

- **Scope & Overlaps**
- **Introduction**
  - **Meaning & Terminology**
  - **Focus**
- **Experiences**
- **Expectations Current Focus**
- **Expectations Pre-Acceptance**
  - Expectations of Authors
  - Expectations of Organisers
  - Expectations of Reviewers
- **Expectations Post-Acceptance / Post-Publishing**
  - Expectations of Authors
  - Expectations of Organisers
  - Expectations of Reviewers
- **Investigation and Retraction To be Completed**
- **Afterward**
  - **Contributors / Authors**
  - **Colophon**

% Replication and Independent Verification

% Laleh Behjat, Bruce Childers, Eric Eide, Grigori Fursin, Yolanda Gil, Simon Harper, Alex K. Jones, Shriram Krishnamurthi, Jimmy Lin, Patrick Madden, Catherine McGeoch, and Michela Taufer.

% 16 October 2015

Replication and independent verification of results are commonplace in many scientific fields. These practices help avoid misunderstandings, provide greater confidence in reported results, and enable the detection of both unintentional error and deliberate fraud. By comparison with fields such as physics, chemistry, and medicine, computer science is clearly lagging in this regard; replication of experiments is rare, and activities such as paper correction or retraction are virtually unheard of.

This document corrects this situation by setting out the view of the ACM SGB in relation to replication and independent verification of results. Therefore SIG focused views and requirements can be adopted into the central ACM publications process.

## **Scope & Overlaps**

# Introduction

The scientific method is based on empirical data and the replicability of those experiments used to collect this data by third parties. However, in Computer Science the ability to replicate experiments is limited because there are few peer reviewed routes for experimental replication.

Our limited ability to replicate experiments – due to a lack of public data – means that some domains do not strictly follow the scientific method, and as such are sometimes seen as less rigorous than other scientific domains. Indeed, we may go further and suggest that:

1. Methodical data collection and experimentation is variable because the data may never be seen by a third party;
2. Experiments to confirm the science are not repeated as the data and experimentation methodology is not available, and without a publication route there is no value in doing so; and
3. Data collection and accurate experimentation is not seen as being important for citation and publication of the data alone, and is seen more as a second class citizen to a system-based technical paper.

Other sciences (Physics, Ecology / Biology, Medicine etc.) recognise the value of their data as it costs (often) enormous amounts of time, effort, and money to acquire. As a discipline we need to plug this gap. In this case, the charge of the task force is to develop proposals on how ACM can bring current replication and verification practices in line with the rest of the scientific community. Specifically, the following questions are of interest.

1. **Experiences:** Across the SIGs, what mechanisms are in place to enable replication and verification efforts? Are they sufficient to catch errors or deliberate fraud (and have they in fact done so?). Can these processes be improved or encouraged in some way?
2. **Expectations:** What are reasonable expectations from authors, with regards to enabling replication and verification? How can these expectations be made compatible with the review process in conferences, symposia, journals? Should there be different expectations during review, and after paper acceptance?
3. **Retractions:** If published work fails to be replicated, appears to contain significant errors, or even appears to be fraudulent – what mechanisms should be in place to investigate the matter? Who would such concerns be communicated to? Who would perform investigation, and decide the matter? And if errors, deception, or fraud are confirmed – what should the ACM, as a publisher, do? Should there be corrections, annotations noting the errors, paper retraction?

The SIGs, with elected leaders representing the various research communities of computer science, are a logical group to engage in setting policy. The SIG leaders will be familiar with typical practices within their areas of expertise. With control of many of the major publishing venues, this group also has the ability to enact change.

## Meaning & Terminology

The critical point for widespread adoption of replication is a shared understanding of what this means, what items actually support replication, and what a flexible framework to house these aspects looks like. Flexible is key if we are to accommodate all ACM SIGs, while having enough commonalities to present a holistic structure.

## Focus

The key issue here is the replication and support or repudiation of the work based on the claims set out in the accompanying paper (on which the evaluation of the research artefacts are based). Along with the ability to convey this information, and the results produced, at a high level to interested parties who may not be technical specialists within the area.

We are trying to convey surety to other researchers and the public at large that this work is reproducible, or has the potential to be reproducible. And that any discussions surrounding repudiation and retraction are open and transparent including (maybe especially) when there is a suggestion of academic malpractice.

# Experiences

Our initial aim was to collect and understand the experiences of different groups within the ACM SIG structure. This was, so that we could decide what approaches to pursue for the subsequent **expectations** and **retractions** best practice, and which approaches could be implemented for all. However as detailed in the “Scope and Overlap” section the collection of best practice and experiences was already well underway as part of the remit of the ACM-DSR. In this case, we chose to adopt their best practices so as not to repeat work and so that we could move quickly to discussing **expectations**.

This said we did contribute best practice and current work around Artefact Evaluation and the experiences of SIGPLAN and SIGSOFT, and the CGO and PPOPP (2015, 2016) conferences. This best practice was included in the ACM-DSR documents.

To summarise, the ACM-DSR have collected around 650 pages of best practice experiences. Have summarised these as a 47 page document, and have further condensed this as a 10 page executive summary. The outcomes of this summarisation process can be thought of as follows.

A One Page Summary will be placed here once the final small document is created

## Expectations **Current Focus**

When discussing replication we need to understand what are reasonable expectations. Reasonable expectations from authors with regards to enabling replication verification of their paper – based on the artefacts supplied as accompanying materials to that paper. We must also understand what expectations we have of the review process for conferences, symposia, and journals. Still further, we must also understand what expectations we have of volunteer reviewers and what benefits to a reviewer this enhanced replication process may have (or can be offered).

Realising that a volunteer reviewer is likely to be another academic and that the currency of this academic may be visibility and citation as well as engagement with the community and volunteerism is critical. Finally we need to answer questions as to the different expectations we may have from papers during review and those after paper acceptance.

Therefore, we are interested in detailing what are reasonable expectations. And what are possible inducements for all three stakeholders (the authors, the organisers, and the reviewers). Further describing, how these can be accommodated across the diverse Special Interest Groups which make up the bulk of the membership of the ACM.

We now need to build a set of expectations and detail them along with rationale here

## Expectations Pre-Acceptance

**Expectations of Authors**

**Expectations of Organisers**

**Expectations of Reviewers**

## Expectations Post-Acceptance / Post-Publishing

**Expectations of Authors**

**Expectations of Organisers**

**Expectations of Reviewers**

# Investigation and Retraction **To be Completed**

## Afterward

The ACM SGB Replication Taskforce was initiated at the request of the SGB Chair Patrick Madden in September 2015. SIG Development Advisor Simon Harper was appointed Taskforce Chair to take this forward as well as integrating it into the ‘DL-Technology: Data, Software, & Reproducibility’ project.

The charge of the task force was to develop proposals on how the ACM can bring current replication and verification practices in line with the rest of the scientific community. Given the range of research topics and problems considered, there was no one-size-fits-all policy. With this in mind this group was open to all voices across the SGB and composed SGB members and other invited experts. The document is also intended to evolve (hence the versioning here, and the versioned releases) and be transparent. By engaging many SIGs in this effort, a degree of consistency was achieved across the ACM.

## Contributors / Authors

- Behjat, Laleh
- Childers, Bruce
- Eide, Eric (SIGOPS)
- Fursin, Grigori
- Gil, Yolanda (SIGAI)
- Harper, Simon (Taskforce Chair)
- Jones, Alex K.
- Krishnamurthi, Shriram (SIGPLAN)
- Lin, Jimmy (SIGIR)
- Madden, Patrick (SGB Chair)
- McGeoch, Catherine

## Colophon

This document is created in standard ‘Markdown’ with pandoc citation additions. Critic Markup is used for comments. The output format is as yet unknown and so a basic markdown document which can be easily converted to many formats (via pandoc or mmd) is most appropriate. Further, version control via Git is easier using basic text formats.