(Re)Use of Research Results ... why should we?

Maria Teresa Baldassarre
Department of Informatics – University of Bari
mariateresa.baldassarre@uniba.it
Who am I

⇒ **Associate Professor** at the Department of Informatics - University of Bari (www.di.uniba.it)

⇒ **Coordinator of the «Process & Product Quality» area** @Software Engineering Research LAB (serlab.di.uniba.it)


⇒ **Member** of the International Software Engineering Research Network (ISERN)

@ mariateresa.baldassarre@uniba.it

@ mtbaldassarre
CS Department – Dipartimento di Informatica
BARI - Puglia
Main research interests:

- SOFTWARE PROCESS AND PRODUCT QUALITY
- HUMAN FACTORS IN SOFTWARE ENGINEERING
- EMPIRICAL SOFTWARE ENGINEERING
Agile User&Quality Oriented Development...

Inception
- PO - Product Owner
- SM - ScrumMaster
- T - Team
- CC - Customer Committee

Input from End-Users, Customers, Context and Field Studies

Customer Committee

Sprint n.0
- Product Backlog (Features)
- Software System Skeleton
- Sprint Backlog (Stories)
- Sprint Planning

Development
- (IN)Sprint Review
- Product Backlog Refinement

Delivery
- GIT
- ASANA
- JENKINS
- GRADLE
- KIUWAN
- SONAR CUBE

ISO/IEC 25011: Quality Model Division
ISO/IEC 25001: Quality Measurement Division
ISO/IEC 25002: Quality Management Division
ISO/IEC 25003: Quality Requirements Division
ISO/IEC 25004: Quality Evaluation Division

Experience Factory: Organization
Feedback Loop to Customer

Customer-Ready Product Increment
- Incremental Product Release
- Project Retrospective
...Agile User&Quality Oriented Development

- **BAMBOO**: continuous integration/deployment
- **JIRA**: application lifecycle management
- **SONARQUBE**: quality management
- **SVN**: version control system
Spin-off of the University of Bari - established in 2006.

30 employees

- 9001:2008 - Quality management systems - Requirements
- 14001: 2004 - Environmental management systems
- 25000:2014 - Systems and software engineering – First in Italy to assess certification of a sofware product
- SERLab carries out research and empirical validation of results
- SER&P transfers the results of these activities to industry; provides data and industrial context for field experimentation
INDUSTRIAL COLLABORATIONS
Is it important for a scientist to Report Research Results so others can (Re)Use them?
"... the ideas we can most trust are those that have been the most tried and tested.

For that reason many of us are involved in this process called ‘science’ which produces trusted knowledge by sharing one’s ideas and trying out and testing the ideas of others ... "

cit. Popper
Produce & Report research results
ReUse results/findings ...

... to improve reproducibility and transparency
«RESULTS PARADOX»

CREATE OBLIVIATE
«RESULTS PARADOX»

«FACTS & TRUTH»
Keep research results at arm’s length

Objective investigator – detective
Follows data with discipline; never indulges in data massaging or cherry picking

«BE PERSUASIVE»
Pressure of publishing clear novel and positive findings on behalf of funding agencies, evaluation committees

Good lawyer
Arguments and produces amounts of beautiful and convincing results

https://doi.org/10.1038/s41562-021-01193-7
Researchers attempt to solve this paradox ... questionable research practices ... reduce confidence of conclusions ... harm reproducibility ...
Questionable Research Practices (QPRs) Hurt Science ...

HARKing (Hypothesizing After Results are Known)
- Neat data, what explains it?
  - Acceptable in explanatory not confirmatory

Post-hoc Rationalizing
- Story-telling to explain the data found in a study
  - Acceptable in explanatory/inductive theory building not confirmatory

... Questionable Research Practices Hurt Science

File-drawer effect

- Hmm, bad outcome, bin it. Negative result – reject. Not published. Do not appear in meta-analysis and SLRs

Forking paths in data analysis choices after seeing the data (Researcher Bias)

- Let’s use a Kruskal-Wallis test and then a Lewandoski-Neymar test of significance (instead of?)

QRPs result when publication **venue** and publication **significance/novelty** are emphasized over replication & soundness of the method
Registered Reports

free researchers from the pressure to engage in QRPs

Avoid the RESULTS-ORIENTATION
Deal with RESEARCHER BIAS

Focus on SOUNDNESS OF THE RESEARCH PLAN & SIGNIFICANCE OF THE RESEARCH QUESTION

Registered Reports ... why?

Pre-registration (clinical trials): register your protocol including planned hypothesis, data collection, data analysis that is «registered» BEFORE the study is conducted.

Protocol commits to analysis and expected outcomes.

Registered Report: Peer-reviewed pre-registration.
... Registered Reports ... why?

⇒ Benefits

- **Feedback**: Provide feedback at early phase of research (before spending $$$)
- **Reduce/eliminate** under-powered, selectively reported, researcher-biased studies
- **Quality**: Help improve study quality and scientific impact
RAPID RISE
Since 2013, the number of journals offering Registered Reports (RRs) has risen to more than 200 titles.

- First multidisciplinary journal launches RRs across 200 sciences (Royal Society Open Science).
- BMC Medicine launches first RRs for clinical trials.
- First journal exclusively for RRs (Comprehensive Results in Social Psychology).
- Publication of 100th completed RR.

(*As of June 2019*)
RR in SW_Engineering

EMSE J. → MSR,
ICSME, then ESEM,
now CHASE, SANER,
ICPC
TOSEM (direct submit)
CSE special issue

(ACM, Springer, T&F)
Fig. 1 Stages of the Registered Reports workflow. Center for Open Science (https://www.cos.io/initiatives/registered-reports/#tabid3) CC-BY-NoDerivs 4.0
Phase 1 – Review Criteria

Is this study novel, significant, able to find effects?

1. **Importance** of the research question(s).
2. Logic, rationale, and plausibility of the proposed hypotheses.
3. **Soundness** and **feasibility** of the methodology and analysis pipeline (including statistical power analysis where appropriate).
4. **Clarity** and degree of methodological detail for replication.
5. Will results obtained **test** the stated hypotheses?
Phase 2 – Review Criteria

Did the authors execute on Phase 1 plan?

1. Whether the data are able to test the authors’ proposed hypotheses by satisfying the approved outcome-neutral conditions (such as quality checks, positive controls)
2. Whether the Introduction, rationale and stated hypotheses are the same as the approved Stage 1 submission (required)
3. Whether the authors adhered precisely to the registered experimental procedures
4. Whether any unregistered post hoc analyses added by the authors are justified, methodologically sound, and informative
5. Whether the authors’ conclusions are justified given the data
Current state of RR in SE

MSR 2020 feedback on IPA:

“I think it is a key principle. However, in a way it also raises the bar significantly for the Registered Reports”

“[...] the fact that the results are missing, helps reviewers and authors focus on the methodological issue, which is a great added value in the review process [...]”
MSR Results - IPA

“During my review, though, I had the feeling that more interaction with the authors could add even further value”
“I think the EMSE paper still needs a careful assessment, as it is still possible that the operation or the application of the protocol turns out to be wrong [...]”
“I felt a bit uncomfortable to have this burden on my shoulders as a reviewer so early in the process.”
No (3 responses):
“A registered report may be, and should be allowed to be, risky and, therefore, may not work out. The ensuing work should be subject to full and normal review.”
In general, would you participate again (as reviewer or authors)?

25 responses

Table 1  RR submissions and publications since inception at EMSE

<table>
<thead>
<tr>
<th>Venue</th>
<th>Stage 1</th>
<th></th>
<th>Stage 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Submissions</td>
<td>IPAs</td>
<td>Submissions</td>
<td>Publications</td>
</tr>
<tr>
<td>MSR 2020</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>MSR 2021</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>MSR 2022</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ICSME 2020</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>ICSME 2021</td>
<td>n/a</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>ESEM 2021</td>
<td>n/a</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ESEM 2022</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that some studies were affected by the COVID-19 pandemic. Data may be incomplete as tracking submissions can be challenging.
Open Issues and Questions
Pros & Cons of RR

RRs provide early-stage feedback to authors and reduce researcher bias problems

Table 2  Benefits and disadvantages of registered reports in SE

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shareable protocols for research replication.</td>
<td>More effort from researchers.</td>
</tr>
<tr>
<td>Focus is on research, not publication.</td>
<td>Limited acceptance by journals so far.</td>
</tr>
<tr>
<td>Improved rigour in reporting.</td>
<td>Rigour can mean different things to different people/communities</td>
</tr>
<tr>
<td></td>
<td>(Storey et al. 2020).</td>
</tr>
<tr>
<td>Early peer review on research approach.</td>
<td>Not all research strategies are registerable.</td>
</tr>
</tbody>
</table>
Three faces of RR

RR to prevent questionable research practices

Tell the world what you will do, then do it

RR as doctoral symposium

Early feedback before expensive data collection

RR as 1st round review

Pre-empt journal review with in-principle acceptance

To what CS studies could it apply?

Most suited to post-positivist, confirmatory studies with clear hypotheses.
Admin Challenges

CS has conference and journals - no one else does
Journals and conference rarely share admin interfaces (HotCRP vs Editorial Manager - and they are usually terrible)
Hard to manage reviewer discussions esp longitudinally
Currently, stick Phase 1 on Arxiv/OSF.io/Github

Have to explicitly coach reviewers (not yet mature, but true of other formats)

Manually track in progress RR on Google Sheets (low vacation factor)
Admin Challenges

**Reviewer/editor burden** is increasingly a problem (overall, not just RR)
Accepting 5 IPAs at 3 conferences a year = 15 journal submissions in the next 12-18 months, with publication 24-36 months after that
+ who is asked to be conference track chair? What freedoms do they have?

**Minor shenanigans** - reviewer COI, authorship incentives
Admin Challenges – J1C2?

Publication models run into journal profit models

First phase - Journal - then present at conference?

Steffen Herbold @stherbold · Sep 28
Replying to @xdevroey @AndreasZeller and 3 others
Fees for registered report tracks are also a problem. @acm_tosem is now an alternative to the (awesome!) @emsejournal registrations at confs (e.g. @msrconf @ESEM_conf @IEEEICSME). The costs were okay for virtual conferences, but full fees to *register* a study is too much.
Enhance Reproducibility
- Standardization of submitted protocols

Are more likely to report Negative Results

Reviewers can help authors improve the protocol beforehand -> prevents flaws

Are a PLAN.... Not a PRISON
- Flexibility is not lost ... rather the possibility of airbrushing changes out of the picture
Department of Reuse

Ultimately RR is about pre-specifying analysis. One way to do that is to reuse analysis protocols from other papers.

Done all the time in medicine; rarely in CS except in benchmarks.

Q: to what extent are artifacts such as protocols reused?

https://reuse-dept.org/
Artifacts Creation, sharing and Reuse

SE researchers share artifacts
Not only publications... Ideas, methods, datasets, tools

Artifacts engage replication and reproducibility

Science produces more types of artifacts than just publications

Researchers use some but not necessarily all artifacts from other work

"HOW DO WE CAPTURE REUSE?"
Badging – Artifact Evaluation Committees

The authors of accepted conference papers submit software packages that, in theory, let others re-execute that work. These evaluation committees award “badges”

Table 1. Badges such as the ones shown in this table are currently awarded at conferences. This table is based on ACM’s badge program, however, analogous badges are used at other conferences. Images used by permission of the Association for Computing Machinery.

<table>
<thead>
<tr>
<th>Available</th>
<th>Functional</th>
<th>Reusable</th>
<th>Reproduced</th>
<th>Replicated</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Available Badge" /></td>
<td><img src="image2.png" alt="Functional Badge" /></td>
<td><img src="image3.png" alt="Reusable Badge" /></td>
<td><img src="image4.png" alt="Reproduced Badge" /></td>
<td><img src="image5.png" alt="Replicated Badge" /></td>
</tr>
<tr>
<td>In a public repository with a long-term retention policy. A DOI needs to be provided.</td>
<td>Artifacts are documented, consistent, complete, exercisable, and include evidence of verification and validation.</td>
<td>Functional, significantly exceed minimal functionality.</td>
<td>Results of this paper have been reproduced by a different team using the original artifact.</td>
<td>Results of this paper have been replicated by a different team without the original artifact.</td>
</tr>
</tbody>
</table>
Badging – Artifact Evaluation Committees

Fig. 4. Artifact evaluation committee sizes 2011-2019. From Hermann et al. [8]
Is the artifact evaluation process is creating reused artifacts?

We queried ACM Portal for ICSE papers between 2011 to 2021, to find 2.4% of papers with an artifact badge.

Of these, 111 available, 74 reusable, 24 functional, NO replicated or reproduced artifacts.

approach to recording Research Reuse -> REUSE GRAPH
 Researchers read 170 SE papers selected from 6 major 2020 conferences

Teams were asked to record six types of reuse

Each edge connects papers to the prior work they are (re)using
This figure shows reuse from Bernal-Cárdenas et al. Edges reflect tool, dataset, and methodology reuse. Red nodes indicate arXiv preprint; green represents a GitHub repository; blue denotes a published paper, and grey indicates other websites or grey literature locations. https://www.reuse-dept.org/doi/10.1145/3377811.3380328.
The Rose Initiative (Recognizing and Rewarding Open Science in Software Engineering) is an international, multi-conference workshop that will continually report updates to the software engineering reuse graphs.
Registered reports in software engineering

Neil A. Ernst¹ - Maria Teresa Baldassarre²

Accepted: 14 December 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Registered reports are scientific publications which begin the publication process by first having the detailed research protocol, including key research questions, reviewed and approved by peers. Subsequent analysis and results are published with minimal additional review, even if there was no clear support for the underlying hypothesis, as long as the approved protocol is followed. Registered reports can prevent several questionable research practices and give early feedback on research designs. In software engineering research, registered reports were first introduced in the International Conference on Mining Software Repositories (MSR) in 2020. They are now established in these conferences and two pre-eminent journals, including this one (EMSE). We explain the motivation for registered reports, outline the way they have been implemented in software engineering, and outline some ongoing challenges for addressing high quality software engineering research.

Keywords Registered report - Research methods - Software engineering

1 Introduction
Registered reports are a model of scholarly publication which prioritize the importance of study design and significance rather than study outcomes. Focusing on whether the study was suitable to support the inferences of interest decouples publication from a focus on headline-worthy ‘significant’ results.

In software engineering (SE) research, empirical methods are now standard. The top conferences in the field emphasize “the extent to which the paper’s contributions and/or

Communicated by: Robert Feld and Thomas Zimmermann

Neil A. Ernst  
nerst@uvic.ca

Maria Teresa Baldassarre  
mariateresa.baldassarre@uniba.it

¹ Department of Computer Science, University of Victoria, Victoria, BC, Canada
² Dipartimento di Informatica, Università degli studi di Bari, Bari, Italy

Published online: 11 March 2023 © Springer
Credits & Special Thanks
References


Figure. Watch the authors discuss this work in the exclusive Communications video. https://cacm.acm.org/videos/reuse-of-research
mariateresa.baldassarre@uniba.it