Project Acronym: DyVirt

Horizon 2020

# Dynamic virtualisation: modelling performance of engineering structures

A MARIE SKŁODOWSKA-CURIE European Training Network (764547)

WP7
Deliverable D7.2:
Data Management Plan (version 1)
Last modified 08/02/2019

Due date of deliverable: 30.06.2018
Actual submission date: 08.02.2019

Start date of project: 01.02.2018
Duration: 48 months

Organisation name of lead contract for this deliverable:
The University of Sheffield, UK

## Table of Contents

## 1. EXECUTIVE SUMMARY

This data management plan has been formulated for the DyVirt project in order that it can contribute to the Horizon 2020 open science initiative. The purpose of the plan is to provide details of the management of data within the project. Specifically, it gives details of how research related data will be dealt with during the lifetime of the project. This includes details of how data is generated, collected, processed, stored, in addition to how it will be shared and made open to the public. Beyond this, details of how it will be curated and persevered are also included.

## 2. DATA SUMMARY

***Introduction to the DyVirt project***

The research carried out through this network will go beyond the now ubiquitous process of creating computer-based simulation models of structural dynamics. Obtaining a valuable virtual model is no longer a question of computing power, but now rests in the more difficult problem of developing trust in the model through the process of Verification and Validation (V&V).

In the DyVirt research programme, a 'virtualisation' will be a computer model validated to previously unreachable levels of trust in its capability. By the invention of new techniques in Verification and Validation (V&V) and the development of new understanding of how substructures and components assemble into full structures, it will be possible for the first time, to trust computer models for design cycle decision-making in the absence of more complete test data from prototypes or full-scale testing. This research will substantially extend the reach of current technologies that can only guarantee trust in models in more restricted circumstances.

For large engineering structures that operate in (often extreme) dynamic environments, such as wind turbines, aircraft, gas turbines and bridges, the process of virtualisation presents particularly difficult challenges. This is because the dynamic behaviour during operation needs to be fully captured by the computational model, but is highly sensitive to very small changes in (or disturbances to) the structure. For example, small differences in manufacturing tolerances, mechanical joints, or the operating environment (temperature, humidity), can all lead to apparently large changes in dynamic performance.

Although the virtualisation process has been attempted in some domains, it has never been properly applied to engineering applications operating in highly dynamic environments; therefore, our research aim is to create a dynamic virtualisation capability that for the first time will address this limitation. The methodology for DyVirt will be to focus on structural dynamics, and bring in new knowledge from the fields of sensor technology, data mining, decision theory, machine learning, optimisation, signal processing, statistics, aerodynamics, fracture mechanics, materials science and computational mechanics.

***Overview of data management***

Virtualisation requires extensive use of datasets for V&V procedures. Our research will include data collected from material, component, substructure and system tests, optimally collected at different levels of assembly using strategically placed sensors/actuators and/or optimal excitation features. Some of these datasets are subject to confidentiality, where they include data from industry. Records of experiments and outputs of simulations and models, including (but not limited to) graphical software package files, other analysis files, reports, analyses and communications including Emails will be collected. These will be produced by the Universities and, in some cases, also Industrial Partners. The data will be collected or

created by secure transfer of data, by agreement from the University and Industrial Partners.

Our data management plan is to collect and preserve, well documented, and well-organised data sets from all work packages in the network. Within WP1, a centralised, shared knowledge base will be created and linked to raw data, metrics, models, methods, requirements and model specifications. To achieve this, a data repository specifically dedicated to structural dynamics and virtualisation will be established using the USFD Library resources that already have the required e-infrastructures in place. It will be informed by developers and end-users and will exploit intranet design-sharing protocols in order to achieve a secure and reliable data management plan. Ontologies will be explored as a semantic web approach for encoding large, complex and heterogeneous domain knowledge. The idea will be to develop a novel V&V ontology using the Ontology Web Language and achieve knowledge storage based on WWW protocols and Resource Description Frameworks (RDF). Confidentiality/security of data will need to be assured with appropriate agreements (NDA, sharing accesses etc.). [1, 2, 3]

We note that for earthquake engineering, that ground motion records have been publically available for many years, we will use our website as a portal to give researchers access to these and other relevant sources of data. Where possible, and as appropriate, we will make this data available for sharing via OpenAIRE. We will also include “open code” to share algorithms and computer code developed as part of the project.

## 3. FAIR DATA

*3.1 Making data findable, including provisions for metadata:*
Metadata refers to “data about data”, i.e., it is the information that describes the data that is being published with sufficient context or instructions to be intelligible for other users. Metadata must allow a proper organization, search and access to the generated information and can be used to identify and locate the data via a web browser or web-based catalogue.

For reports and communications the documentation itself will be produced such that its purpose is explained within the document. For graphical software package and analysis files, the metadata will be included in all available metadata tags provided by the software package used or developed. In the context of data management, metadata will form a subset of data documentation that will explain the purpose, origin, description, time reference, creator, access conditions and terms of use of a

[1]Koo K-Y et al., 2011. SHM data management system using mySQL database and MATLAB interfaces. SHMII-5 2011 Cancun.
[2]Antoniadou, I et al. 2014. Towards and ontology for verification and validation in structural dynamics. ISMA 2014 Leuven.
[3]Kucuk, D and Arslan, Y 2014. Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles. RenewableEnergy 62, 484-489.

data collection. The metadata that would best describe the data depends on the nature of the data. For research data generated in DyVirt the metadata will be based on a generalised metadata schema, which includes elements such as:

- Title: free text
- Creator: Last name, first name
- Date:
- Contributor: It can provide information referred to the EU funding and to the DyVirt project itself; mainly, the terms "European Union (EU)" and "Horizon 2020", as well as the name of the action, acronym and the grant number
- Subject: Choice of keywords and classifications
- Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data.
- Format: Details of the file format
- Resource Type: data set, image, audio, etc.
- Identifier: DOI
- Access rights: closed access, embargoed access, restricted access, open access.

Additionally, a readme.txt file could be used as an established way of accounting for all the files and folders comprising the project and explaining how all the files that make up the data set relate to each other, what format they are in or whether particular files are intended to replace other files.

*3.2 Making data openly accessible:*

The H2020's open access policy pursues that the information generated by the projects participating in that programme is made publicly available. However, as stated in EC guidelines on Data Management in H2020, "As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex I, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access." In line with this, the DyVirt consortium will follow the strategy in Figure 1 to decide what information is made public according to aspects as potential conflicts against commercialisation, IPR protection of the knowledge generated (by patents or other forms of protection), meaning a risk for obtaining the project objectives/outcomes, etc. There may be restrictions on data sharing required whereby experimental data or datasets are produced by or in collaboration with the Industrial partners.

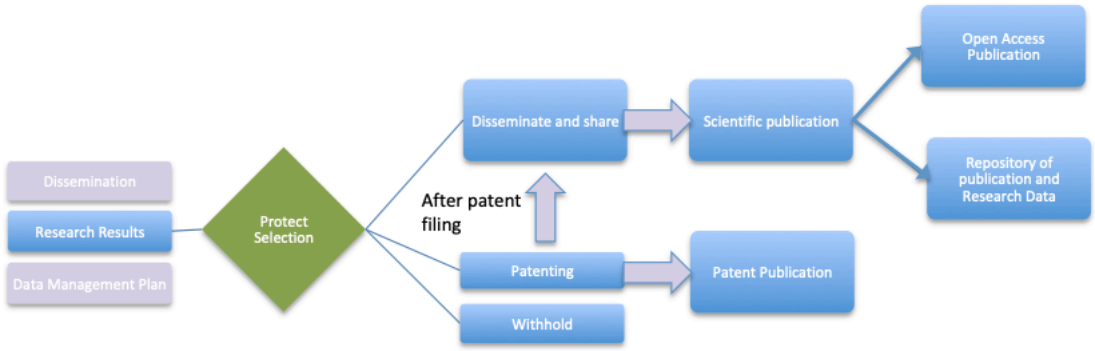


**Fig 1.** Process for determining which information is to be made public (from EC's document "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 – v1.0 – 11 December 2013")

Institutional repositories will be used by the project consortium to make the project results (i.e. publications and scientific data) publicly available and free of charge for any user. According to this, several options are considered/suggested by the EC in the frame of the Horizon 2020 programme to this aim:

**For depositing scientific publications:**

- Institutional repository of the research institutions (e.g. Open Access at The University of Sheffield uses the White Rose repository https://eprints.whiterose.ac.uk/)
- Subject-based/thematic repository
- Centralised repository (e.g. ZENODO)

**For depositing generated research data:**

- A research data repository which allows third parties to access, mine, exploit, reproduce and disseminate free of charge (e.g. ORDA at The University of Sheffield https://www.sheffield.ac.uk/library/rdm/orda)
- Centralised repository (e.g.ZENODO)

The academic institutions participating in DyVirt have available appropriate repositories. Data will be stored during the research according to the best practice of the host university of the researchers involved. Outputs will be stored in the form of analysis files or incorporated into reports, and shared with the Academic and Industrial partners by saving them on a secure project data repository. Access to primary data will be managed according to the host university of the researchers involved. Access to data shared under Open Data Policy will be managed according to the recommended best practice and capabilities of the repository of the host university producing the data. Data will be shared amongst the University and Industrial partners by means of the project Google site, and, for larger datasets, via the data repository of the host university producing the data. Data selected for long-term preservation and sharing will be stored on centrally provisioned University of Sheffield virtual servers and research storage infrastructure

(https://www.sheffield.ac.uk/cics/research) for at least ten years. Records of these data will be published in ORDA, a registry of research data produced at the University of Sheffield.

The ESRs will have the opportunity to publish results arising from their IRP as publications in the highest-ranking journals in this subject area (e.g. Journal of Sound and Vibration; Mechanical Systems and Signal Processing; Structural Control & Health Monitoring; Journal of Wind Energy; Renewable Energy; IEEE/ASME/ASCE Journals). The majority of these journals allow an open access modality and the author's post-print version can be deposited in a repository. This is in line with the Horizon 2020 requirements.

The DyVirt project website will act as a portal for (i) an ESR and supervisor intranet, that will enable sharing of information and project documentation, (ii) the data management platform for the project, and (iii) open access publications from the project, via a link to the OpenAIRE portal. Some of the academic institutions participating in DyVirt have available appropriate depositories for scientific publications which are linked to OpenAIRE. Institutions/organisations that don't have appropriate depositories or repositories that link to OpenAIRE will ensure that publications are deposited directly in OpenAIRE (https://www.openaire.eu):

**The University of Sheffield**
Type: Publication Repository
Website URL: https://www.sheffield.ac.uk/library/openaccess

**The University of Liverpool**
Type: Publication Repository
Website URL: https://livrepository.liverpool.ac.uk/

**Eidgenoessische Technische Hochschule Zurich**
Type: Publication Repository Website
URL: https://www.research-collection.ethz.ch/

**Gottfried Wilhelm Leibniz Universitaet Hannover**
Type: Publication Repository
Website URL: https://www.repo.uni-hannover.de

**Liege Universite**
Type: Publication Repository
Website URL: https://orbi.uliege.be/

**Panepistimio Thessalias**
Type: Publication repository
Website URL: http://ir.lib.uth.gr

**Akademia Gorniczo-Hutnicza IM. Stanislawa Staszica W Krakowie**
Type: Publication repository

Website URL: http://www.bg.agh.edu.pl/en/node/1376

Apart from these repositories, the DyVirt project will also use the centralised repository ZENODO to ensure the maximum dissemination of the information generated in the project (research publications and data), as this repository is the one mainly recommended by the EC's OpenAIRE initiative in order to unite all the research results arising from EC funded projects. ZENODO is an easy-to-use and innovative service that enables researchers, EU projects and research institutions to share and showcase multidisciplinary research results (data and publications) that are not part of existing institutional or subject-based repositories. ZENODO enables users to:

- Easily share the long tail of small data sets in a wide variety of formats, including text, spreadsheets, audio, video, and images across all fields of science
- Display and curate research results, get credited by making the research results citable, and integrate them into existing reporting lines to funding agencies like the European Commission
- Easily access and reuse shared research results
- Define the different licenses and access levels that will be provided

ZENODO also assigns a Digital Object Identifier (DOI) to all publicly available uploads, in order to make content easily and uniquely citable and this repository also makes use of the OAI- PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting) to facilitate the content search through the use of defined metadata. Metadata Schema according to OpenAIRE Guidelines. The short- and long-term storage of the research data in ZONODO is secure and it uses digital preservation strategies to storage multiple online replicas and to back up the files (Data files and metadata are backed up on a nightly basis). This therefore fulfils the requirements of the EC for data sharing, archiving and preservation of the data generated in DyVirt.

*3.3 Making data interoperable:*

Records of datasets will be published in ORDA, the University of Sheffield's registry of research data produced at the University, which will issue Data Cite DOIs for registered datasets and promote discovery

*3.4 Increase data re-use (through clarifying licenses):*

Apart from earthquake ground motion records, mentioned above, there has until now been only limited publically available data on dynamics of infrastructure. We anticipate, that the DyVirt data portal will radically improve the digital science activity in this field, specifically by encouraging more open collaborations amongst researchers, giving SMEs and other industrial organisations wider access to research data, and allowing policy makers and governmental bodies to have more access to well documented and archived data on this important area of policy. The University of Sheffield's Good Research and Innovation Practice (GRIP) Policy follows UKRI

principles for data sharing (https://www.ukri.org/funding/information-for-award-holders/data-policy/).

All DyVirt team members will ensure that a data access statement is put in the Acknowledgements section of all their research publications.

Examples of data access statements:
"All data created during this research are openly available from the University of X data archive at http://XXXXXXX"
"All data supporting this work are provided as supplementary information accompanying this paper."
"All data are provided in full in the results section of this paper."
"This publication is supported by multiple datasets, which are openly available at locations cited in the references"
"No new data were created during this study"
"This study was a re-analysis of existing data that are publicly available from http://XXXXXXX"
Commercial restrictions
"Supporting data will be available from the University of X data archive at http://XXXX after a Y month embargo from the data of publication to allow for commercialisation of research findings"
"Due to confidentiality agreements with research collaborators, supporting data can only be made available to researchers subject to a non-disclosure agreement. Details of the data and how to request access are available at the University of X data archive: http://XXXXX."

It is also encouraged to use publications that allow supplementary files to be uploaded with the publication in order to more directly link the data files being published to the publication itself.

Data should be retained, shared and / or preserved where it may be used to reproduce conclusions from publications arising from the research program, where it may provide additional information, or where it may be considered useful. The long-term preservation plan for the dataset is for it to be preserved as long as it is considered useful and economic within the data policy of the host institution that originally produced the data.

## 4. ALLOCATION OF RESOURCES

The resources required to deliver the plan are available within the budget of the DyVirt project in conjunction with the facilities provided by the project Partners.
The University of Sheffield research data storage facility allocates 10TB storage free to research groups during the lifetime of a project. If a larger quota is required then this will involve charges. Long-term archiving of data may involve charges also. ORDA (https://www.sheffield.ac.uk/library/rdm/orda), the University of Sheffield research data repository for managing and sharing research data is free to use.

Contacts for project data:
Project Manager – Dr Victoria Hand (University of Sheffield)
Project Administrator – Grace Stokes (University of Sheffield)

**5. DATA SECURITY**

Data and definitive project documentation will be stored on centrally provisioned University of Sheffield virtual servers and research data storage infrastructure throughout the lifetime of the project. Both Windows and Linux Virtual Servers with up to 10TB of storage are made available to research projects. Access control is by authorised University computer account username and password. Off-site access is facilitated by secure VPN connection authenticated by University username and remote password. By default, two copies of data are kept across two physical plant rooms, with a 28-day snapshot made of data and backed up securely offsite at least daily. This service is maintained by the University's Corporate Information and Computing Services. Google Drive is used for more flexible collaborative working but only where non personal-sensitive information is involved. Where Google Drive is used, copies of complete and definitive documents will be transferred to the main project repository on the University research storage infrastructure.

**6. ETHICAL ASPECTS**

Copyright and IPR issues shall be managed as noted in the Project Collaboration Agreement. The Ethics committee will be asked to oversee the data management process to ensure that data are collected and preserved in accordance with all ethical considerations, including any potential confidentiality and copyright issues.

**7. OTHER**

**Data Management Policy & Procedures; Data Security Policies and Data Sharing Policies.**

**University of Sheffield:**
**Data Management Policy & Procedures:** https://www.sheffield.ac.uk/govern/data-protection
**Data Security Policies:** https://www.sheffield.ac.uk/cics/policies/infosec
**Data Sharing Policy:** https://www.sheffield.ac.uk/library/rdm/expectations

**University of Hannover:**
https://www.uni-hannover.de/en/datenschutzerklaerung/

**University of Liverpool:**
**Data Management Policy &**
**Procedures:** https://www.liverpool.ac.uk/library/research-data-management/ **AND** https://www.liverpool.ac.uk/media/livacuk/computingservices/research-data-management/researchdatamanagementpolicy.pdf

**Data Security Policies:** https://www.liverpool.ac.uk/csd/security/information-security/ **AND**
http://www.liverpoolac.uk/media/livacuk/computingservices/regulations/informationsecuritypolicy.pdf
**Data Sharing Policy:**
https://www.liverpool.ac.uk/media/livacuk/computingservices/research-data-management/researchdatamanagementpolicy.pdf

**Eidgenoessische Technische Hochschule Zurich**
**Data Management Policy & Procedures:** https://www.library.ethz.ch/en/ms/Digital-Curation-at-ETH-Zurich/Research-data/Research-data-management
**Data Security Policies:** https://www.ethz.ch/en/footer/data-protection.html
**Data Sharing Policy:** https://www.library.ethz.ch/en/ms/Digital-Curation-at-ETH-Zurich/Research-data/Publishing-research-data