CENTAUR: Cost Effective Neural Technique for Alleviation of Urban flood Risk

D4.4 Data Management Plan

Lead Partner: USFD

Revision: 29th October 2018

Report Details

Title: Data Management Plan

Deliverable Number (If applicable): 4.4

Author(s): Simon Tait, Will Shepherd

Dissemination Level: Public

Document History

Version	Date	Status	Submitted by	Checked by	Comment
1.0	9/3/2016	Draft	Simon Tait	Will Shepherd	
1.1	24/3/2016	Final	Will Shepherd	Simon Tait	Agreed at GA meeting, uploaded to Sygma
2.0	12/9/2016	Draft of updated version 2.0	Will Shepherd	Simon Tait	Changes requested by Project Officer
2.1	30/09/16	Draft	Simon Tait	Will Shepherd	Added new material
2.2	30/9/16	Draft	Will Shepherd	Simon Tait	Minor revisions made
2.3	04/10/16	Final	Simon Tait	Will Shepherd	Emailed to PO
3.0	15/05/17	Draft update	Will Shepherd	Simon Tait	Update following acceptance of version 2.3 and GA meeting and prior to upload to Sygma.
3.1	25/05/17	Final	Simon Tait	Will Shepherd	Uploaded to Sygma
4.0	28/08/17	Draft of update	Simon Tait	Will Shepherd	Revisions requested by Evaluator
4.1	01/09/17	Final	Will Shepherd	Simon Tait	Uploaded to Sygma
4.2	30/08/2018	Final update	Will Shepherd	Simon Tait	Dataset register updated, reference to data published on Zenodo added, uploaded to Sygma
4.3	29/10/18	Revision after review meeting	Simon Tait	Alma Schellart	Amend document as appropriate

Acronyms

AC	AC Aguas De Coimbra EM
DoA	Description of the Action (Annex 1 to the Grant Agreement)
DOI	Digital Object Identifier
EAWAG	Eidgenoessische Anstalt Fuer Wasserversorgung Abwasserreinigung Und Gewaesserschutz
EMS	Environmental Monitoring Solutions
FCD	Flow Control Device
IPR	Intellectual Property Rights
UoC	Universidade de Coimbra
USFD	University of Sheffield

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 641931.

Executive Summary

CENTAUR was part of a pilot on open access being run within the H2020 research program. As part of the pilot, CENTAUR was required to produce a Data Management Plan. The H2020 research program is promoting open access of data and publications as the European Commission believes that the wide availability of data will lead to optimal use of public funding by reducing duplication and encouraging and supporting future research and innovation in a cost efficient manner. CENTAUR was an innovation project rather than a research and development project. The project's Data Management Plan attempts to follow the principle of open data access whilst accepting the need for confidentiality to address privacy needs to protect personal data, and to provide for Intellectual Property Rights (IPR) protection and the commercial confidentiality of the partners, especially for the non-University partners who have contributed financially to the project activities. These constraints and how the partners acted regarding these constraints were clearly set out in the Project Consortium Agreement. The Data Management Plan now describes how the consortium managed the competing needs of the partners with the aspirations of the European Commission.

The Data Management Plan addresses how the partners collected data, catalogued it and, when appropriate, made it available on an open access basis during and after the project. The plan also described the review mechanism the consortium used to ensure that as much of the data collected during the project was made available as soon as was practicable. All academic publications from the project were made available in an open access repository.

The Lead Beneficiary provided facilities for storage of open access data and archived this data and deposited it in an enduring open access data repository before the end of the project.

The Data Management Plan was reviewed at each General Assembly meeting. A revised plan was issued annually. Each revision of the Data Management Plan listed the open access data sets and also the data that was held confidential and the reason for this categorisation was also described. This approach was intended to provide an appropriate balance between the aspiration for open access data and the need to retain some data within the consortium to support effective market replication and exploitation so that public benefit, in terms of jobs growth and enhanced flood protection, could be obtained via readily available CENTAUR systems.

CONTENTS

1 Introduction

1.1 Partners Involved in Deliverable

USFD – this deliverable has been drafted by USFD and has been commented on by all partners in the CENTAUR consortium.

1.2 Project Details

CENTAUR - Cost Effective Neural Technique to Alleviate Urban Flood Risk

Funded by: European Commission – Contract No. 641931

Start Date: 01 September 2015

Duration: 36 months

Contact Details: centaur@sheffield.ac.uk

Co-ordinating Institution: University of Sheffield

Website: www.sheffield.ac.uk/centaur

1.3 Project Summary

The project developed a radically new market-ready approach to real time control (RTC) to be used within sewer networks with the aim of being able to reduce local flood risk in urban areas in a highly cost effective manner. Existing RTC projects (e.g. in the cities of Vienna, Dresden, and Aarhus) are characterised by complex sensor networks, linked to high cost centralised control systems governed by calibrated hydrodynamic modelling tools and often fed by high cost and complex radar rainfall technology. Such systems are expensive and complex to install and operate, requiring a high up-front investment in new infrastructure, communication equipment and control systems, and require highly trained staff. In contrast, this CENTAUR has developed a novel low cost decentralised, autonomous RTC system. The concept is to be able to install such low cost RTC systems in existing infrastructure and for these to require low levels of maintenance and staff input. During the project the CENTAUR system was installed, tested and demonstrated in two networks, a combined sewer network in Coimbra, Portugal and a stormwater network in Toulouse. This RTC approach utilised data driven distributed intelligence combined with local, low cost monitoring systems installed at key points within existing sewer infrastructure. The system utilised mechanically robust devices to control flow in order to reduce flood risk at vulnerable sites. This system was informed and its control governed directly by sensors distributed within the local network, without the need for an expensive hydrodynamic model or real time rainfall measurements. The system delivered many of the benefits of existing RTC systems, but avoided the high costs and complex nature of extensive sensor networks, centralised control systems, communications systems and infrastructure modifications. The developed system has therefore proven to be of significant benefit to operators of small to medium sized sewer networks, because of its low up-front capital cost and its high cost benefit when used to control localised flooding.

2 Policies

The project participants at all times met their obligation on the access rights and non-disclosure of data as set out in the project Consortium Agreement. Nothing in the Data Management Plan removed any rights or obligations as set out in the Consortium Agreement

The project aimed to follow the H2020 guidelines as regards open access and data management and also adhered to the principles of the data management policy of the co-ordinating institution, the University of Sheffield.

H2020 Guidelines:
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

University of Sheffield Guidelines on the Management of Research Data:
http://www.sheffield.ac.uk/library/rdm.

The Data Management Plan was reviewed by all partners at each General Assembly meeting and a revision was re-issued every 12 months.

3 Data Collection, Documentation Sharing and Storage

3.1 Overview

The European Commission has recognised that to achieve the best return for their funding of research and development activities, any of the resultant research data should be capable of re-use. This is best achieved by making data and publications openly accessible. The data from CENTAUR was made openly accessible, subject to any constraints set out in the Consortium Agreement on data ownership and its use by other parties. These constraints related to compliance with any national legal requirements (e.g. Personal Data), the protection of IPR and commercial confidentiality in order to achieve effective market replication and exploitation of the CENTAUR technology and supporting knowledge developed during the project.

Subject to the above constraints data created within the project was made available, once it had been processed into a final formal, organised and catalogued and was free of error. Appendix A contains a table of all completed data sets, including whether the data is open access or not. Partners used current best practice in terms of data collection processing and validation and ensured that sufficient resources were made available from the project funds to complete these tasks. Adequate description of the context, measurement and processing methods were also made available for the data that was made publically available. Detailed information was linked to each open data set so that it was clear how it was structured. Adequate documentation was also provided so that the open data sets were searchable by a 3rd party. Each open access data set included information on the sensors used, their calibration and validation, and the file and parameter naming conventions.

The co-ordinator listed the available open access data sets.

The co-ordinator hosted the open access data electronically and transferred all accessible open access data to Zenodo (http://www.zenodo.org) an enduring open access repository before the completion of the project. An open access software tool, used to locate potential locations for the installation of a CENTAUR system was stored on GitHub – (https://github.com/) making it readily available.

The peer-reviewed scientific publications arising from the work in CENTAUR followed the requirements set out in the Grant Agreement and Consortium Agreement. They were all openly accessible as this was a requirement of the Grant Agreement. All publications were stored in an OpenAIRE compliant repository and listed on the CENTAUR page in the OpenAIRE portal. The co-ordinator also listed the details of all publications on the project website, along with links to access the publications.

Appendix A of the Data Management Plan lists the completed data sets produced during the project. It also lists those data sets that are open, and those that are restricted to the members of the consortium along with the reason why any data set had been restricted. Data sets were only restricted for one of three reasons: to comply with national regulations for the protection of personal data; to protect IPR for future exploitation; and for data that was commercially confidential and the release of which would be financially damaging to a partner.

After generating a data set partners were required to list it in Appendix A of the Data Management Plan and then state whether the data was to be open access, if this was not possible then the reason was given as to why the data set was not to be open access. These decisions were reviewed periodically at the subsequent General Assembly meetings. If any objections was raised as to the status of any data set, this was discussed at a General Assembly and then a final decision on the status of a data set was taken by the General Assembly following the decision making process described in the Consortium Agreement.

3.2 Data handling during and after the project

The project data was collected or generated primarily by the University of Coimbra (UoC), University of Sheffield (USFD) and later in the project by Veolia. These partners were supported in the field data collection by Environmental Monitoring Solutions (EMS). Steinhardt generated data on the system design, EAWAG generated simulation data. Aguas de Coimbra were involved in the data being collected by UoC, but did not generate any data themselves.

As a general principle, the primary responsibility for storage and handling of the data lay with the partner originally collecting it. An understandable data structure was always used for any data collected. For field and laboratory data collection, filenames incorporated the date of collection and where appropriate the sensor id. This information was then linked to a spreadsheet providing further details, including any calibration parameters and comments on any issues affecting data quality. For both laboratory and virtual testing datasets the filename incorporated the date of the test or simulation and/or number of the

test or simulation. The date and run number was linked to a spreadsheet summarising the testing carried out and including the relevant parameters for the test or simulation run. For field data, the datasets covered a longer period, hence the filename included both start and end date if applicable, but otherwise conformed to the same basic standards as the laboratory and virtual testing data sets.

Key metadata was stored alongside the data, for field and laboratory measurements this included calibration data, sensor details, sensor location, and details of the tests that were carried out. For virtual testing the metadata included information on the hydrodynamic model, the version of the algorithm and parameters used and the rainfall event(s) run. All data was checked prior to storing, these checks were primarily 'sense checks' such as mass balances and where practical, cross-checking data between sensors for consistency.

Data was backed up on a regular (weekly) basis, with the backup stored at a different site by the partner that had collected it.

Some of the data was useful to other partners and was shared as needed via the project's user controlled Google Drive folder. This data store was provided by the University of Sheffield and is password protected and provided an appropriate level of protection for data used within the project. The folder was managed by the Project Co-ordinator.

It was the responsibility of the partner collecting the data to deem it open access or restricted within the consortium or restricted to within the partner organisation following the principles outlined above. For a datasets which deemed to be suitable for open access, the Project Co-ordinator worked with the partner that generated the data and organised its placement in the Zenodo data repository which enabled the data set to be assigned a unique DOI (Digital Object Identifier). The data was linked to the CENTAUR community on Zenodo (https://zenodo.org/communities/centaur/) and the data was also linked to the CENTAUR page on OpenAIRE (https://www.openaire.eu/search/project? projectId=corda__h2020:: a468749db757b4bb290b04b284706d8a). The project co-ordinator ensured that the data sets uploaded to this repository were quality checked and placed in a structured manner that provided 3rd parties the ability to search and use the data. Discoverability of the data sets was ensured by including a clear abstract / description and relevant keywords within the Zenodo record, any publications referencing the data would use the DOI. Keywords included the project name acronym, keywords listed in the DoA and any additional keywords specific to the data set (e.g. laboratory water depths). For software tools these deposited on GitHub, where version control is a core feature of this platform.

After the project finished the coordinator collected all the internal data on the project's Google Drive folder, archived it and will store it on at an institutional secure storage area. This will be for a period of at least 5 years. This is the definitive record of the project data, the Google Drive service is subscribed to institutionally by USFD, hence there are no direct costs associated with the project. This data will be available to project partners for this 5 year period, during which any follow on publications or studies are most likely be completed. There is no need for data recovery as the Google Drive is mirrored across multiple sites, accidental deletion is very unlikely as files are removed to a 'trash' folder and only completely deleted if subsequently removed from the trash folder.

Appendix B includes an example of information that has been included as part of the metadata for the Open Access data sets. The open access data is expected to be useable for the foreseeable future after the project ends, the repository used is free to use, hence no costs are involved, it is publically funded so will be expected to be enduring.

3.3 Summary of data being collected, processed and generated

A number of separate datasets were generated during the CENTAUR project. The majority of datasets had common features in that the parameters recorded related to flows and depths in sewer pipes, or on an urban catchment surface or to the status of the flow control device, these data sets were time series collected at a single location. The other types of data are the Steinhardt flow control device designs, the EMS LCMS designs and specifications and the data created by EAWAG from the use of site selection methodologies and tools.

3.3.1 Flow survey data for development of the dual drainage model

3.3.1.1 Purpose

To calibrate and verify the dual drainage model of the Coimbra pilot study catchment.

3.3.1.2 Relation to project objectives

A calibrated dual drainage model was required to allow the performance of the urban drainage network to be better understood and allow selection of a site to install the flow control device for pilot testing. The model was used in virtual testing to assess the performance of the flow control device (see 3.3.2).

3.3.1.3 Timescale

Winter 2015 and Spring 2016.

3.3.1.4 Types and formats

Observational data from installed pressure transducers and flow monitors. The data was stored uncompressed and unencrypted in ASCII and/or spreadsheet formats.

3.3.1.5 Methodologies and standards

Data collection and analysis was guided by the document ‘A guide to short term flow surveys of sewer systems’ (WRc, 1987).

3.3.1.6 Access to data and required metadata

This data was made accessible as the associated metadata required to make the data re-usable includes location details of the sewerage network, which is the confidential property of the water company which owns the sewer network. This data can also be used to identify the flood risk of individual properties, and its release can therefore have a significant financial impact on individuals. This data was retained securely by the partners that collected and initially used it (EMS and UoC). It was shared with UFSD and EAWAG,

as they required it to complete their tasks. This sharing was done via password protected files and via the password protected project Google site folder. The key metadata included the locations of the data collection, information on the surrounding drainage network, the sensor specifications and calibration details. This information was stored alongside the stored flow data.

3.3.2 Virtual testing simulation data

3.3.2.1 Purpose

To develop and test the CENTAUR control algorithm using previously calibrated hydro-dynamic sewer network models.

3.3.2.2 Relation to project objectives

Prior to implementing the flow control device on an operational sewer network it was tested both in the laboratory and using hydrodynamic models to confirm that the control algorithm was stable and safe.

3.3.2.3 Timescale

From Spring 2016 until Spring 2018.

3.3.2.4 Types and formats

Simulation data from calibrated hydro-dynamic models. The data was stored uncompressed and unencrypted in ASCII and/or spreadsheet formats.

3.3.2.5 Methodologies and standards

The models were produced in accordance with the 'Code of Practice for the Hydraulic Modelling of Sewer Systems' (WaPUG, 2002).

3.3.2.6 Access to data and required metadata

This data was not made accessible as the associated metadata required to make the data re-usable included details of sewerage networks, which is the confidential property of the water companies which own the sewers. This data can also be used to identify the flood risk of individual properties. This data was retained securely by the partners that collected and used it (EMS and UoC). It was shared via password protected files and via the password protected project Google site folder. The key metadata included details of the network model, the version of the software and the model calibration parameters used in the simulations. This information was stored in a spreadsheet format alongside the results produced.

3.3.3 Laboratory testing

3.3.3.1 Purpose

To test the CENTAUR flow control device hardware and the control algorithm.

3.3.3.2 Relation to project objectives

Prior to implementing the flow control device on an operational sewer network it was tested both in the laboratory and using hydrodynamic models to confirm that the control algorithm was stable and safe and that the hardware was reliable and operated as expected for the pilot study.

3.3.3.3 Timescale

Summer 2016 to Autumn 2017

3.3.3.4 Types and formats

Experimental data from the laboratory test facility constructed for CENTAUR. The data was stored uncompressed and unencrypted in ASCII and/or spreadsheet formats.

3.3.3.5 Methodologies and standards

There are no relevant standards, however the data was collected by calibrated sensors and checked for consistency before being accepted.

3.3.3.6 Access to data and required metadata

This data has been made accessible via the Zenodo data repository, it can be accessed via the DOI 10.5281/zenodo.1406296.

Metadata concerning the laboratory rig dimensions and information on the sensors was provided. Detailed technical information on the control algorithm which operated the flow control device was commercially sensitive and was not provided.

The data will primarily be of interest to anybody wishing to replicate results presented in published papers, there is unlikely to be a significant amount of re-use as the data is very context specific. The total amount of data shared was 500 MB, the measured data was compressed on Zenodo reducing the download to 80 MB.

The data was made available on a Creative Commons Attribution-ShareAlike licence (https://creativecommons.org/licenses/by-sa/4.0/).

At the time of writing Zenodo listed 48 unique views of the data and 56 unique downloads.

3.3.4 Coimbra/Veolia pilot and demonstration testing

3.3.4.1 Purpose

To test the CENTAUR flow control device hardware and the control algorithm.

3.3.4.2 Relation to project objectives

Following virtual and laboratory testing, the flow control device and control algorithm was tested in the Coimbra sewer network in Portugal and then in a demonstration site in Toulouse managed by Veolia.

3.3.4.3 Timescale

From 2016 to September 2018.

3.3.4.4 Types and formats

Observational data from the installed pressure transducers and the flow control device status. The data was stored in uncompressed and unencrypted in ASCII and/or spreadsheet formats.

3.3.4.5 Methodologies and standards

There are no relevant standards, however the data was collected by calibrated sensors and checked for consistency before being accepted.

3.3.4.6 Access to data and required metadata

This data was not made accessible as the associated metadata required to make the data re-usable included details of sewerage networks, which is the confidential property of the water companies which own/manage the sewers. This data can also be used to identify the flood risk of individual properties. This data was retained securely by the partners that collected and used it (EMS, UoC and Veolia). It was shared via password protected files and via the password protected project Google site folder. The performance data from the demonstration site was also commercially sensitive as it can be used to develop the commercial business case for the deployment of CENTAUR. The key metadata included the locations of the data collection, information on the surrounding drainage network, the sensor specifications and calibration details. This information was stored alongside the data by UoC and Veolia.

3.3.5 Flow Control Device design data

3.3.5.1 Purpose

Design information for the developed flow control device.

3.3.5.2 Relation to project objectives

The flow control device was a key part of the CENTAUR system, allowing flows in the drainage network to be controlled.

3.3.5.3 Timescale

The design developed between the start of the project and the finalisation of the design for the demonstration site, i.e. September 2015 to December 2017.

3.3.5.4 Types and formats

The data consisted of drawings, written specifications and tables showing the calculated flow rates under different conditions. These were archived in pdf format.

3.3.5.5 Methodologies and standards

N/A

3.3.5.6 Access to data and required metadata

This data was not made accessible as the design is a key part of the CENTAUR IP and know-how. It was shared via password protected files and via the password protected project Google site folder, for partners that required technical information on the FCD (Veolia, Aguas de Coimbra, UFSD, EMS). There was not any requirement for metadata beyond what was already within stated within the design documents.

3.3.6 LMCS design data

3.3.6.1 Purpose

Design information for the developed Local Monitoring and Control System (LMCS).

3.3.6.2 Relation to project objectives

The LMCS was a key part of the CENTAUR system, allowing monitoring of the water levels, processing of data and communication of control actions to the FCD.

3.3.6.3 Timescale

The design of the LMCS developed throughout the project. The CE and ATEX certification was completed in August 2018.

3.3.6.4 Types and formats

The data consisted of circuit diagrams, code and written specifications. These were archived in a pdf format.

3.3.6.5 Methodologies and Standards

N/A

3.3.6.6 Access to data and required metadata

The data was not made publically available, as the design was a key part of EMS's commercially valuable intellectual property and know how. Any data required to be shared among partners was shared via password protected files. There was not any requirement for metadata beyond that stated within the design documents.

3.3.7 Site selection methodology and results

3.3.7.1 Purpose

Developing a methodology to select optimum sites for the deployment of CENTAUR.

3.3.7.2 Relation to project objectives

In order to efficiently market CENTAUR, a methodology to select sites from commonly available catchment and drainage network data was required.

3.3.7.3 Timescale

This part of Task 3.4 commenced early and developed throughout the project between October 2016 and April 2018.

3.3.7.4 Types and formats

The data output form the methodology scored/ranked the suitability of different parts of the drainage network for installation of a CENTAUR system and was in ASCII format.

The methodology is in the form of a java based software tool.

3.3.7.5 Methodologies and standards

There are no relevant standards for the output data. The software tool was version controlled through a GitHub repository.

3.3.7.6 Access to data and required metadata

The output data was not made accessible as the associated metadata required to make the data re-usable included details of sewerage networks, which is the confidential property of the water companies which own the sewers. This data was retained securely by EAWAG. It was shared via the password protected project Google site folder with UoC, UFSD and EMS who required access to complete some their tasks.

The software tool is openly access through a GitHub repository (https://github.com/ldesousa/centaur.loc), this repository includes the relevant metadata to allow the tool to be run (i.e. instructions). This tool can utilised by other researchers and practitioners to systematically investigate potential in sewer storage.

4 Legal and Ethical Compliance

At all times the partners complied with national legal requirements as regards the protection of personal data. The co-ordinating institution has a rigorous policy on the collection and storage of personal data (http://www.sheffield.ac.uk/library/rdm/expectations). This was adhered to by all partners. After an assessment at the start of the project by the Project Co-ordinator found that no personal data was planned to be collected in this project. No partners generated personal data during the project.

5 Long Term Storage and Archiving

The co-ordinator provided electronic storage facilities for open access data and its metadata created by any partner during the project. Open access data was uploaded to the Zenodo data archive. Any open access software tool produced was stored on GitHub.

The co-ordinator did not provide long term storage for any personal data, or data that is required to protect IPR and commercially confidential information.

At the end of the project, the co-ordinator archived any files and data (not containing personal data or commercially confidential information) on the shared project Google drive and made this available to all the partners.

All peer-reviewed scientific publications relating to the results of CENTAUR were openly accessible. The partner producing any publication was responsible for storing these publications in an enduring repository which is compatible with OpenAIRE (it can be institutional, subject-based or centralised) as soon as possible, and at the latest on publication. Such publications were listed and linked to OpenAIRE (at https://www.openaire.eu/search/project? projectId=corda__h2020:: a468749db757b4bb290b04b284706d8a) and also provided links for access on the project website.

6 Data security

Data was stored securely, to ensure its integrity and also to ensure compliance with personal data protection regulations, IPR protection and commercial confidentiality.

Devices that contained data were password protected and securely stored when not in use. Data sets available online were in a password protected folder, such as the project's Google Drive.

Data that was open access was not password protected and was made available via the open access data repository Zenodo.

7 Summary

The CENTAUR project endeavoured to make the data produced open access following the H2020 guidelines. The partners took into account any constraints on data availability described in the Consortium Agreement and any national legal requirements on the protection of data.

The project beneficiaries ensured that sufficient resources were made available from the project funds to ensure that all the data sets that are uploaded onto the open access repository Zenodo are organised, catalogued and practically free of error, and that sufficient metadata was provided so that a third party can use the data.

The partners ensured all peer-reviewed scientific publications relating to the results of CENTAUR were available through an open access route and were listed on the CENTAUR page of the OpenAIRE portal.

The co-ordinator collated a list of all data collected during the project and required partners to declare whether data was open access or restricted in line with the policy outlined in the Data Management Plan. Access to Open data was unrestricted, apart from where an embargo period is deemed necessary to allow academic publications to be finalised. Appendix A of the Data Management Plan lists all completed data sets and their

availability. The co-ordinator ensured that all open access data produced during the project was appropriately archived and deposited in an enduring open access repository.

The Data Management Plan has been reviewed periodically at each General Assembly and contains a record of the data sets collected and the status of each data set as regards its availability.

8 References

WaPUG, 2002. *Code of Practice for the Hydraulic Modelling of Sewer Systems*. Version 3.001. Available at: http://www.ciwem.org/wp-content/uploads/2016/05/Code-of-Practice-for-the-Hydraulic-Modelling-of-Sewer-Systems.pdf

WRc, 1987. *A guide to short term flow surveys of sewer systems.* Swindon: WRc Engineering.

Appendix A. Register of Completed Datasets

Date added to Data Register	Lead Partner	Description of Data (Data, metadata, format and size)	Status of Data (Open / Restricted)	Justification of Status / Other comments	Location of Data
03/10/2016	UoC	Flow survey data for development of the dual drainage model. Flow and water level data at various locations in the Coimbra sewer network. Text and spreadsheet format, several hundred megabytes in total.	Restricted	Cannot share vital metadata, see 3.3.1.6. Data may identify individual property's flood risk – can have a significant impact on an individual.	UoC
29/09/2017	EAWAG	Site selection methodology and results. Site selection tool developed in java. Results from the tool in ASCII text format. Total 40 MB.	Site selection tool Open Access; Results restricted	Cannot share vital metadata relating to results, see 3.3.6.6. The metadata includes details of sewerage networks, which is the confidential property of the water companies which own the sewers.	EAWAG; site selection tool available at https://github.com/ldesousa/centaur.loc
26/01/2018	Steinhardt	Flow Control Device design data. Autocad drawings and calculations, stored as dwg files, pdfs and spreadsheet format. Total 20 MB.	Restricted	This data is a key part of the CENTAUR IP.	Steinhardt

Date added to Data Register	Lead Partner	Description of Data (Data, metadata, format and size)	Status of Data (Open / Restricted)	Justification of Status / Other comments	Location of Data
30/04/2018	UoC with USFD	Virtual testing simulation data. Hydrodynamic model outputs. ASCII text files, around 20 Gigabytes in total.	Restricted	Cannot share vital metadata, see 3.3.2.6. Data may identify individual property's flood risk – which can have a significant impact on an individual.	UoC
25/05/2018	UoC	Coimbra pilot and demonstration testing. Recorded data from CENTAUR system, and additional flow, depth and rainfall sensors, stored as ASCII text files, Matlab .mat files and spreadsheet files. Total 800 MB.	Restricted	Cannot share vital metadata, see 3.3.4.6. Data may identify individual property's flood risk – personal data.	UoC
28/06/2018	USFD	Laboratory testing data. Experimental data and associated metadata, stored as ASCII text files with some metadata in pdf format. Total 500 MB.	Open Access	All recorded laboratory data is available. Information on the control algorithm is commercially sensitive and has not be provided.	USFD and DOI 10.5281/zenodo.1406296
31/08/2018	Veolia Eau	Veolia demonstration testing. Recorded data from CENTAUR system, and additional flow, depth and rainfall sensors, stored as ASCII text files, Matlab .mat files and spreadsheet files. Total 350 MB.	Restricted	Cannot share vital metadata, see 3.3.4.6. Data may identify individual property's flood risk.	Veolia Eau