Clean Sky 2 Joint Undertaking (CS2JU)

Data Management Plan

Call Reference No: H2020-CS2-CPF06-2017-01

Dissemination Level: PUBLIC

October 2017

Information notice

Introduction

This Data Management Plan (DMP) is based on the H2020 Data Management Template .The template is aimed at assisting the applicants selected by the CS2JU in the preparation of the DMP of the selected proposal, where applicable.

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

FAIR data management

In general terms, research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This DMP is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept.

More information about FAIR:

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

Structure of the DMP

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, this DMP has clear version number and include a timetable for updates. As a minimum, the DMP will be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update will be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 4 contains a summary of these elements in bullet form. This DMP itself will be updated as the policy evolves.

<u>When is the DMP applicable to a project?</u>

A DMP is mandatory when at the stage of the application the single entity or applying consortium requested in the application template the "Open access to research data" to be applicable to its proposal and did not select the "opt-out" option in the application template or requested that during the grant preparation phase. In such a case of "opt-in", Article 29.3 of the CS2JU model grant agreement will be added to the grant agreement and the DMP will become mandatory and shall be delivered as a deliverable to the CS2JU within 6 months from the start of the action.

In the case of proposals for which an "opt-out" was chosen at application stage or requested and agreed by the CS2JU at grant preparation stage, please note that the DMP may still be applicable and used by the single beneficiary or consortium on a voluntary basis but its costs will not be eligible under the grant.

<u>Specificities of CS2JU projects</u>

Please note that considering the complementary nature of the topics of the CS2JU which are linked to an ITD/IADP demonstrator and the access rights regime and exchange of data that may be applicable during project implementation between the beneficiary and the Members' topic manager to implement the project, the coordinator must consult the responsible topic manager during the grant preparation phase and related technical sessions to determine the scope and perimeter of a possible open access regime and identify in written the data that will be generated out of the action or exchanged in the course of the action implementation which may be released and made available on a public repository under the open access regime. Please note in particular the provision of Section 8.2.2 of the Implementation Agreement (v2 March 2017) applicable to the complementary topics launched under the CS2JU calls for proposals.

<u>Legal disclaimer</u>

Please note that this "DMP" is meant to provide advice and support the beneficiaries of a CS2JU project in the case the project is covered by the "open access to research data" under Article 29.3 of the grant agreement. The CS2JU bears no responsibility for its use and implementation *vs* the grant beneficiaries, the other participants in the project and any other third party.

Project[1] Number: [785349]

ITD/IADP/TA:

Project Acronym: [PROTEUS]

Project title: [Performance & Operability of Turbofan Engines Under Sub-Idle]

DATA MANAGEMENT PLAN

Update of the DMP record

Revision	Date	Description	Reviewer
1.0	04/08/2018	Initial outline	
2.0	**12/06/2019**	**Interim Review**	
3.0	**01/02/2020**	**Interim Review**	
4.0	**01/02/2021**	**Final Review**	

Prepared:	C. Mourouzidis, L. Ferrer-Vidal, H. Azamar Aguirre (CU)
Reviewed:	V. Pachidis (CU)

---

[1] The term 'project' used in this template equates to an 'action' in certain other Horizon 2020 documentation

Executive Summary

The data management plan (DMP) outlines how data is to be generated and handled, both during and after the project is completed. The DMP plan considers the many aspects of data management, generation, preservation, and analysis. This ensures that data are well-managed in the present, and prepared for preservation in the future. The objective of the DMP is to establish a detailed structure and framework on how the research data used during the project timeframe is originated, handled, maintained and preserved to guarantee findable, accessible, interoperable and re-usable (FAIR) data principles. The nature of the dataset generated, handled and preserved as part of the PROTEUS project is numerical, experimental and a set of software tool codes.

1. DATA MANAGEMENT AND RESPONSABILITY

1.1. DMP Internal Consortium Policy

The data management activities are going to be managed by a specific person per consortium partner, while the overall plan will be managed by Cranfield University. The PROTEUS project coordinator organization, Cranfield University, is responsible for adopting and procuring the strategy for data maintenance, selecting the data repository, uploading the data and keeping up-to-date the information. Given these responsibilities, Cranfield University is the responsible entity for the relation and communication with the topic manager.

The list of the organizations and individuals responsible is presented in Table 1.

Table 1 – Responsible Organizations and Individuals

Organization Name	Organization Role	Responsible Contact
Cranfield University	Coordinator/ Overall Contact	Prof Vassilios Pachidis
University of Cambridge	Beneficiary	Prof Epaminondas Mastorakos
Karlsruhe Institute of Technology	Beneficiary	Prof Rainer Koch
Rolls-Royce PLC	Topic Manager	Mr Richard Tunstall

1.2. DATA MANAGEMENT Responsible

The project data contact responsible and in charge of data procurement, upload, maintenance and update is given in Table 2.

Table 2 –Project Data Contact Responsible

Project Data Contact (PDC)	**Prof Vassilios Pachidis**
PDC Affiliation	**Cranfield University**
PDC mail	**v.pachidis@cranfield.ac.uk**
PDC telephone number	**+44 (0) 1234 75 4663**

1.3. DATA nature, link with previous data and potential users

The nature of the PROTEUS project dataset generated, handled and preserved is numerical, experimental and software toolsets. PROTEUS project involves the generation of numerical data from three different simulation sources: 3-D CFD (Computational Fluid Dynamics), 1-D mean-line and 0-D whole-engine performance cycle. Experimental data is provided by the topic manager for validation purposes collected from the agreed engine demonstrator program test-rig. The several in-

house methods developed by the Consortium partners within the PROTEUS project will be capitalised to predict the engine performance and operability during idle and sub-idle conditions leading to software toolset codes to be delivered to the topic manager.

1.4. Data Summary

The data generated within the PROTEUS project will stem from 3-D CFD simulation results of engine component models, 0-D whole-engine performance simulation models, as well as in-house software tools developed by the Consortium partners. The high-fidelity data generated (3-D CFD) is to be reduced into component characteristics for integration into whole-engine performance tools. Data formats will include for example numerical datasets, computer codes, text data, NPSS outputs and technical figures, as summarized in Table 3.

Table 3 - Data formats and file extensions to be used within the project

Data format	Format extension
C, C++ files	.c, .h
Comma separated value files	.csv
Text files	.txt
Simple data files	.dat
Portable document format files	.pdf
Tagged image files	.tif
Portable network graphic files	.png
FORTRAN 90 files	.f90
PYTHON files	.py
MATLAB files	.m
NPSS files	.int, .mdl, .map

In order to ensure portability and accessibility of the numerical data generated, comma separated value files (.csv) will be the primary format for data storage and transfer between partners. Final datasets are estimated to occupy broadly 1 Tb, although, the expected data size will be frequently re-defined during the project.

2. FAIR data

To manage the numerical and experimental research data, as well as the computational toolsets generated and/or collected during the project, the partners will follow the fundamental IPR rules defined in the H2020 Grant Agreement.

Following guidelines on data management in Horizon 2020, PROTEUS partners will ensure that the research data from the project is Findable, Accessible, Interoperable and Re-usable (FAIR).

2.1. Making data findable, including provisions for metadata

On the basis of a FAIR practice, the management of the numerical and experimental data, and the computational toolsets will possess a file name and nomenclature according to a standard identification mechanism explained and detailed in metadata. The research data as well as the metadata file names will have a series of subject identifiers separated between them by an underscore (_).

A metadata *readme* .txt file will accompany the data repositories of the project, including a list describing the contents of each directory and standard file nomenclature. The metadata will be clearly identified by the directory name and the label *readme* at the end, to permit the user to identify this file as metadata.

The generated data will be archived in general by engine type and component, including subdirectories for each case. The file names contained in every subdirectory will be consistent with their location address as a way of finding and tracking the files and containing any possible misplacement. Hence, the file name standard convention to be used is described by the engine type, component and analysis case in the format: *enginetype_component_analysiscase* .file extension. The engine type is given by the engine model. The analysis case is composed of the engine condition, mass flow rate and rotational speed.

At the end of the project, all available public domain data will be uploaded by Cranfield University onto CORD within 3 (three) months of the project end. CORD is an institutionally-approved secure repository, which retains data for at least 10 years. CORD uses the *figshare* platform, which is ISO27001 certified and assigns a *DOI* to each item, to ensure that datasets are findable. Metadata such as keywords will be added to optimize possibilities for re-use. All data, prior uploading, will be checked for its validity and quality by each member of the consortium responsible for it and the Topic Manager.

2.2. Making data openly accessible

The work undertaken within the PROTEUS project in general is sensitive and of commercial value to the Topic Manager, relating to recent and intended future products. Datasets with background Intellectual Property (IP) supplied by the Topic Manager to the PROTEUS partners, or data that may include third-party data licence from other companies, will not be made openly available. For instance, IP datasets might include geometry, rig and engine test data, performance model information and software. Similarly, results obtained from the PROTEUS project that contain background IP, such as geometry information, will not be made openly available.

Any other datasets that do not contain the background IP supplied by the Topic Manager or without a third-party data licence from other companies will be made openly available. In this context, any PROTEUS project output data will be post-processed in such a way as to remove any Topic Manager's background IP to make it publicly accessible.

Making the data openly accessible requires formal clearance and consent from the Topic Manager. All peer-reviewed scientific publications resulting from the project will be Open Access where possible with previous clearance from the Topic Manager.

The data output and results generated from the PROTEUS project will be obtained from commercial software such as ANSYS CFX, in-house developed toolsets and methods, and modelling-system tool (NPSS) supplied by the topic manager. Main output data will be post-processed in such a way that no software interface is required to access it; however, if any raw data is required to be stored, documentation will be included for the case of the in-house software and NPSS. For any other raw data derived from commercial programmes, the documentation to access is openly available.

As a general principle, the software source codes from the different partners will not be made available. Partners will produce publications describing the new methods applied in the software rather than publishing the source code nor the software directly. It is preferable from the Topic Manger not to publish any PROTEUS-derived software directly; however, if any foreground software from PROTEUS is to be made public then it must be ensured that this software does not contain any background Topic Manager's IP, including Topic Manager's IP software.

A special case is related to the NPSS thermodynamic modelling-system supplied by the Topic manager to Cranfield University as a background IP. Neither the software itself nor its source code should be made public, due to the Topic Manager's background IP, and because the NPSS code carries a US Export-Control rating which prevents publication.

Secured data will be stored in the Exostar ForumPass 6 information repository for the specific purposes and duration of the project. ForumPass 6 is a trusted workspace built upon the Microsoft Office SharePoint Server 2013 platform and is managed by the Topic Manager. The Exostar ForumPass 6 repository will allow access only to project participants. Metadata will be allocated in the lowest level of subdirectories to describe the datasets.

The topic manager is responsible for managing and granting access to the Exostar ForumPass 6 platform. Consortium partners will have access to the storage though a security system based on username and password. Since the Topic Manager's background IP protection has been agreed within the partners no data access committee is foreseen. Nevertheless, in the case of an external request for data access, Cranfield University as leading partner will select an internal data access committee.

2.3. Making data interoperable

All research-generated data is exchangeable and can be re-used between the Consortium partners. The exchangeable data shared among the project participants is mainly of numerical nature and is transferred as a tabular data package in text or table-processor files as CSV (comma separated values) files. Due to the secure data storage and exchange platform managed by the Topic Manager, Exostar ForumPass 6, all experimental and numerical information can be formally shared and be accessible to all Consortium partners.

The vocabulary and nomenclature used within the exchangeable data follows the international standards for engine station numbering and nomenclature to avoid ambiguous data transfer and misinterpretation. The vocabulary reference for this interoperable data is according to the SAE (1974) Gas Turbine Engine Performance Station Identification and Nomenclature Aerospace Recommended

Practice (ARP) 775A. The nomenclature used in the software toolset and method codes will follow the SAE Aerospace Standards (AS) for digital computer programs to ensure interchangeability between the partners and the Topic Manager. For the case of gas turbine engine performance steady-state and transient simulation codes, the nomenclature standard used is according to the latest revised version SAE AS681K. Any additional necessary parameters related to time-based simulations will follow the SAE ARP1257 to cover for transient performance analyses.

2.4. Increase data re-use (through clarifying licences)

To permit the widest reuse possible, the relevant experimental and numerical data derived from the PROTEUS project without any Topic Manager's background IP and without a third-party data licence will be clear and accessible through a data usage licence. Any re-use activity will have to be cleared and with previous consent by the Topic Manager in advance and may be used by external entities for comparison and validation.

Sensitive data containing the Topic Manager's background IP or holding a third-party data licence will remain restricted and will not be licenced for re-use or disseminated by the PROTEUS partners unless written permission is obtained from the Topic Manager in each case. In the case of every consortium partner, their own source code, in-house software, and technical data involving the Topic Manager's background IP will not be licenced.

To this end, the datasets are expected to be made publicly available for re-use immediately after the full completion of the project. The Topic Manager encourages the partners to apply for patents for any novel technologies developed under the PROTEUS project. In this context, the Topic Manager prefers the partners to delay publication of any materials necessary to allow time for all relevant inventions to be protected.

To guarantee the re-use of data, a clear history of origin, methodology, data workflows and references will be detailed in the final reports. The utilization of metadata, standard files, vocabulary and nomenclature previously described in sections 2.1 and 2.3 guarantees an easier re-use of the datasets. Any diversion from the well-established good practices for the data submission will be addressed in the metadata.

The time for the data to remain re-usable includes the timeframe for the current technology developed to enter in service and indefinitely after this stage. The time of the data to remain re-usable is subjected to the restrictions given by the Topic Manager's background IP. If any other improvement or upgrade to the methods developed herein is achieved within the timeframe given, a re-assessment of the data must be done to obtain up-to-date information, and in consent and agreement with the Topic manager's background IP.

3. Allocation of resources

The costs involved for making the data FAIR are considered within the Horizon2020 CS2 (Clean Sky 2) Joint Technology Initiative (JTI) grant agreement under the work package (WP) 1.3 for deliverable preparation and progress reporting. The tasks within this WP are focused on assuring an appropriate FAIR data management.

Cranfield University in its role of leading coordinator will be responsible for the data management of the project. The preservation of the resources and research-generated data will be discussed in advance between the leading organization and the Topic Manager to decide the future of the datasets in terms of its potential value and preservation time.

4. Data security

As described in the previous sections, data will be stored on a Cranfield University institutional network drive as well as on the Exostar ForumPass 6 trusted workspace. Cranfield's network drives are backed up automatically by Cranfield's IT on a daily basis to two separate data centers. Cranfield University uses Microsoft Active Directory controls to ensure 'role-based access controls' are enforced, which means that user permissions to content are controlled by their User ID and password. Access to these drives has been authorized only to the consortium participants. Cranfield University employs perimeter *Firewalls* to protect internal networks and IT systems, including the *FileStore* that hosts the shared drives. The university also undertakes monthly vulnerability scans to ensure systems remain patched and up-to-date and has a penetration testing schedule for all its major systems.

Final datasets will also be stored on CORD, where the data can be retained for at least 10 years. Cranfield University also has an ongoing project looking to implement processes and systems to optimize the digital preservation of such datasets.

5. Ethical aspects

PROTEUS-related research will be checked for ethical compliance locally by each participating University and will need to obtain necessary approvals. Any data used within the PROTEUS project will be strictly selected from either publicly available sources or data generated by the project itself. There is no plan to use third party data or to collect any kind of personal data within PROTEUS. Data to be published will be reviewed by the Topic Manager before publication.

6. Other issues

Cranfield University has allocated approximately 8k Euro to support open access publications and other dissemination activities during the 3-year period. Cranfield also has a policy in place to support Open Access publications through additional funds if needed.

Acknowledgement

SUMMARY TABLE 4

FAIR Data Management at a glance: issues to cover in your DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation The nature of the project datasets are experimental, numerical and a set of software tool codes. The purpose of the collected experimental data is to validate the numerical data and the generated computational tools. The several methods developed by the Consortium partners will be capitalised into software toolsets to predict the engine performance and operability during idle and sub-idle conditions. • Explain the relation to the objectives of the project The data generated within the project will stem from simulation results of engine component models as well as whole-engine performance simulation models. The high-fidelity 3-D CFD data is to be reduced into component characteristics for integration into whole-engine performance tools and obtain 0-D data for the entire gas turbine engine. • Specify the types and formats of data generated/collected Data format files to be used are: C (.c), C++ (.c, .h), comma separated values (.csv), text (.txt), simple data (.dat), portable document format (.pdf), tagged image (.tif), portable network graphics (.png), FORTRAN90 (.f90), PYTHON (.py), MATLAB (.m), NPSS (.int, .mdl, .map) • Specify if existing data is being re-used (if any) No analytical or numerical data is being re-used. The origin of all numerical data is coming from the current project itself. Test data is re-used from the Topic Manager. • Specify the origin of the data Datasets will be produced from 3-D CFD analyses, 0-D whole-engine analysis and from in-house software tools developed by the Consortium partners. • State the expected size of the data (if known) Final datasets are estimated to occupy broadly 1 Tb, although the expected data size will be frequently re-defined during the project.

	• Outline the data utility: to whom will it be useful
	The final software, methods and toolkits for the whole-engine performance simulation during idle and sub-idle conditions will be delivered to the Topic Manager, envisaged to be matured to TRL6 through validation against test data collected from the engine demonstrator program.
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) The management of the numerical and experimental data, and computational tools will possess file names and nomenclature according to a standard identification mechanism detailed in the metadata files. • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? At the end of the project, all available public domain data will be uploaded by Cranfield University onto CORD within 3 (three) months of the project end. CORD is an institutionally-approved secure repository, which retains data for at least 10 years. CORD uses the *figshare* platform, which is ISO27001 certified and assigns a *DOI* to each item, to ensure that datasets are findable. • Outline naming conventions used The research data as well as the metadata file names will have a series of subject identifiers separated between them by an underscore (_). The generated data will be archived by engine type and component including subdirectories for each case. The file names contained in every subdirectory will be consistent with its location address as a manner of finding and tracking the files and contain any possible misplacement. The file name standard convention to be used is decried by the engine type, component and analysis case in the format: *enginetype_component_analysiscase* .file extension. The engine type is given by the engine model. The analysis case is composed of the engine condition, mass flow rate and rotational speed. • Outline the approach towards search keyword Metadata such as keywords will be added to optimise possibilities for re-use. • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how The metadata *readme* .txt files that will accompany the data repositories of the project, including a list describing the contents of each directory and standard file nomenclature. The metadata will be clearly identified by the directory name and the label *readme* at the end to permit the user identify this file as metadata.

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
	Work undertaken within the PROTEUS project is sensitive and of commercial value to the Topic Manager, relating to recent and intended future products, and may include third-party data, supplied on the basis of a third-party licence from other companies. Datasets with background Intellectual Property (IP) supplied by the Topic Manager to the PROTEUS partners require to remain closed. These IP datasets include geometry, rig and engine test data, performance model information and software.
	Results of the PROTEUS project that are made openly available and public must not reveal background IP from the Topic Manager, such as geometry information. The PROTEUS project output data will be post-processed in such a way as to remove any Topic Manager's background IP to make it publicly accessible.
	• Specify how the data will be made available
	Making the data openly accessible requires formal clearance and consent from the Topic Manager. All peer-reviewed scientific publications resulting from the project will be Open Access where possible with previous clearance from the Topic Manager. The clearance given by the Topic Manager will make sure that no data to be made openly available carries background IP or is subject to a third-party license.
	• Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
	The data output and results generated from the PROTEUS project will be obtained from commercial software such as ANSYS CFX, in-house developed toolsets and methods, and modelling-system tool (NPSS) supplied by the topic manager. Main output data will be post-processed in such a way that no software interface is required to access it; however, if any raw data is required to be stored, documentation will be included for the case of the in-house software and NPSS. For any other raw data derived from commercial programmes, the documentation to access is openly available.
	As a general principle, the software source codes from the different partners will not be made available. Partners will produce publications describing the new methods applied in the software rather than publishing the source code nor the software directly. It is preferable from the Topic Manager not to publish any PROTEUS-derived software directly; however, if any foreground software from PROTEUS is to be made public then it must be ensured that this software does not contain any background Topic Manager's IP, including Topic Manager's IP software.
	A special case is related to the NPSS thermodynamic modelling-system supplied by the Topic manager to Cranfield University as a background IP. Neither the software itself nor its source code should be

made public, due to the Topic Manager's background IP, and because the NPSS code carries a US Export-Control rating which prevents publication.

- Specify where the data and associated metadata, documentation and code are deposited

Secured data will be stored in the Exostar ForumPass 6 information repository for the specific purposes and duration of the project. ForumPass 6 is a trusted workspace built upon the Microsoft Office SharePoint Server 2013 platform and is managed by the Topic Manager. The Exostar ForumPass 6 repository will allow access only to project participants. Metadata will be allocated in the lowest level of subdirectories to describe the datasets.

- Specify how access will be provided in case there are any restrictions

The topic manager is responsible for managing and granting access to the Exostar ForumPass 6 platform. Consortium partners will have access to the storage though a security system based on username and password. Since the Topic Manager's background IP protection has been agreed within the partners no data access committee is foreseen. Nevertheless, in the case of an external request for data access, Cranfield University as leading partner will select an internal data access committee.

2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
	All research-generated data is exchangeable and can be re-used between the Consortium partners. The exchangeable data shared among the project participants is mainly of numerical nature, and is transferred as a tabular data package in text or table-processor files as CSV (comma separated values) files. Due to the secure data storage and exchange platform managed by the Topic Manager, Exostar ForumPass 6, all experimental and numerical information can be formally shared and be accessible to all Consortium partners.
	• Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
	The vocabulary used within the exchangeable datasets is standard and follows the international standards for engine station numbering and nomenclature. The interoperable data vocabulary is according to the SAE (1974) Gas Turbine Engine Performance Station Identification and Nomenclature Aerospace Recommended Practice (ARP) 775A. The nomenclature used in the software toolset codes is according to the SAE Aerospace Standards (AS) for digital computer programs. For engine performance steady-state and transient simulations codes, the nomenclature standard is according to the SAE AS681IK. Any additional necessary parameters related time-based simulations will follow the SAE ARP1257.

2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible
	The relevant experimental and numerical data derived from the PROTEUS project without any Topic Manager's background IP and without a third-party data licence, will be clear and accessible through a data usage licence that will be released, requiring formal clearance and consent from the Topic Manager. No Topic Manager's background IP will be licenced or disseminate by the PROTEUS partners unless written permission is obtained from the Topic Manager in each case. In the case of every consortium partner, their own source code, in-house software, and technical data involving the Topic Manager's background IP will not be licenced.
	• Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
	Datasets are expected to be available for re-use at the end of the project with previous authorization and clearance from the Topic Manager and provided they are free of any Topic Manager's background IP. The Topic Manager encourages the partners to apply for patents for any novel technologies developed under the PROTEUS project. In this context, the Topic Manager prefers the partners to delay publication of any materials necessary to allow time for all relevant inventions to be protected.
	• Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
	Sensitive data that requires Topic Manager's background IP or holding a third-party data licence will remain restricted and will not be available for re-use due to the sensitive and commercial value, which may include third-party data. Any other re-use activity that does not involve the Topic Manager's background IP will have to be cleared and with previous consent by the Topic Manager in advance and may be used by external entities for comparison and validation. These datasets may be made publicly available for re-use immediately after the full completion of the PROTEUS project.
	• Describe data quality assurance processes
	To guarantee the re-use of data, a clear history of origin, methodology, data workflows and references will be detailed in the final reports. The utilization of metadata, standard files, vocabulary and nomenclature previously described in sections 2.1 and 2.3 guarantees an easier re-use of the datasets. Any diversion from the well-established good practices for the data submission will be addressed in the metadata.
	• Specify the length of time for which the data will remain re-usable
	Naturally, the time for the data to remain re-usable includes the timeframe for the current technology developed to enter in service and indefinitely after this stage. The time of the data to remain re-usable is subjected to the restrictions given by the Topic Manager's background IP. If any other improvement or upgrade to the methods developed herein is achieved within the timeframe given, a re-assessment of

	the data must be done to obtain up-to-date information, and in consent and agreement with the Topic manager's background IP.
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs The costs involved for making the data FAIR are considered within the Horizon2020 CS2 (Clean Sky 2) Joint Technology Initiative (JTI) grant agreement under the work package (WP) 1.3 for deliverable preparation and progress reporting. The tasks within this WP are focused on assuring an appropriate FAIR data management. • Clearly identify responsibilities for data management in your project Cranfield University in its role of leading coordinator will be responsible for the data management of the project. • Describe costs and potential value of long term preservation The preservation of the resources and research-generated data will be discussed in advance between the leading organization and the Topic Manager to decide the future of the datasets in terms of its potential value and preservation time.
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data As described in the previous sections, data will be stored on a Cranfield University institutional network drive as well as on the Exostar ForumPass 6 trusted workspace. Cranfield's network drives are backed up automatically by Cranfield's IT on a daily basis to two separate data centers. Cranfield University uses Microsoft Active Directory controls to ensure 'role-based access controls' are enforced, which means that user permissions to content are controlled by their User ID and password. Access to these drives has been authorized only to the consortium participants. Cranfield University employs perimeter *Firewalls* to protect internal networks and IT systems, including the *FileStore* that hosts the shared drives. The university also undertakes monthly vulnerability scans to ensure systems remain patched and up-to-date and has a penetration testing schedule for all its major systems. Final datasets will also be stored on CORD, where the data can be retained for at least 10 years. Cranfield University also has an ongoing project looking to implement processes and systems to optimize the digital preservation of such datasets.
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

	PROTEUS-related research will be checked for ethical compliance locally by each participating University and will need to obtain necessary approvals. Any data used within the PROTEUS project will be strictly selected from either publicly available sources or data generated by the project itself. There is no plan to use third party data or to collect any kind of personal data within PROTEUS. Data to be published will be reviewed by the Topic Manager before publication.
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any) Cranfield University has allocated approximately 8k Euro to support open access publications and other dissemination activities during the 3-year period. Cranfield also has a policy in place to support Open Access publications through additional funds if needed.

HISTORY OF CHANGES		
Version	**Issue date**	**Change**
1.0	August 2017	▪ Based on H2020 Initial version 1.0 October 2017
2.0	February 2018	▪ Update to Data Management Plan