Ref. Ares(2019)353191 - 22/01/2019

RADICLE

Real-Time Dynamic Control System for Laser Welding

Grant Agreement Number: 636932		
Project Title: Real-time dynamic control system for laser welding		
Project Acronym: RADICLE	**Funding Scheme:** Collaborative Project	
Date: December 2018	**Project Website Address:** www.radiclelaser.com	
EC Project Officer: Christoph Helmrath	**Email:** christoph.helmrath@ec.europa.eu	
Deliverable Number: D7.20	**Deliverable Name:** Final Data Management Plan	
Work Package Number: 7		
Date of Delivery: M45	Actual	M47
Status	Draft ☐ Final ☒	
Nature	Prototype ☐ Report ☒ Specification ☐ Tool ☐ Other ☐	
Distribution Type	Public ☒ Restricted ☐ Consortium ☐	
Authoring Partner: EWF		
Contact Person: André Cereja		
Email: afcereja@ewf.be	**Phone:** (+351) 215 815 204	**Fax:** N.A.
Abstract (for dissemination)	N.A.	
Keywords	N.A.	
Name of the Scientific Representative of the Project's Co-ordinator, Title and Organisation:	**Name:** Nicholas Blundell **Tel:** +44 247 670 1803 **E-mail:** nicholas.blundell@the-mtc.org	

Content

1. Summary

The Horizon 2020 Work Programme introduces the Data Management Plan (DMP), in which a project consortium specifies the research data that will be open access and how it will be handled during the project.

Open Access refers to the practice of giving online access to all relevant information that is free of charge to the end-user. In this way, data becomes re-usable, and the benefit of public investment in the research will be improved.

A DMP is a report created to provide an analysis of the main elements of the data management policy that is used by the RADICLE Consortium with regard to the project public research data.

The DMP spans across the complete research data life cycle. It contains information related to the types of public research data that were generated during the project, how the research data is preserved and shared for verification or reuse.

The DMP describes in detail what public data the project has generated, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

The RADICLE project beneficiaries have deposited public data in a research data repository and taken measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any user.

The DMP covers several topics established by the EC and it describes the Data Management life cycle for all data sets that will be collected, processed or generated by this research project.

The DMP was a live document that must *reflected the current status of reflection within the consortium about the data that was produced*. This report is its last iteration and is also an evolution of D7.18 and D7.19.

2. Research data

RADICLE's DMP aims to provide an analysis of certain elements of the data management policy that will be used by the Consortium with regard to the project research data.

The DMP covers the complete research data life cycle. It describes the selected types of public research data that are collected during the project, the used data standards, the preservation of research data and what parts of the datasets are publically shared for verification or reuse. It also reflects the current state of the Consortium agreements on data management and must be consistent with exploitation and IPR requirements.

The DMP deals with how the project participants managed the research data generated and/or collected during the project. As agreed by the RADICLE partners, the type of data that was generated related specifically to:

- Characterisation of welding joints;
- Sensor outputs and how they relate to detection of defects;
- Data collection and manipulation;
- Specific knowledge relating to the particular end-user samples.

Data created during the project development was held on secure servers either at local or cloud level (or both) depending on partner preference. Access has been provided to non-confidential results through the gold open access procedures (including online repositories).

All aspects of the data are covered by the Consortium Agreement. The appropriate structure of the consortium to support exploitation is addressed in section 2.3.2. The consortium is working work as a part of the Pilot on Open Research Data in Horizon 2020 on a voluntary basis.

This report is only focused on data that the consortium has agreed to make publicly available.

2.1. Data Identification

Data Identification consists in a Dataset reference and a Dataset name.

2.2. Dataset description

The Dataset description may include the following items:

- Data Description
- Type (Collected/Processed/Generated)
- Origin (if Collected/Processed)
- Format
- Nature
- Scale
- Useful to Whom
- Does it underpin a scientific publication
- Information on existing similar data
- Possibility for integration and reuse
- Storage and Backup

2.3. Data Sharing

The Data Sharing description may include the following items:

- Steps to Protect Privacy
- Security
- Confidentiality
- IPR
- How the Data will be Shared
- Access Procedures
- Who controls It
- Embargo Periods
- Outlines of Technical Dissemination
- Software and Tools to Enable Re-Use
- Widely Open Access or Restricted to Specific Groups
- Repository Where Data will be Stored
- Type of Repository (institutional, standard repository for the discipline, etc.)

2.4. Archiving and Preservation (including storage and backup)

In addition to the internal project database for IPR-restricted data, the public dataset is stored in ZENODO, which is the open access repository of the Open Access Infrastructure for Research in Europe, OpenAIRE.

Items deposited in ZENODO will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for the at least next 20 years. Data files and metadata are backed up on a nightly basis, as well as replicated in multiple copies in the online system.

ZENODO was built and developed by researchers, to ensure that everyone can join in Open Science.

The OpenAIRE project, in the vanguard of the open access and open data movements in Europe was commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and ZENODO was launched in May 2013.

Figure 1 – Zenodo Logo

3. RADICLE Public Dataset

Based on the type of data generated during the development of the technical work of the RADICLE project, the consortium has identified one Dataset that has been shared with other researchers, with an Open Access policy. Other types of data that are produced during the project, for instance relating to the end-user samples, are not subject to release to the public, due to IPR restrictions.

Table 1 identifies the dataset to be made public within the RADICLE project.

Table 1 – RADICLE Datasets

Dataset	Main responsible for data	Related WP(s)
S355 test data	TWI, LOE	WP2, WP3

3.1. Dataset description - S355 test data

The description of Dataset number 2, S355 test data, is presented next, on Table 2:

Table 2 - Dataset description - S355 test data

Dataset Characterisation	Description
Dataset reference	RADICLE_S355_test_data
Dataset name	RADICLE_S355_test_data
Data description	The RADICLE_S355_test_data dataset consists on the raw and treated data related to testing the laser welding process on S355 steel (a high-strength low-alloy structural grade), in a butt-weld configuration. S355 is a material commonly applied throughout the 'heavy industry' sectors, including transport (road, rail and marine), yellow goods (earth-moving and construction machinery), civil engineering and energy sectors. Hence, the H2020 RADICLE project consortium is providing open access to the data collected during the test trials, to allow for further analysis and re-use of this information. The files included in the dataset are: • Digital photographs.rar • Digital radiographs.rar • Jan 18 S355 TWI cam analysis.rar • LDD.zip • LOE Camera.zip • Permanova Camera.zip

Dataset Characterisation	Description
	• RADICLE readme 6mm S355.xlsx • Rough photos.rar • S355 LOE Diode Data.zip
Type	Generated
Origin	This dataset contains both Collected and Processed data
Format	This dataset contains various types of files. Larger files are uploaded in .rar archives. Inside these, raw data is stored in .csv, .jpg, .tdms, .avi, .ldsd and .ldrd file formats.
Nature	Academic/Industrial research data
Scale	NA
Useful to whom	European Industry and Academia
Does it underpin a scientific publication?	No
Information on existing similar data	NA
Possibility for integration and reuse	Yes, data is available for integration and reuse. Data is hosted online and has been publicly available made through the ZENODO platform.
Storage and Backup	These matters are covered by the ZENODO repository, which ensures sustainable archiving of the research data. All data is available via the link: https://zenodo.org/record/2277818 Or via the DOI: 10.5281/zenodo.2277818

3.2. Dataset Sharing - S355 test data

The characterisation of the sharing protocol for Dataset number 2, S355 test data, is presented next, on Table 2:

Table 3 - Dataset Sharing - S355 test data

Dataset Sharing	Description
Steps to Protect Privacy	The Zenodo collaboration does not track, collect or retain personal information from users of Zenodo, except as otherwise provided herein. In order to enhance Zenodo and monitor traffic, non-personal information such as IP addresses and cookies may be tracked and retained, as well as log files shared in aggregation with other community services (in particular OpenAIREplus partners). User provided information, like corrections of metadata or paper claims, will be integrated into the database without displaying its source and may be shared with other services. Zenodo will take all reasonable measures to protect the privacy of its users and to resist service interruptions, intentional attacks, or other events that may compromise the security of the Zenodo website.
Security	ZENODO takes security very serious: • CERN Data Centre: Data centres are located on CERN premises and all physical access is restricted to a limited number of staff with appropriate training and who have been granted access in line with their professional duties (e.g. Zendo staff do not have physical access to the CERN Data Centre) . • Servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to servers are restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with latest security patches via our automatic configuration management system Puppet. • Network: CERN Security Team runs both host and network based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks in order to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.

Dataset Sharing	Description

	• Data: Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.
	• Application: ZENODO is employing a suite of techniques to protect the user's session from being stolen by an attacker when logged in and run vulnerability scans against the application.
	• Staff: CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that
	  ○ staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties.
	  ○ access to user data must always be consistent with the professional duties and only permitted for resolution of problems, detection of security issues, monitoring of resources and similar.
	  ○ staff are liable for damage resulting from any infringement and can have access withdrawn and/or be subject to disciplinary or legal proceedings depending on seriousness of the infringement.

Confidentiality	The RADICLE_S355_test_data dataset is open-access.
IPR	No IPR is associated to the RADICLE_S355_test_data dataset.
How the Data will be Shared	RADICLE_S355_test_data is available via the DOI (10.5281/zenodo.2277818) or via the https://zenodo.org/record/2277818 weblink.
Access Procedures	NA
Who controls It	NA
Embargo Periods	NA
Outlines of Technical Dissemination	NA
Software and Tools to Enable Re-Use	All files are publicly available via the open access Zenodo platform, and are not encrypted.

Dataset Sharing	Description
Widely Open Access or Restricted to Specific Groups	Widely Open Access
Repository Where Data will be Stored	https://zenodo.org/record/2277818
Type of Repository (institutional, standard repository for the discipline, etc.)	Institutional

3.3. ZENODO data

The dataset RADICLE_S355_test_data is available to the public via the address: https://zenodo.org/record/2277818 or via the DOI 10.5281/zenodo.2277818

Next is shown a screen capture of the data uploaded to the platform.

Figure 2 – Contents of the RADICLE_S355_test_data dataset available in the ZENODO platform

4. Other activities

Besides the effort to identify Datasets which could be made available to the public, the RADICLE partners (particularly LOE) also made extra efforts to share the generated data to an external audience.

As part of the funding of the RADICLE project it is intended that weld signal data generated should become publicly available for wider learning and understanding. NTU used internal funding to enable three of their staff to work with LOE on behalf of RADICLE to investigate the limitations of the data formats and how to make it accessible for wider sharing.

A report on the above is included in Annex I of this document. This Annex outlines a data dissemination and analysis exercise carried out by LOE working with Nottingham Trent University (NTU).

5. Conclusions

This document is the last iteration of RADICLE's Data Management Plan (DMP). The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regards to the project research data.

The DMP is not a fixed document: on the contrary, it has and will evolve during the lifespan of the project. This last version of the DMP includes an overview of the datasets produced by the project, and the specific conditions that are attached to them.

The final version of the DMP describes the practical data management procedures implemented by the RADICLE project, with the goal of complying with the requirements set out by RADICLE's participation in the Pilot on Open Research Data launched by the European Commission along with the H2020 programme.

Laser Optical Engineering Ltd
pioneering laser technology

Annex I

RADICLE data Dissemination and testing exercise with NTU

Prepared for: RADICLE consortium

Date: 16th November 2018

Job: P0118

Laser Optical Engineering Ltd.
Building 72a
The Air Cargo Centre
Argosy Road
East Midlands Airport
DE74 2SA
United Kingdom

Tel: +44 (0) 1332 814612
Fax: +44 (0) 1332 850855
e-mail: enquiries@laseroptical.co.uk
web: www.laseroptical.co.uk

Laser Optical Engineering Ltd
pioneering laser technology

Version control

Version	Date	Compiled by	Reviewed by
c	19/11/18	D. Lloyd	F. Warner

Contents

1 Annex I – Introduction

This report outlines a data dissemination and analysis exercise carried out by LOE working with Nottingham Trent University (NTU). As part of the funding of the RADICLE project it is intended that weld signal data generated should become publicly available for wider learning and understanding. NTU used internal funding to enable 3 of their staff to work with LOE on behalf of RADICLE to investigate the limitations of the data formats and how to make it accessible for wider sharing.

2 Annex I – Work outline

LOE have worked with Nottingham Trent University (NTU) to explore how the open access data set can be analysed by third parties to the RADICLE project. At EWF's suggestion, TWI undertook a set of welding experiments on steel S355 as this a conventional material with little proprietary knowledge involved, but of much wider interest than some of the specialised aerospace grade materials welded for end users.

Lecturer Dr Georgina Cosma and Research Associates Tolulope Oluwafemi and Sadegh Mousaabadi Salesi were given a set of photo diode, acoustic, LDD and camera data recorded during the welding. Weld quality data was provided in the form of x-rays of the welded part and a 'traffic light' status of the weld.

Due to the limited time available it was decided that rather than try to make a feedback process, a more limited quality prediction tool would be considered. This allows the computational algorithm more time to analyse the data it is fed so that less optimisation is required. As a result, the only outcome is a quality prediction, rather than a specific defect location. Each of the welds had one of three statuses: good, questionable or bad.

The first observation was that the data set was very limited in size making it difficult to sufficiently train a machine learning algorithm. A number of the data recordings were incomplete or had other issues, meaning that just 8 welds were suitable for use. None of those considered useable had the questionable status. As a result, classification and prediction of the data set can only produce 2 outcomes a positive or negative status. This means that very high statistical predictions are achieved, a success which would not be replicated with a much larger and more representative data set.

In order to be able to make any predictions about a data set it is first necessary to 'train' a model and to prepare the data. Preparing the data includes stages such as defining the start and end point to remove any 'non-signal' data and aligning the different signal types. LOE provided x-ray data produced by TWI in image form and in normalised numerical form which required development of an image processing tool.

Training the model then requires the data set to be classified. A number of these methods are available and are detailed in the NTU reports attached. Clustering the datasets breaks down the raw data into computer defined significant features. The number of these clusters varies and the derivation of them is statistically derived. Using the best approach tested, it was found that 3 clusters showed the best quality indicator for the available data. If the majority of data points are in Cluster 1, then the quality is 'bad'. If Clusters 2 and 3 are bigger, then the quality is 'good'. Due to the limited data set it has not been determined if there is more meaning that could be derived from clusters 2 and 3.

Once the clustering has been performed, the true machine learning can be applied. Again a number of algorithms can be applied. These can be tested by removing the quality indicator and processing one of the data sets. The algorithm then compares it to the datasets for which it has a quality status and looks for similarities. By matching it to the closest data set, a predicted quality status can be generated. In testing, a number of algorithms showed very similar and very good performance. This is misleading as the data set is very small and the range of statuses tested binary. The prediction made is therefore binary; anything other than a near 100% prediction would be closer to guess work than a statistical indictor.

3 Annex I – Future work

To aid wider use of the data set annotations will need to be made to explain the data. At present the start and end points of the weld are observed manually: for wider use there needs to be an explanation of where these occur and how to identify them. This is complicated by the length of path travelled while the beam is on being different to the heat affected area: the keyhole was a width which extends beyond the nominal point of incidence of the beam.

A major challenge that still remains is to align a higher resolution status with the signal data. This means converting the resolution of an x-ray, typically 600 pixels (for a 70mm weld), to the same resolution as the video data (3000) and photodiode data (180,000) points. To achieve this, interpolation is required. It also requires interpretation of the x-ray to define acceptable pores or clusters and bad areas for each defect category. While x-rays showing the presence of low density areas was available, the cause of these was not available during this study.

For dissemination purposes, the method of sharing the data remains a challenge. A typical set of signal data generated by the LOE system equates to around 100mB per weld. The Permanova seam tracking video potentially adds several hundred mB. In addition to this the quality data, X-ray, CT scans and surface profiles add significantly more data, especially in their raw form. Sharing this data on physical memory or the server time and space to make it available on a ‘sharepoint’ ftp site are costs which have not been factored in to this aim.

This study has simply attempted to find a way to predict global quality of a weld. To turn this into a system which can predict the cause of a change quality and make a process parameter change to correct it is a significantly harder challenge. This work suggests that there is useful data in all the signals generated which allows quality predictions to be made. This provides evidence that the signals gathered could, in theory, allow real time control and analysis to be achieved.

LOE would like to thank VTT for help critiquing NTUs work and for helping provide feedback to direct the study.