Findable, Accessible, Interoperable and Reusable (FAIR) Digital Image Correlation Data

M. Sampson[1a], L. Fletcher[1] , A. Tayeb[1], A. Harte[1], and C. Hamelin[1]

[1]UK Atomic Energy Authority, Culham Campus, Abingdon, OX14 3DB, United Kingdom

[a]megan.sampson@ukaea.uk

Abstract.
The FAIR (findable, accessible, interoperable and reusable) data principles aim to maximise the research value of data. However, researchers in the field of experimental mechanics may struggle putting this into practice without specific guidance. In this work we develop a FAIR digital image correlation (DIC) data storage guide, designed to support researchers in structuring their data and metadata that reflects data management best practice.

Possible Sessions

19. Optical and DIC Techniques, 14. Model Validation, 6. Data-Driven Testing

Introduction

With the adoption of digital image correlation in the experimental mechanics community, researchers are producing large amounts of data which will only increase as camera technology improves. Whilst in the short term, these outputs can be handled manually, with time it can become unwieldy. As details of the data and experiment fall out of the mind, it becomes impossible to retrieve and reuse the data. This data can be valuable for things such as machine learning and meta-analyses, so it is important to increase the reproducibility of experiments and reusability of data. With this, the public get more value from their research funding, and companies can spend less money obtaining material tests data which is not shared.

The FAIR principles provide guidance to researchers on how to produce data that is findable, accessible, interoperable, and reusable (FAIR) for people and computers [1]. This should mean that data should be able to be found by both humans and machines, should reside in a place that users can reach, should be able to be integrated with other data so it can be included in pipelines and used with other applications, and should be well described/documented to make it possible for others to use. These guidelines are applied with a three-point framework, referring to the data, metadata, and infrastructure.

The implementation of these principles has benefits for multiple parties involved in experimental mechanics. The researchers themselves will have an easier time handling data, the cost of accessing and storing data will decrease for companies currently dealing with data management problems, and the research output from funded projects will see more value as it becomes more reusable.

There is already evidence of consideration of data management in the experimental mechanics community. The DIC good practice guide [2] provides a point of reference on how best to report DIC experimental results. Some experimental mechanics journals have started to implement checks against these reporting requirements as part of the peer review process. However, there are still some important elements to consider, such as creating metadata that is machine readable, infrastructure appropriate for storing and sharing terabytes of data, as well as the accessibility of closed file formats in commercial software.

In this work we develop a FAIR DIC data storage guide for experimental mechanics research undertaken at UKAEA. Our goal was to create a data storage structure in a way that would make it understandable and reusable. As part of this guide, we are developing a hierarchical meta-data schema that describes the key experimental parameters (as outlined in the DIC Good Practice Guide) in a machine-readable format. We also provide recommendations on appropriate cross platform and open file formats that will allow for processing of DIC data across software providers. We are also working on software tools that automate as much of the data storage principles as possible to reduce the burden on researchers implementing it.

A FAIR Data Storage Guide for DIC

DIC Data Overview. DIC data has multiple components that need to be considered, such as: the calibration images; information about the calibration target; static images pre/post testing; image deformation analyses for uncertainty quantification; convergence studies; noise floor analysis; the raw experimental data; and post-processed data. See Fig. 1 for a summary of the main types of data and associated data formats. We have developed a template JavaScript Object Notation file (JSON) that is generated and stored with each of these types of data allowing for automated machine retrieval of the associated meta-data. This meta-data JSON file includes all information required to reproduce the associated data such as the software used and links to bespoke code.

We tested our data storage guide with several experimentalists on recent campaigns using DIC at UKAEA have been tested on real data, such as a recent experiment testing two fusion components under high heat flux loads where we gathered 1.7TB of raw image data. Feedback was taken from these tests, and the guidelines updated to reflect the comments.

Meta-Data Schema & File Formats. Metadata references data that describes the other data, including how it is collected and stored. This metadata should have the details required to reproduce the experiment from scratch and should make note of any software used and needed to read any data produced along with the version. This should be captured by filling out a template data summary file. The file format should ideally be human and machine readable to adhere to the FAIR principles. In this case, JSON has been chosen to fulfill this purpose.

Naming conventions have been specified in the storage guide so that files can be easily recognised and understood by anyone viewing them. This is implemented by using standard name structures that include numbers and phrases to identify the campaigns, experiments, and tests.

Data should be stored in open and widely used formats where possible to maintain accessibility once software is deprecated. We suggest the following file types *.tiff/*.bmp for images, *.csv/*.hdf5 for post-processed data, and *.xml/*.yaml/*.json for meta-data and configuration specifications. In the case that using open format is not possible, the tools required to read the data should be stated in the metadata.

Key Principles of DIC Data Storage. We have developed a series of key principles for DIC data storage to help experimental mechanics researchers store their data in a reusable manner. The key principles of DIC data storage have been identified as follows:

1) Data should be backed up to a cloud storage system as soon as possible after collection, with data that does not yet conform to the guidance provided being placed in a scratch area and moved to a long term archive as soon as it is appropriately formatted.

2) A meta-data summary must be completed that captures the relevant data to the experiment, with enough information to reproduce the campaign and associated data.

3) Data should be stored in open-source formats where possible, with preference for human readable formats such as *.csv.

4) Data that must be stored in proprietary formats should have information about the software, including software version, detailed in the data summary file.

5) The data should be stored in a consistent folder structure. 6) Raw experimental data must be stored such that post processing is reproducible, so calibration data should be stored with the raw data, and static data traces for noise floor analysis must be stored.

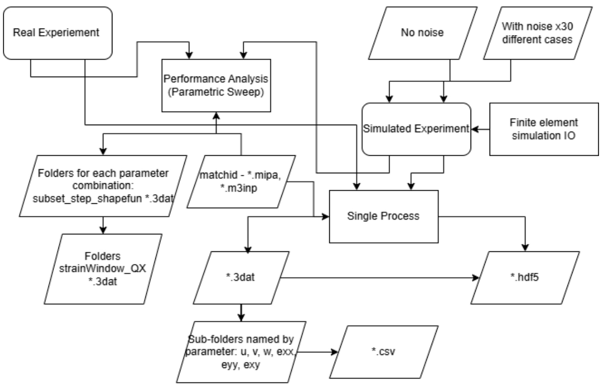

Figure 1: A flow chart for DIC data management

Conclusion & Future Work

In this work we have developed a FAIR data storage guide that we are using at UKAEA to ensure reusability of our DIC data. This should increase data lifespan and allow for more extensive use of data by the wider research community. For future work we are developing software to support the implementation of these guidelines. For instance, some metadata may automatically be produced during experiments, which could be extracted for use in the data summary.

Acknowledgements. We would like to acknowledge the support of UKRI through the Future Leaders Fellowship of Dr Lloyd Fletcher, grant number MR/Y015916/1. We would also like to acknowledge support from the EPSRC energy programme through grant number EP/W006839/1.

References

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[2] International Digital Image Correlation Society, Jones, E.M.C. and Iadicola, M.A. (Eds.) (2018). A Good Practices Guide for Digital Image Correlation. DOI: 10.32720/idics/gpg.ed1