Engineering Fracture Mechanics 197 (2018) 21–26

Contents lists available at ScienceDirect

Engineering Fracture Mechanics

journal homepage: www.elsevier.com/locate/engfracmech

Technical Note

Long term availability of raw experimental data in experimental fracture mechanics

Patrick Diehl[a,*], Ilyass Tabiai[a], Felix W. Baumann[c,b], Daniel Therriault[a], Martin Levesque[a]

[a] Department for Multi Scale Mechanics, Polytechnique Montreal, C.P. 6079, succ. Centre-Ville, Montreal, Quebec H3C 3A7, Canada
[b] TWT GmbH Science & Innovation, Stuttgart, Germany
[c] Institute of Computer-Aided Product Development Systems, University of Stuttgart, Stuttgart, Germany

ARTICLE INFO

Keywords:
Long term availability of data
Experimental fracture mechanics
Open data

ABSTRACT

Experimental data availability is a cornerstone for reproducibility in experimental fracture mechanics, which is crucial to the scientific method. This short communication focuses on the accessibility and long term availability of raw experimental data. The corresponding authors of the eleven most cited papers, related to experimental fracture mechanics, for every year from 2000 up to 2016, were kindly asked about the availability of the raw experimental data associated with each publication. For the 187 e-mails sent: 22.46% resulted in outdated contact information, 57.75% of the authors did received our request and did not reply, and 19.79 replied to our request. The availability of data is generally low with only 11 available data sets (5.9%). The authors identified two main issues for the lacking availability of raw experimental data. First, the ability to retrieve data is strongly attached to the possibility to contact the corresponding author. This study suggests that institutional e-mail addresses are insufficient means for obtaining experimental data sets. Second, lack of experimental data is also due that submission and publication does not require to make the raw experimental data available. The following solutions are proposed: (1) Requirement of unique identifiers, like ORCID or ResearcherID, to detach the author(s) from their institutional e-mail address, (2) Provide DOIs, like Zenodo or Dataverse, to make raw experimental data citable, and (3) grant providing organizations should ensure that experimental data by public funded projects is available to the public.

1. Introduction

Reproducibility is the ability to obtain the same research results as another researcher, given the same analysis is done on the same raw data. Reproducibility is crucial to the scientific method [1,2]. Reproducibility in experimental mechanics, can hardly be achieved without access to the raw data used by fellow researchers in their publications. The lacking of scientific reproducibility has been shown for basic and preclinical research [3] and psychological science [4]. In biology [5], a study revealed that raw data sets could be obtained from 19% of 516 papers containing experimental data and published from 1991 to 2011.

Different stakeholders addressed the lacking availability of experimental data. The organization for economic co-operation and development (OECD) was commissioned by different governments to develop a set of guidelines to provide cost-effective access to publicly funded research data [6,7]. Publishers are currently investigating means to strengthen data-access practices [8] or support

* Corresponding author.
E-mail address: patrick.diehl@polymtl.ca (P. Diehl).

https://doi.org/10.1016/j.engfracmech.2018.04.030
Received 20 March 2018; Accepted 17 April 2018
Available online 30 April 2018
0013-7944/

open data [9].

According to a recent study, more than 70% of researchers, out of more than 1500 polled, have tried and failed to reproduce another scientist's experiments [10]. However, the study also shows that physicists and engineers are confident that peer reviewed published data is reproducible.

Most publications in experimental fracture mechanics rely on data gathered from experiments. It has been our experience (and practice) that only quantities of interest are presented and the raw experimental data is usually missing. Moreover, even when experimental data is available, information related to the experimental setup itself is usually sparse (e.g., calibration of the measurement unit, software used, etc.), which prevents the experiment's replication.

The modelling community is also highly interested in high fidelity and well documented experimental data to validate model predictions [11]. The data published in the literature usually lacks information about boundary conditions, etc., to ensure that the models reproduce, at least conceptually, the experiments they aim to reproduce.

This short communication focuses on the accessibility and long term availability of raw experimental data, as well as supporting information, in experimental fracture mechanics. We have contacted the authors of the eleven most cited papers related to experimental fracture mechanics for every year from 2000 up to 2016. We kindly asked these authors about the availability of raw experimental data associated with each publication. The up-to-dateness of the e-mail addresses were studied and the reply-behaviors for working e-mail addresses were considered. Finally, the availability of raw data out of the positive responses from the authors was emphasized.

Section 2 deals with the methodology for this study and the data collection. Section 3 presents the main results. In Section 4, discusses the results and Section 5 concludes the paper.

2. Methodology

2.1. Data collection

The *Web of Science* database[1] was queried with the following fields on September 18th 2017: `TOPIC: (CRACK) AND TOPIC: (DAMAGE) AND TOPIC: (EXPERIMENTAL) AND YEAR PUBLISHED`: (*X*), where *X* varied from 2000 to 2016.

The top eleven cited papers containing experimental data generated by the authors, included as a reference to an online resource or as an appendix, were selected for each of the respective publication year.

We investigated: *(1)* the document's meta-data provided by *Web of Science* (WoS), (*2*) the PDF document itself and finally (*3*) the publisher's website to identify the corresponding authors email contacts. Less than 25% of the papers published between 2000 and 2004 contained an e-mail address in the meta data from WoS while 80% of the papers published after 2004 contained that information. The first corresponding author was selected for the communication attempt. The full list of references investigated can be found on Github under the BibTEX format.[2]

The generic email detailed in Appendix B was sent to the selected authors on October 16th 2017. We asked the authors if they were willing to share their experimental data and if so, how long, in minutes, would it take to gather.

The prepared prescribed answers were: *(a)* data is not available, (*b*) the data is confidential, (*c*) one of the co-authors should be contacted to obtain the data. Furthermore, an open answer was available to the authors that were unable, or unwilling, to share this information.

A reminder was sent on November the 6th (three weeks after the first attempt) to all authors for which we did not receive a reply and for which the e-mail did not bounce. The communication contained the detailed query shown in Appendix C.

Eight e-mails were sent a few weeks after the first iteration, to correct an error in the automated data acquisition. All responses received before December the 15th 2017 were considered in this survey.

All e-mails were sent using the institutional e-mail address of one of the authors, as in other studies. The possibility for a biased reply-behavior when sending the e-mail as a student or as a professor was not addressed.

3. Results

Out of the 187 papers selected, only one publication provided the experimental raw data attached as supplementary data on the journal's website.

Table 1 lists the data analysis for the 187 e-mails sent. The first column presents the number of e-mails that bounced. The second column shows the number of replies to either the first or second e-mail. Note that there is no distinction between a positive or negative reply with respect to sharing the data. Only 30 authors that did not respond to the first e-mail responded after receiving the second. The third column presents the number of no replies, which means that we did not obtain an error from the mail server and no answer 6 weeks after sending the first e-mail.

Table 1 also lists the time, in minutes, required for the authors to retrieve the data (for those willing to share it). The following columns list the reasons the authors invoked for not providing the requested data. The last column lists the amount of available data sets.

[1] https://webofknowledge.com.

[2] https://github.com/OpenDataExpMechanics/Survey.