GRANT AGREEMENT No.: 764902		
Project Acronym: TOMOCON		**TOMOCON** Smart Tomographic Sensors for Advanced Industrial Process Control
Project title: Smart tomographic sensors for advanced industrial process control		
Deliverable Rel. No. **D7.6.**	Lead Beneficiary **HZDR**	
EU Del. No. **D30**	Type **ORDP: Open Research Data Pilot**	
WP No. **WP7**	Date: **28.02.2018**	Revision: **0**

Innovative Training Network

TOMOCON

Deliverable Title

Data Management Plan

Description

This deliverable describes the purpose of data collection and the types and formats of generated data. It also outlines the discoverability and identifiability of data and specifies how the data will be made openly available.

Prepared by:	Prof. Dr. Uwe Hampel	*U. Hal*
Approved by:	Prof. Dr. Uwe Hampel	*U. Hal*
Approved by Supervisory Board:	19.03.2018	

Dissemination Level: **Public**

Revision Sheet

Revision Number	Purpose of Revision	Effective Date
0	Initial Issue	19.03.2018

Dissemination Level: **Public**

Content

Dissemination Level: **Public**

TOMOCON Data Management Plan

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: To whom will it be useful**

Data within TOMOCON is generated via experimental studies and numerical multi-physics simulation of exemplary industrial processes. The data serves the following purposes with the TOMOCON project:

1)   Data is being generated to enhance the understanding of fundamental physical and chemical sub-processes in an industrial process scenario. This concerns transport of momentum, mass and energy in diverse fluid-flow dominated processes, propagation of electromagnetic fields or sound fields as well as kinetics of processes like crystallization.

2)   Data is being used to simulate and assess the performance and interplay of tomographic sensors and control systems in given industrial process model scenarios.

The TOMOCON project considers exemplarily the following industrial processes as model processes for the demonstration of the new technologies: Continuous steel casting, batch crystallization, inline fluid separation and microwave drying of porous products.

Data being generated typically represents adequately sampled four-dimensional physical parameter fields and is complemented by data for geometry specifications (e.g. CAD files) and boundary as well as initial conditions. It originates from the following sources:

1)   Numerical simulation data originates from the computational calculation of physical field quantities with specific commercial or proprietary simulation codes.

2)   Experimental data is digitized measurement data from specific sensors and instrumentation, e.g. for temperature, pressure, flow rate or filling-level, and from particle image velocimetry, high-speed video imaging, infrared thermography or diverse tomographic imaging techniques.

It is expected that data being produced within the TOMOCON project is essentially new data as it is based on novel methods, technologies and sub-models.

By its nature the data being generated within TOMOCON will be of large size (typically tens of megabytes to few gigabytes per data set) and of diverse and often proprietary digital formats and encoding.

Dissemination Level: **Public**

TOMOCON partners will share data in order to commonly develop new models, sensors and process control systems. Moreover, some of the data may be of interest for other scientists' groups to use them for code validation and own sensor or model developments.

The latter is subject of this data management plan and further referred to as TOMOCON Open Access Data. TOMOCON Open Access Data shall undergo a dedicated quality assurance before publication.

2. FAIR data

2.1. Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **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?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

TOMOCON Open Access Data will be made accessible by a unique digital object identifier.

Significant metadata will be provided to ensure discoverability and identifiability. Generally metadata provision follows the DataCite Metadata Scheme. Provision of specific metadata beyond that scheme is in the responsibility of the individual partners. However, the following type of metadata is suggested:

Numerical data: Software and version used for data generation; input data including boundary conditions and numerical grid; methods of simulation time step control; used sub-models e.g. for turbulence, kinetics, heat transfer etc.; digital output data format.

Experimental data: Description of the experimental setup; reference to geometry data, boundary conditions, specifications of relevant materials and fluids; instrumentation and their specifications (sampling rate, operational limits, accuracy/uncertainties).

Naming of data has to be done in a way that is most clearly indicating the type of data and the general background of its generation (e.g. experimental vs. numerical, type of experiment, related industrial application, purpose of the study).

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**
- **Specify how the data will be made available**

Dissemination Level: **Public**

- **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)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Making open access data accessible is in the responsibility of the individual partners. Many partners have their own institutional data repositories with specific procedures and access rules. If partners have no own repository they may either use public repositories, such as for example the Zenodo repository at CERN, or repositories of other TOMOCON partners. For the latter HZDR as the Coordinator will offer its RODARE repository. As other platforms HZDR's RODARE is interconnected to research data harvesters like OpenAire of the EU to ensure most efficient retrievability of data.

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.**
- **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?**

Alignment with standardized ontologies, such as DCAT Data Catalog Vocabulary is strongly encouraged.

2.4. Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest re-use possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **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**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

TOMOCON Open Access Data is public and recommended to be licensed according to Creative Commons Attribution 4.0 International (CC BY 4.0). For public software licenses GPLv3 is recommended. Embargo periods of up to 12 months may be imposed to restrict the use of data within the TOMOCON consortium.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

Dissemination Level: **Public**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long-term preservation**

TOMOCON Open Access Data data will be stored in repositories bound to the FAIR principles. Costs incurring to the partners in form of labour expenditure for preparing the data publication is covered by the EU funding. It is expected that costs in form of labour expenditure after the funding period will be minimal (e.g. by any kind of updating) and can be born by the respective partners via their institutional funding. The responsibility for data management is with the individual principal investigators of the research groups where the data has been produced.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

It is expected that the chosen institutional and public data repositories provide an adequate frame for secure data storage and recovery. No personal data will be stored with TOMOCON Open Access Data sets.

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

There are no ethics issues with TOMOCON data according to the DoA. All work including data generation will follow best practice guidelines of the EU and the existing national rules.

Dissemination Level: **Public**