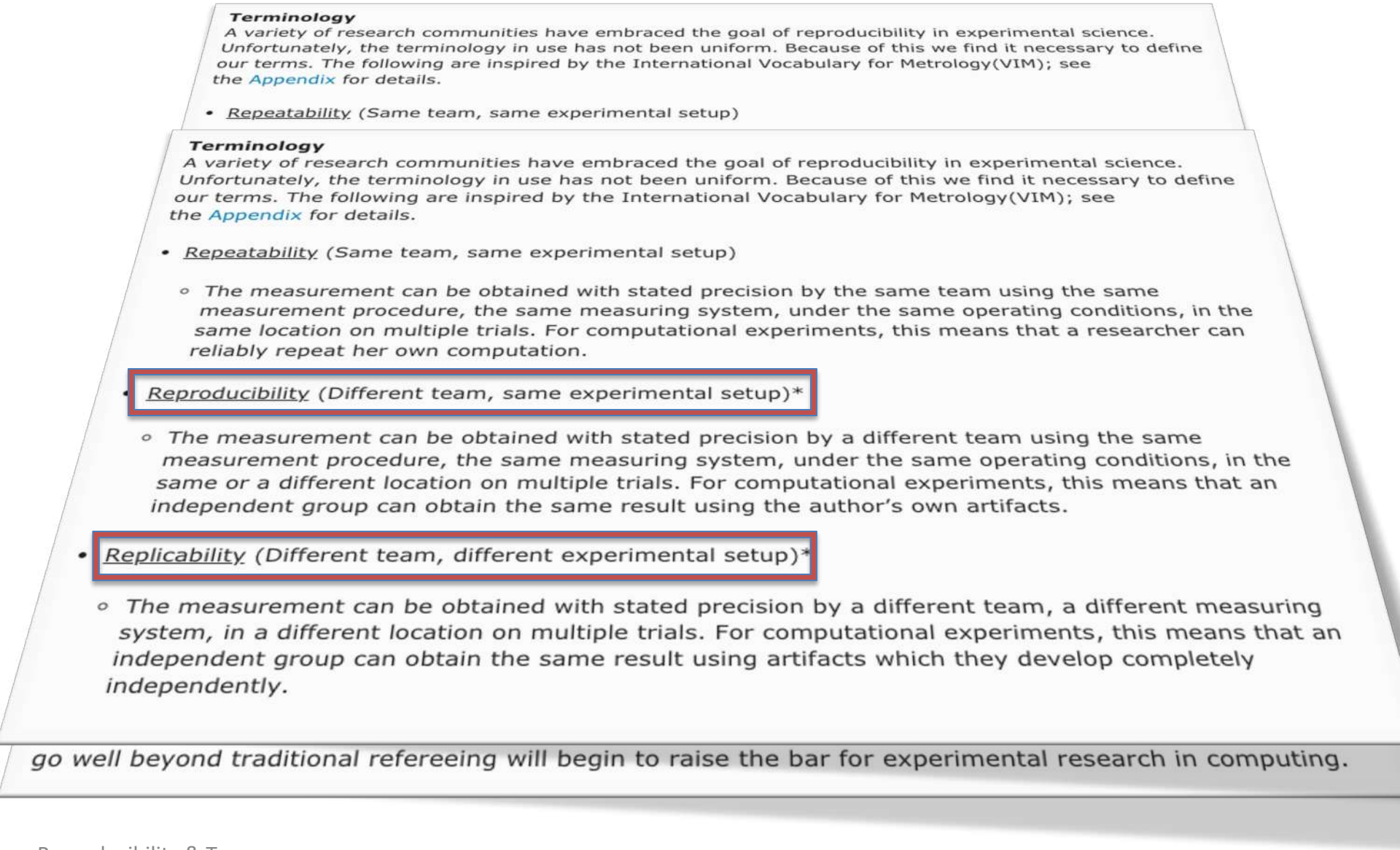
Terminology

A variety of research communities have embraced the goal of reproducibility in experimental science. Unfortunately, the terminology in use has not been uniform. Because of this we find it necessary to define our terms. The following are inspired by the International Vocabulary for Metrology(VIM); see the Appendix for details.

- *Repeatability* (Same team, same experimental setup)
  - *The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation.*
- *Reproducibility* (Different team, different experimental setup)*
  - *The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts.*
- *Replicability* (Different team, same experimental setup)*
  - *The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently.*

The concepts of repeatability and reproducibility are taken directly from the VIM. Repeatability is something we expect of any well-controlled experiment. Results that are not repeatable are rarely suitable for publication. The proposed intermediate concept of replicability stems from the unique properties of computational experiments, i.e., that the measurement procedure/system, being virtual, is more easily portable, enabling inspection and exercise by others. While reproducibility is the ultimate goal, this initiative seeks to take an intermediate step, that is, to promote practices that lead to better replicability. We fully acknowledge that simple replication of results using author-supplied artifacts is a weak form of reproducibility. Nevertheless, it is an important first step, and the auditing processes that go well beyond traditional refereeing will begin to raise the bar for experimental research in computing.

ACM Initiative … *reloaded?*



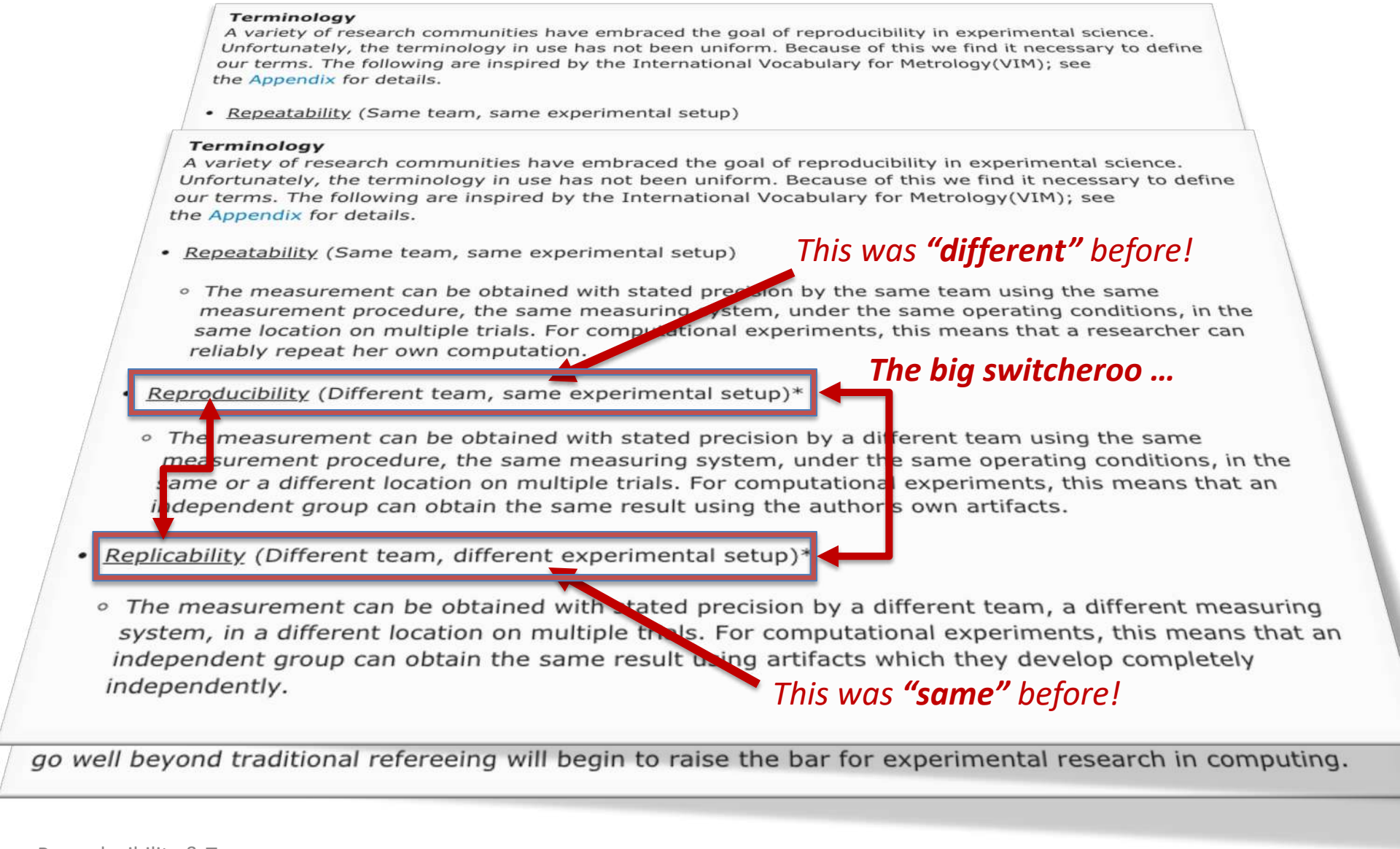
Terminology

A variety of research communities have embraced the goal of reproducibility in experimental science. Unfortunately, the terminology in use has not been uniform. Because of this we find it necessary to define our terms. The following are inspired by the International Vocabulary for Metrology(VIM); see the Appendix for details.

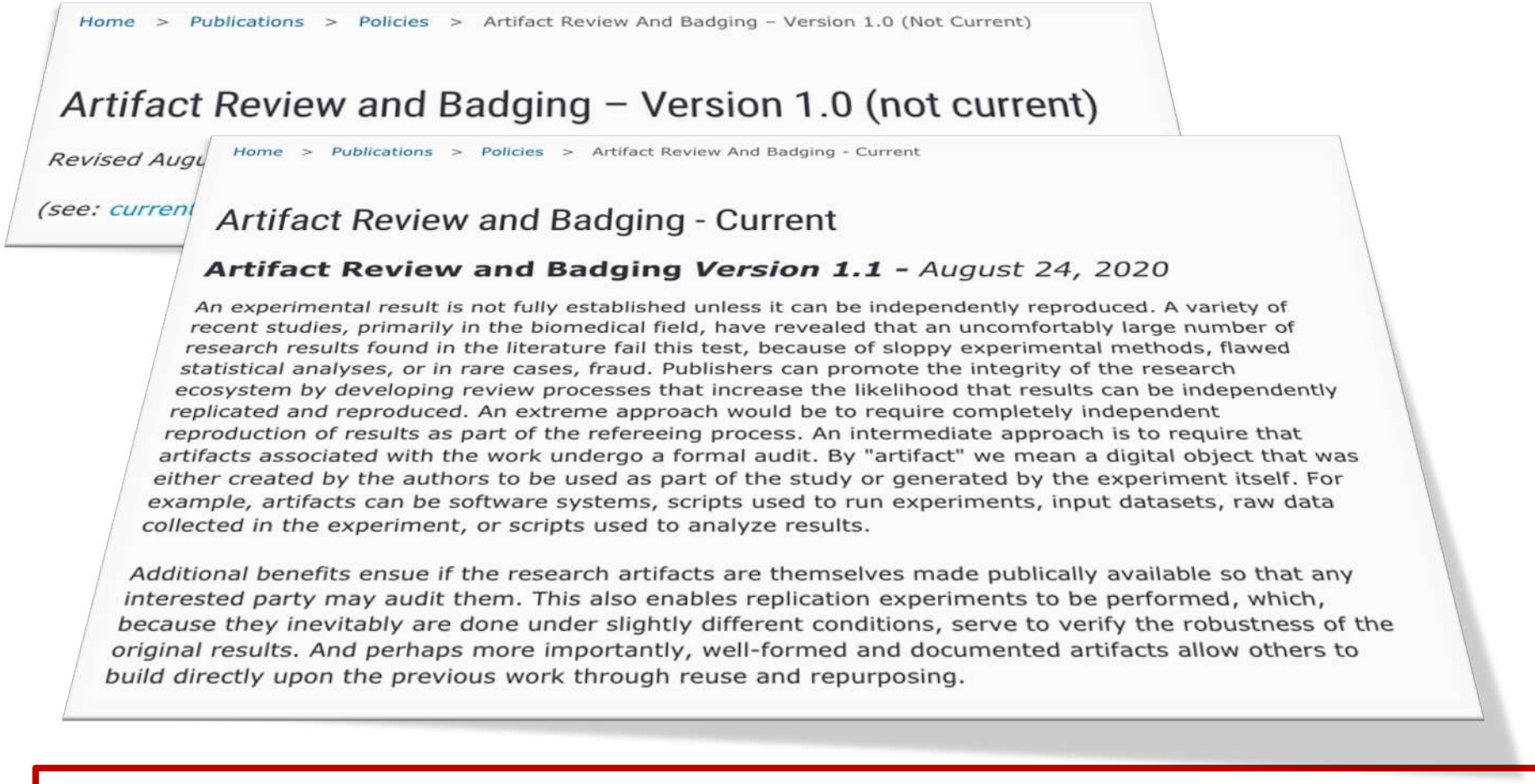
- *Repeatability (Same team, same experimental setup)*
  - *The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation.*
- *Reproducibility (Different team, same experimental setup)**
  - *The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts.*
- *Replicability (Different team, different experimental setup)**
  - *The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently.*

go well beyond traditional refereeing will begin to raise the bar for experimental research in computing.

ACM Initiative … *reloaded?*



ACM caves to new terminology policy?



*As a result of discussions with the National Information Standards Organization (NISO), it was recommended that ACM harmonize its terminology and definitions with those used in the broader scientific research community, and ACM agreed with NISO's recommendation to swap the terms "reproducibility" and "replication" with the existing definitions used by ACM as part of its artifact review and badging initiative. ACM took action to update all prior badging to ensure consistency.

Reproducibility badges and verification workflows

… choices & options galore …

- ACM SIGMOD defines a defines a procedure for **assessing database research reproducibility**.
- ACM awards (currently) **four different reproducibility badges** distinct from the SIGMOD reproducibility assessment.
- ACM has defined **eight versions** of the guidelines for awarding its badges **since 2015**.
- The workflow used by the American Journal of Political Science (AJPS) to verify computational artifacts also is versioned.
- Does the meaning of reproducibility badges may change from year to year even within a single organization? Is there light at the end of the terminology tunnel?

If we want these badges to have any meaning at all they should be mapped to something that isn't constantly changing.

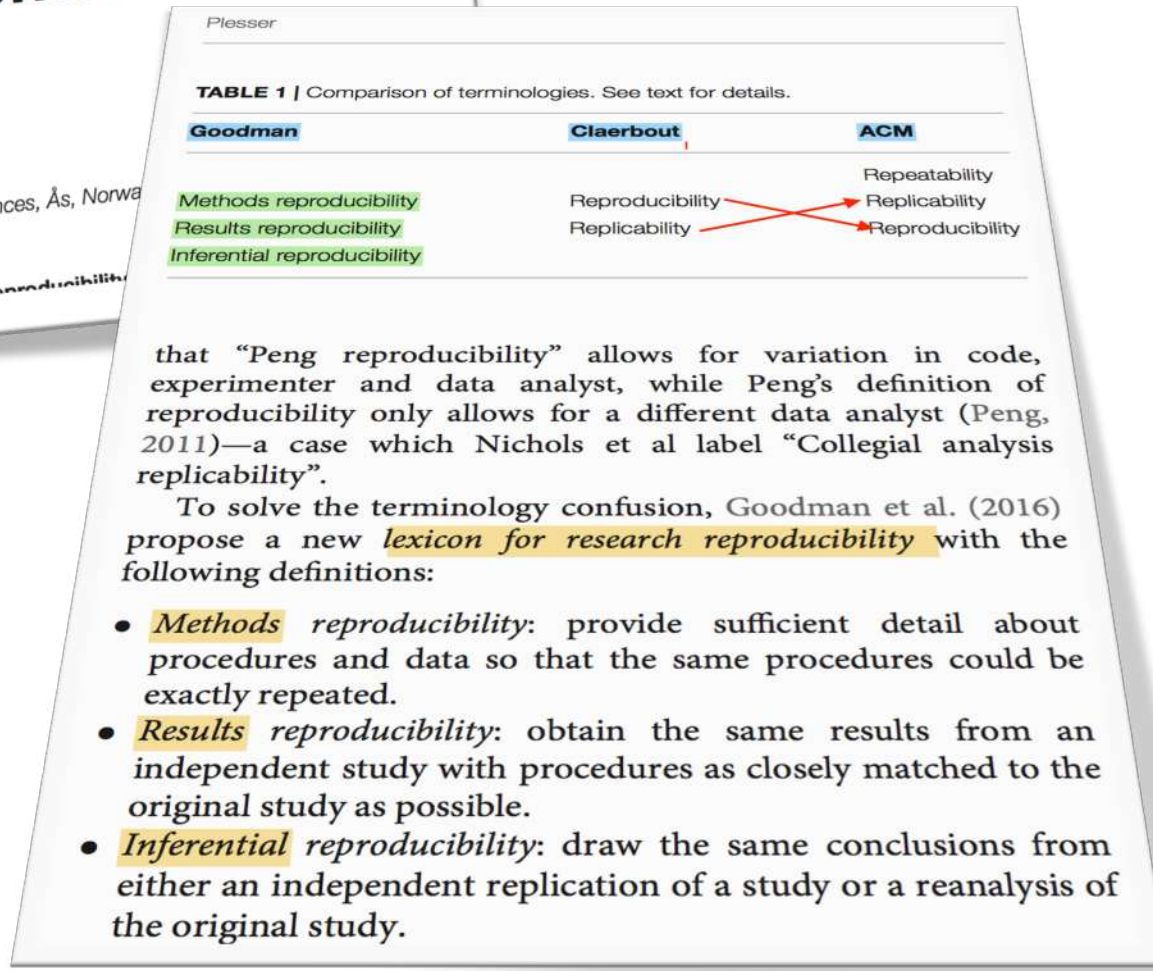
db-reproducibility.seas.harvard.edu, www.acm.org/publications/policies/artifact-review-badging , ajps.org/wp-content/uploads/2019/01/ajps-quant-data-checklist-ver-1-2.pdf

Yes, we need to **Mind our Vocabulary!**

Reproducibility vs. Replicability: A Brief History of a Confused Terminology

Hans E. Plesser[1,2*]

[1] Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norwa
Medicine (INM-6), Jülich Research Centre, Jülich, Germany



Plesser

TABLE 1 | Comparison of terminologies. See text for details.

Goodman	Claerbout	ACM
		Repeatability
Methods reproducibility	Reproducibility	Replicability
Results reproducibility	Replicability	Reproducibility
Inferential reproducibility		

that "Peng reproducibility" allows for variation in code, experimenter and data analyst, while Peng's definition of reproducibility only allows for a different data analyst (Peng, 2011)—a case which Nichols et al label "Collegial analysis replicability".

To solve the terminology confusion, Goodman et al. (2016) propose a new *lexicon for research reproducibility* with the following definitions:

- *Methods reproducibility*: provide sufficient detail about procedures and data so that the same procedures could be exactly repeated.
- *Results reproducibility*: obtain the same results from an independent study with procedures as closely matched to the original study as possible.
- *Inferential reproducibility*: draw the same conclusions from either an independent replication of a study or a reanalysis of the original study.

*ACM **was** aligned - just not "in harmony" with NAS committee … Now it's a more aligned with NAS, but no longer with FASEB, …*

(some crossed wires are now aligned; some previously aligned wires are now crossed …)

***with namespaces**: NAS:reproducibility ~ FASB:replicability*
NAS:replicability ~ FASB:reproducibility

Chaos is a ladder.

Is reproducibility a staircase?

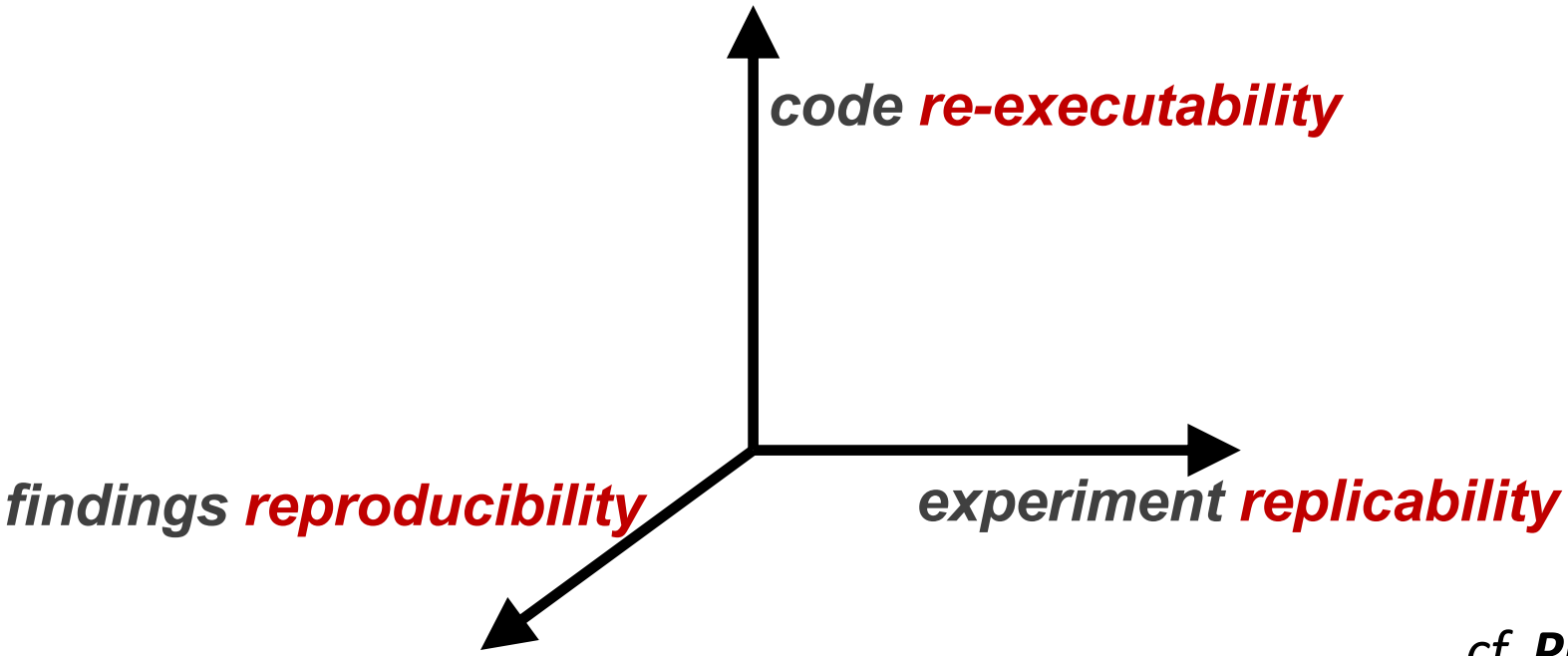
Greater reproducibility?

- Study **fully reproducible** !
- Code **reusable** !
- Computed artifacts **support** paper
- Code produces **expected artifacts**
- **Computing environment** repeatable
- **Code** shared and freely licensed
- **Data** published and accessible to all

*It is tempting to think about reproducibility **one-dimensionally** …*

But isn't scientific reproducibility multidimensional?

- Do the ***R-words*** have an obvious **order**, where **achieving one must precede achieving the next**??
- Or might they represent base vectors of a multidimensional space?



cf. ***PRIMAD***

Modeling reproducibility as multidimensional may offer way out of the terminology quagmire

- Recognize that **different terminologies refer to different sets of dimensions**; communities focus on different subspaces, or different choices of basis vectors.
- **Map conflicting definitions onto shared dimensions**; use mappings to convert claims made using one terminology to claims using a different terminology.
- Allow each community to focus on dimensions of interest to them using the most intuitive terminology; **use namespaces to eliminate ambiguity**.
- Use ***Research Objects*** to **attach claims about reproducibility to research artifacts**, to **disambiguate** these **claims**, and to **support queries** using terminology of the user's choosing.

Transparent Research Objects

- Transparency in the natural sciences **enables research to be evaluated**—and reported results used with confidence—**without actually repeating others' work**.
- How can **ROs extend the advantages of transparency to computational research** and the computational components of experimental studies?
- Researchers need to be able to **query** the **reproducibility characteristics** of artifacts in ROs.
- These queries need to be poseable **using terminology familiar to the researcher**—terminology likely **different** from that used by the author of the RO (minimizing headaches no matter which terminology you grew up with..)
- Queries about computational reproducibility need to take the **longevity of technological approaches to reproducibility** into account.

Food for Thought: Research Objects & Information Gain

- An *object of research* is the primary target of scholarly investigation.

In contrast, we may think of a **research object** as an artifact that

(a) performs a specific function,

(b) is guided by and underlying theory

(c) whose objective might be to **allow information gains towards falsifying a particular hypothesis,** and

(d) Which admits representation through a metalanguage that captures its role in a science-driven discourse.

PRIMAD *(what have you "primed"?)*

6.1.2 The PRIMAD Model

As a starting point, we defined a preliminary list of "variables" that could potentially be changed:

- (R) or (O) Research Objectives / Goals
- (M) Methods / Algorithms
- (I) Implementation / Code / Source-Code
- (P) Platform / Execution Environment / Context
- (A) Actors / Persons
- (D) Data (input data and parameter values)

This spells: OMIPAD. Rearranging the letters that we use to represent the several aspects that can be changed, it can be remembered as PRIMAD: (P)latform, (R)esearch Goal, (I)mplementation, (M)ethod, (A)ctor, (D)ata (both input and parameter data), which allows us to ask: What variables have you "primed" in your reproducibility study?

Dagstuhl Seminar #16041 Report

Outputs = Exec(M,I,P,D) | RO, A

- M = parsimony/bootstrap/..
- I = package XYZ
- P = MacOS ..
- D = (Params, Files)

PRIMAD & Information Gain

- ***Original*** *study:* $Y = F_P(X)$ ***Reproduction****:* $Y' = F'_{P'}(X')$
  - ***Y' ≈ Y*** *=>* ***Reproduction Success*** *else* ***Reproduction Failure***

NOTE:
This does ***NOT*** *mean that a small delta in a parameter results couldn't have a large change in the output …*

no wiggle ***biggest wiggle***

no wiggle *biggest wiggle*

PRIMAD *(what have you "primed"?)*

Label	Data		Platform / Stack	Implementation	Method	Research Objective	Actor	Gain
	Parameters	Raw Data						
Repeat	-	-	-	-	-	-		Determinism
Param. Sweep	x	-	-	-	-	-		Robustness / Sensitivity
Generalize	(x)	x	-	-	-	-		Applicability across different settings
Port	-	-	x	-	-	-		Portability across platforms, flexibility
Re-code	-	-	(x)	x	-	-		Correctness of implementation, flexibility, adoption, efficiency
Validate	(x)	(x)	(x)	(x)	x	-		Correctness of hypothesis, validation via different approach
Re-use	-	-	-	-	-	x		Apply code in different settings, Re-purpose
Independent *x* *(orthogonal)*							x	Sufficiency of information, independent verification

Figure 1 PRIMAD Model: Categorizing the various types of reproducibility by varying the (P)latform, (R)esearch Objective, (I)mplementation, (M)ethod, (A)ctor and (D)ata, analyzing the gain they bring to computational experiments. x denotes the variable primed i.e. changed, (x) a variable that may need to be changed as a consequence, whereas – denotes no change.

Dagstuhl Seminar #16041 Report

Back to computational reproducibility:
Journal verification workflows in Whole Tale

- Important **new** use case for Whole Tale
- Study of **journal reproducibility initiatives** *(Willis, 2020a)* -- **FINDINGS**:
  - Initiatives have **common, basic requirements** for **transparency** and **computational reproducibility**
  - Initiatives rely on **established research repositories** for artifact **preservation** and **long-term access** (so does WT)
  - Editorial **infrastructure is lacking** (tools to support packaging, access to computational infrastructure) -- WT provides this, but they need more
  - Need for **standards** for the *description* and *packaging* of reproducible and transparent computational Research Objects (our ***Tale format***)

Willis, C. *(2020a).* ***Trust, but verify: An investigation of methods of verification and dissemination of computational research artifacts for transparency and reproducibility*** *(Ph.D. thesis). University of Illinois at Urbana-Champaign*

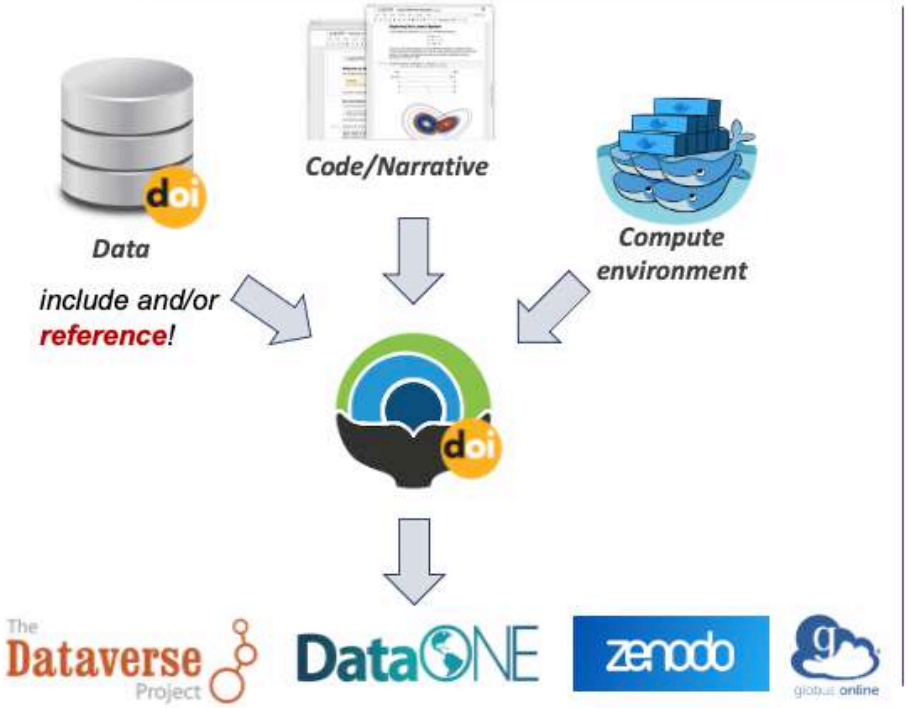
Whole Tale & the Elements of a …

Reproducible Computational Research Platform

Develop	Analyze	Share	Package	Reproduce
Easy-to-access **cloud-based** computational **environments**	**Transparent** access to research **data**	**Collaborate** and **share** with others	**Export** or **publish executable research objects**	**Re-execute** **Review** **Verify** **Re-use**

Support users (researchers, scientists) & the tools they already use!

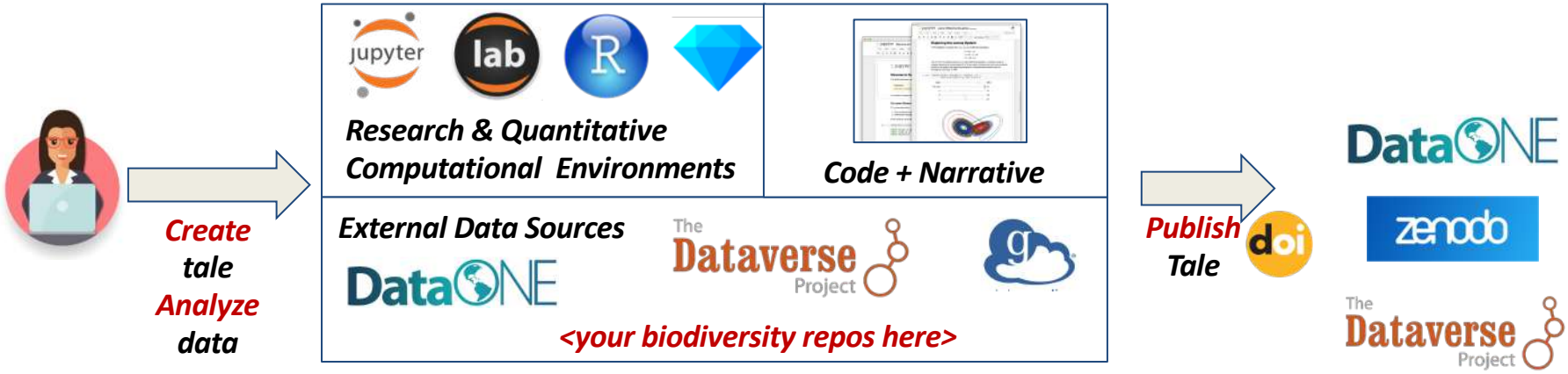
What's in a tale?



- ✓ Tale ~ **Research Object** with
  - *data, code, narrative, compute environment*
- ✓ Executable
- ✓ Publishable

- **Standards-based**
- **Remixable**
- **Transparent (=> *Provenance*!)**
- **Verifiable**

Whole Tale Platform Overview



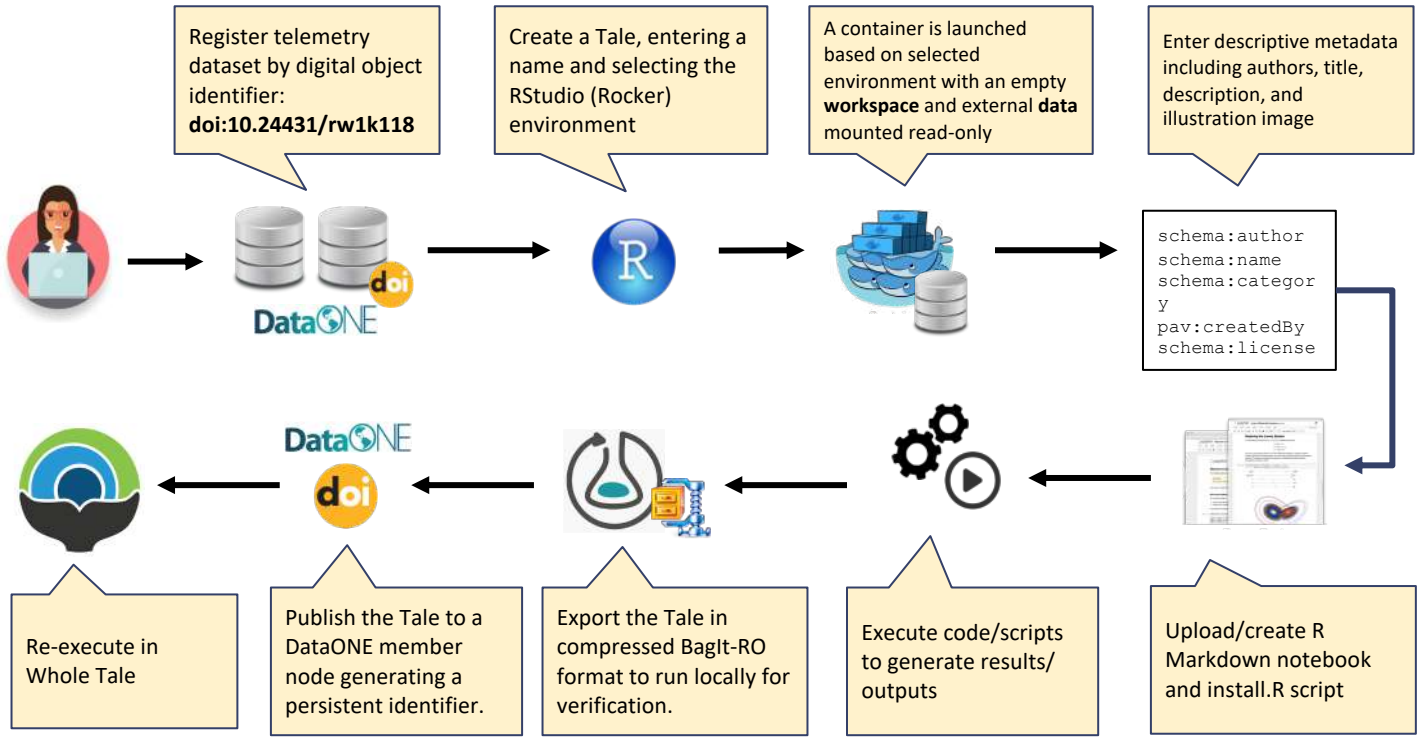
- **Authenticate** using your institutional identity
- **Access** commonly-used **computational environments**
- Easily **customize** your environment (via repo2docker)
- Reference and access externally **registered data**

- Create or upload **your data and code**
- Add **metadata** (including **provenance** information)
- Submit code, data, & environment to **archival repository**
- Get a **persistent identifier**
- **Share** for **verification** and **re-use**

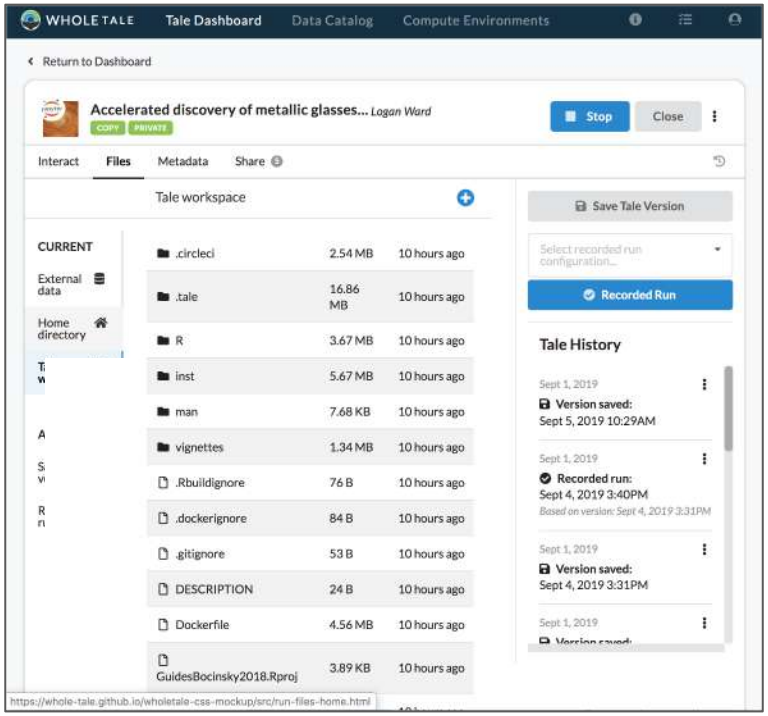
Upcoming Whole Tale releases & new features:

- *WT-v1.1: Git integration; Tale Sharing & Versioning; Support for licensed software (MATLAB and STATA)*
- *WT-v1.2: **Recorded Runs**; Publishing Images*

Tale Creation Workflow

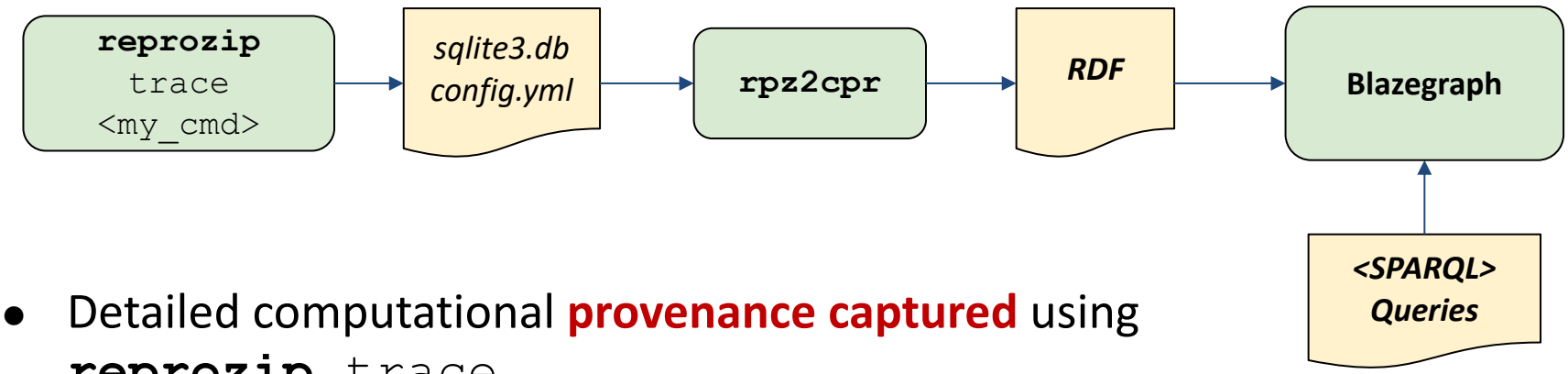


Some new, related features: Recorded Run* to support Transparency



- **Automated** workflow execution with **provenance capture**
- User specified **execution entrypoint**
- System provenance captured using **ReproZip**
- Converted to **comprehensive provenance record** (CPR) => **query** and **reason** about provenance => **provenance reports**
- Each recorded run is a **version**
- User can access **past runs**
- **Standards-based Provenance** information included in **published tale**

Recorded Run:
Provenance Capture*



- Detailed computational **provenance captured** using **`reprozip`** `trace`
- ReproZip output converted to CPR as **RDF triples**
- Imported to Blazegraph for **queries** and **reports**

Comprehensive Provenance Record* (CPR)

- **General provenance** model that supports querying & reasoning across multiple "worldviews" => **hybrid** provenance model
- **Retrospective provenance** (system/runtime provenance) (… ptrace/strace via ReproZip …)
- **Prospective provenance** (e.g., YesWorkflow, CWL, …)
- **Language-level provenance** (e.g., SDTL, …)

Recorded Run: Example Queries*

- Q1: Show me all **inputs** and **outputs** of a given **run**
- Q2: Show me what software was **installed** at the time of the run
- Q3: Show me what software packages were **actually used** by the run
- Q4: Show me the **packages**/**versions** used by a **particular script**
- Q5: Show me **scripts** that **use** a particular package/version
- Q6: Show me which **inputs** where **used** or **outputs created** by a particular script
- …

➔ *Through queries and inference rules: additional information can be derived for reports (e.g. Deltas: what was installed by **not** used, ...)*

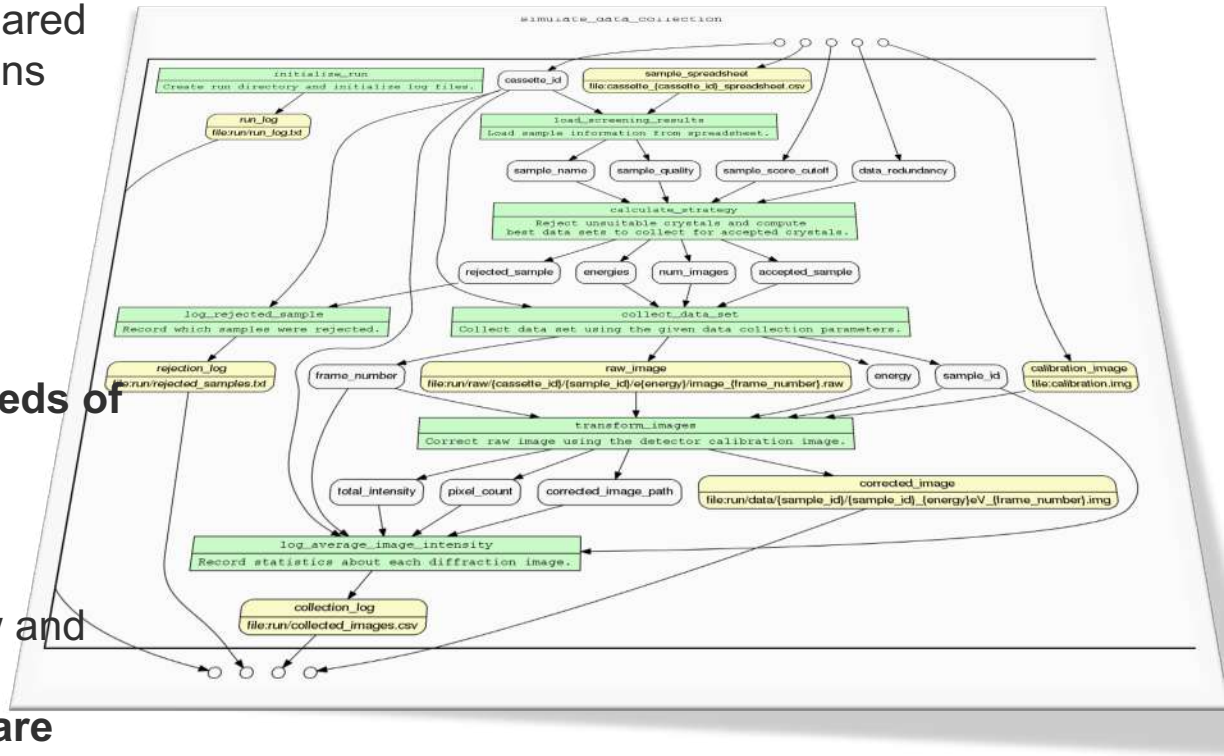
Prospective and **retrospective** provenance: *better together*

- **Prospective** provenance declared using **YesWorkflow** annotations e.g. in Python.
- **Retrospective** provenance captured at run time using **noWorkflow** (or: *Reprozip*, *recordR*, …)
- Script run can produce **hundreds of output files.**
- Each output has a **distinct provenance**.
- **Jointly** querying YesWorkflow and noWorkflow yields answers to **provenance questions that are meaningful to scientists**.

```
…
for energy, frame_number, intensity, raw_image_path in collect_next_image(
        cassette_id, sample_id, num_images, energies,
        'run/raw/{cassette_id}/{sample_id}/e{energy}/image_{frame_number:03d}.raw'):

# @end collect_data_set

# @begin transform_images @desc Correct raw image using the detector calibration image.
# @param sample_id  energy  frame_number
# @in raw_image_path @as raw_image
# @in calibration_image @uri file:calibration.img
# @out corrected_image @uri
file:run/data/{sample_id}/{sample_id}_{energy}eV_{frame_number}.img
# @out corrected_image_path  total_intensity pixel_count

corrected_image_path = 'run/data/{0}/{0}_{1}eV_{2:03d}.img'.format(sample_id, energy,
frame_number)
(total_intensity, pixel_count) = transform_image(raw_image_path, corrected_image_path,
'calibration.img')

# @end transform_images

# @begin log_average_image_intensity @desc Record statistics about each diffraction image.
…
average_intensity = total_intensity / pixel_count
…
```

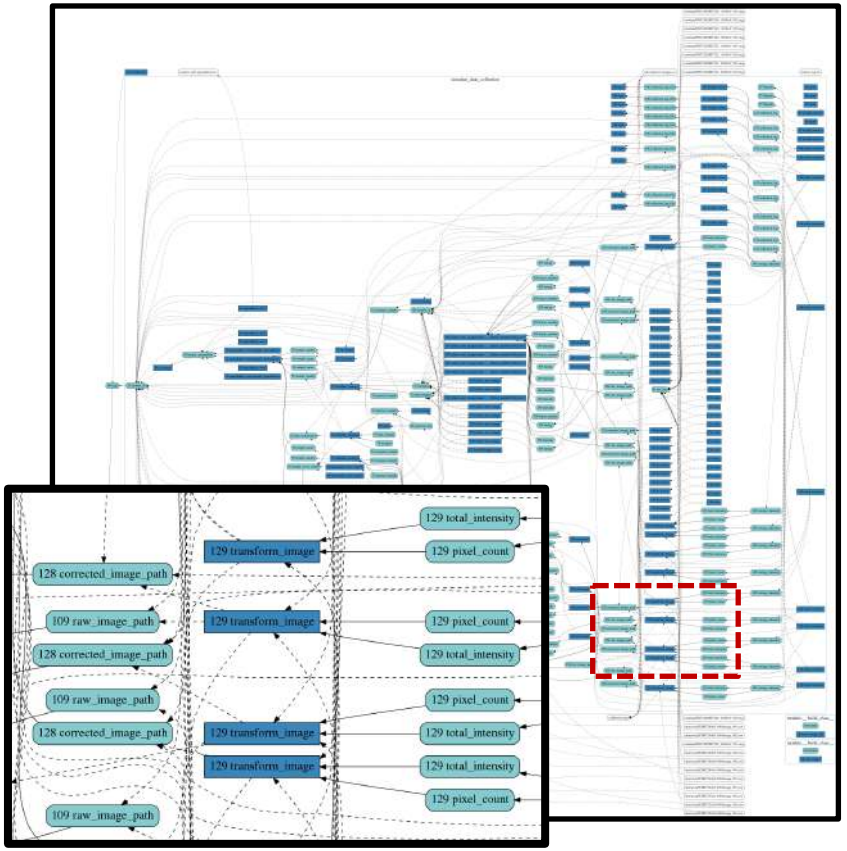
Prospective and **retrospective** provenance: *better together*

- **Prospective** provenance declared using **YesWorkflow** annotations e.g. in Python.
- **Retrospective** provenance captured at run time using **noWorkflow** (or: *Reprozip*, *recordR*, …)
- Script run can produce **hundreds of output files.**
- Each output has a **distinct provenance**.
- **Jointly** querying YesWorkflow and noWorkflow yields answers to **provenance questions that are meaningful to scientists**.



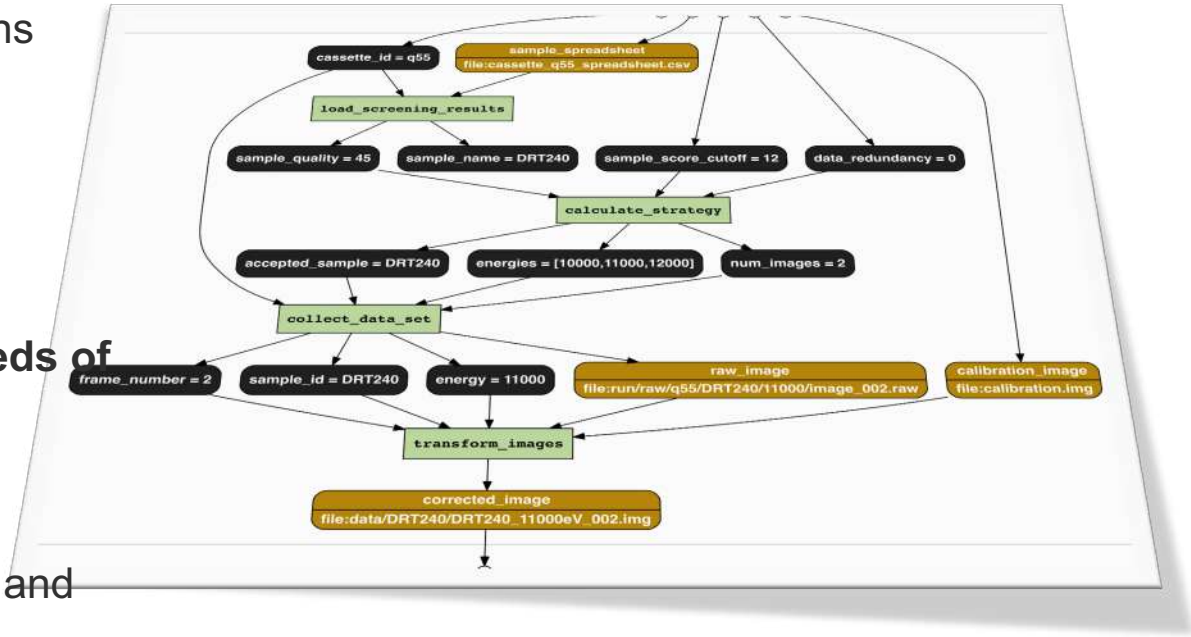
Prospective and **retrospective** provenance: *better together*

- **Prospective** provenance declared using **YesWorkflow** annotations e.g. in Python.
- **Retrospective** provenance captured at run time using **noWorkflow** (or: *Reprozip*, *recordR*, …)
- Script run can produce **hundreds of output files.**
- Each output has a **distinct provenance**.
- **Jointly** querying YesWorkflow and noWorkflow yields answers to **provenance questions that are meaningful to scientists**.



Prospective and retrospective provenance: *better together*

- **Prospective** provenance declared using **YesWorkflow** annotations e.g. in Python.
- **Retrospective** provenance captured at run time using **noWorkflow** (or: *Reprozip*, *recordR*, …)
- Script run can produce **hundreds of output files.**
- Each output has a **distinct provenance**.
- **Jointly** querying YesWorkflow and noWorkflow yields answers to **provenance questions that are meaningful to scientists**.

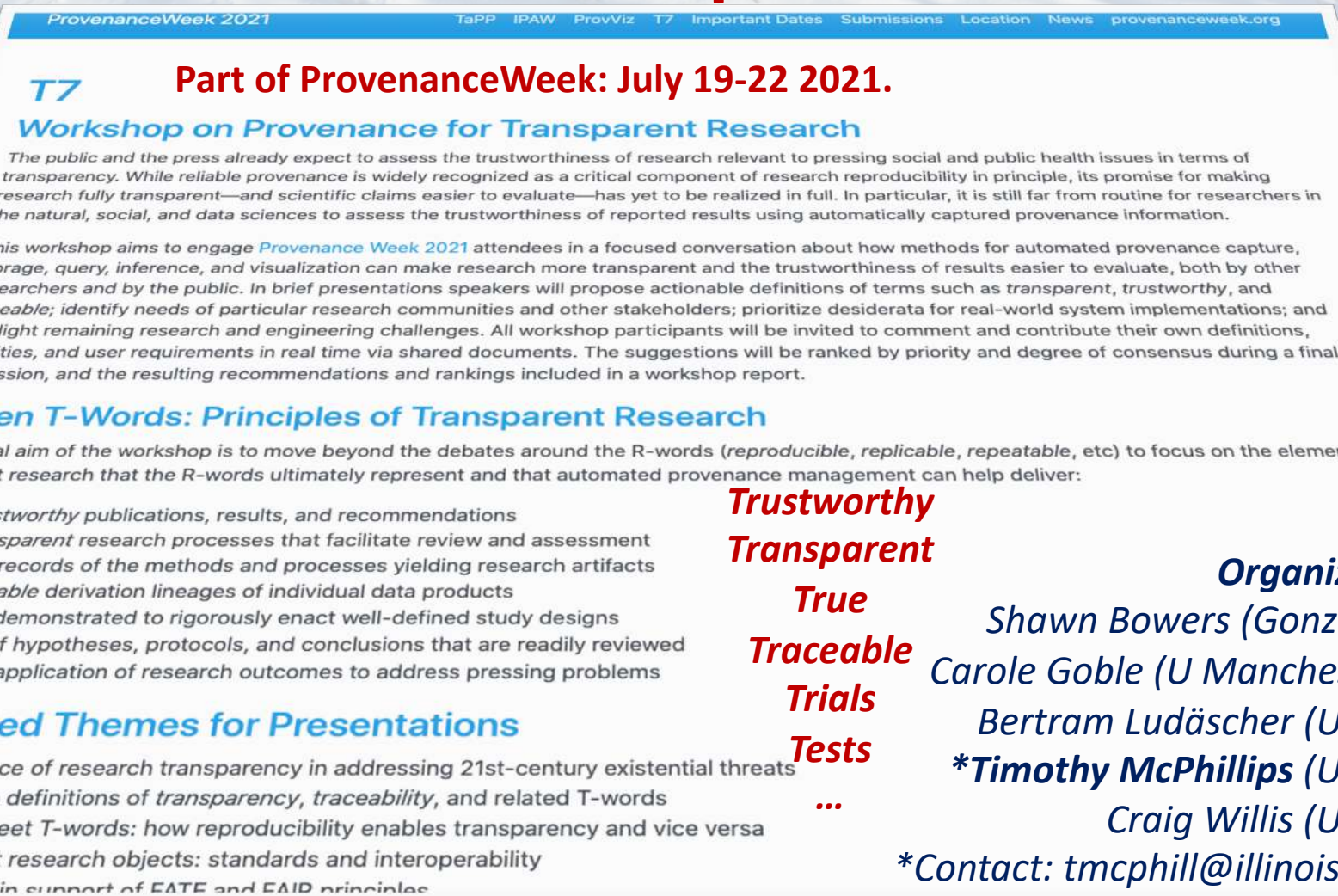


Takeaway Points

- ***Computational reproducibility*** doesn't mean what you might think it means (≈ ***re-executability***)
- ***Computational reproducibility*** is **not required** for ~~reproducible~~ science
- ***Transparency*** on the other hand, **is required** for science.
- Both have a place in (data- and compute-intensive) scientific publishing
  - *You still need to read & understand the paper! (and maybe the code!?)*
  - *Special use cases, e.g. Craig Willis' thesis: Trust but verify => support for "validation workflows" (cf. "badging")*
  - *In economics, social sciences => cf. Lars Vilhuber's work*
- Opportunity cost by getting stuck with R-words =>

Shifting attention from R-words to T-words

T7 Workshop on
Provenance for Transparent Research



ProvenanceWeek 2021 TaPP IPAW ProvViz T7 Important Dates Submissions Location News provenanceweek.org

Part of ProvenanceWeek: July 19-22 2021.

T7 Workshop on Provenance for Transparent Research

The public and the press already expect to assess the trustworthiness of research relevant to pressing social and public health issues in terms of transparency. While reliable provenance is widely recognized as a critical component of research reproducibility in principle, its promise for making research fully transparent—and scientific claims easier to evaluate—has yet to be realized in full. In particular, it is still far from routine for researchers in the natural, social, and data sciences to assess the trustworthiness of reported results using automatically captured provenance information.

This workshop aims to engage Provenance Week 2021 attendees in a focused conversation about how methods for automated provenance capture, storage, query, inference, and visualization can make research more transparent and the trustworthiness of results easier to evaluate, both by other researchers and by the public. In brief presentations speakers will propose actionable definitions of terms such as transparent, trustworthy, and traceable; identify needs of particular research communities and other stakeholders; prioritize desiderata for real-world system implementations; and highlight remaining research and engineering challenges. All workshop participants will be invited to comment and contribute their own definitions, priorities, and user requirements in real time via shared documents. The suggestions will be ranked by priority and degree of consensus during a final discussion, and the resulting recommendations and rankings included in a workshop report.

Seven T-Words: Principles of Transparent Research

A central aim of the workshop is to move beyond the debates around the R-words (reproducible, replicable, repeatable, etc) to focus on the elements of excellent research that the R-words ultimately represent and that automated provenance management can help deliver:

- *Trustworthy* publications, results, and recommendations
- *Transparent* research processes that facilitate review and assessment
- *True* records of the methods and processes yielding research artifacts
- *Traceable* derivation lineages of individual data products
- *Trials* demonstrated to rigorously enact well-defined study designs
- *Tests* of hypotheses, protocols, and conclusions that are readily reviewed
- *Timely* application of research outcomes to address pressing problems

Suggested Themes for Presentations

- *Significance of research transparency in addressing 21st-century existential threats*
- *Actionable definitions of transparency, traceability,* and related T-words
- *R-words meet T-words: how reproducibility enables transparency and vice versa*
- *Transparent research objects: standards and interoperability*
- *Provenance in support of FATE and FAIR principles*

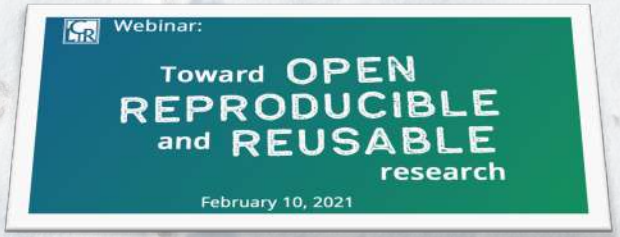
Trustworthy
Transparent
True
Traceable
Trials
Tests
…

Organizers:
Shawn Bowers (Gonzaga)
Carole Goble (U Manchester)
Bertram Ludäscher (UIUC)
***Timothy McPhillips** (UIUC)*
Craig Willis (UIUC)
**Contact: tmcphill@illinois.edu*

https://iitdbgroup.github.io/ProvenanceWeek2021/t7.html

Opportunities for future work …

- There are many opportunities, e.g., …
- 1) Sorting out **terminological** issues (NAS vs FASEB vs ACM …)
- 2) … **Information Gain** / **PRIMAD**$^{+}$ (PRIMAD 2.0) !?
- 3) **Provenance Tools R&D** : *Provenance* => *Transparency* => *Science*
  (… for a suitable definition of "=>" …)
- 4) Join ***T7 Workshop on Provenance for Transparent Research!***



References

- McPhillips, Timothy, Craig Willis, Michael R. Gryk, Santiago Nunez-Corrales, and Bertram Ludäscher. **Reproducibility by other means: Transparent research objects**. In *2019 15th International Conference on EScience (EScience)*, pp. 502-509. IEEE, 2019
- Rauber, A; Braganholo, V; Dittrich, J; Ferro, N; Freire, J; Fuhr, N; Garijo, D; Goble, C; Järvelin, K; Ludäscher B; Stein B; Stotzka R: **PRIMAD: Information gained by different types of reproducibility**. In: *Reproducibility of Data-Oriented Experiments in e-Science (Seminar 16041)*. Vol. 6, Leibniz-Zentrum für Informatik, Schloss Dagstuhl.
- Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M.B., Kowalik, K., Kulasekaran, S., Ludäscher, B., Mecum, B.D., Nabrzyski, J. and Stodden, V., 2019. **Computing environments for reproducibility: Capturing the "Whole Tale"**. *Future Generation Computer Systems*, *94*, pp.854-867.
- McPhillips, Song, Kolisnik, Aulenbach, Belhajjame, Bocinsky, Cao, Cheney, Chirigati, Dey, Freire, Jones, Hanken, Kintigh, Kohler, Koop, Macklin, Missier, Schildhauer, Schwalm, Wei, Bieda, Ludäscher (2015). **YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts**. *International Journal of Digital Curation (IJDC)* 10, 298-313.
- João Pimentel, Saumen Dey, Timothy McPhillips, Khalid Belhajjame, David Koop, Leonardo Murta, Vanessa Braganholo, Bertram Ludäscher. **Yin & Yang: Demonstrating Complementary Provenance from noWorkflow & YesWorkflow**. *Intl. Workshop on Provenance and Annotation of Data and Processes (IPAW)* LNCS 9672, 2016.
- Craig Willis. **Trust, but verify: An investigation of methods of verification and dissemination of computational research artifacts for transparency and reproducibility**. *PhD Thesis*, University of Illinois, Urbana-Champaign, 2020.