HORIZON 2020	***The Framework Programme for Research & Innovation*** ***Research and Innovation actions (RIA)***

Project Title:

Multimodal Scanning of Cultural Heritage Assets for their multi-layered digitization and preventive conservation via spatiotemporal 4D Reconstruction and 3D Printing

Scan4Reco

Grant Agreement No: 665091

[H2020-REFLECTIVE-7-2014] Advanced 3D modelling for accessing and understanding European cultural assets

Deliverable

D1.2. Data Management Plan

Deliverable No.		D1.2	
Work package No.	**WP1**	Work package Title and task type	**Project Management**
Task No.	**T1.3**	Task Title	**Management of Data, Knowledge and Intellectual Property Right Issues.**
Lead beneficiary		**OF-ADC**	
Dissemination level		**PU – Public**	
Nature of Deliverable		**Report**	
Delivery date		**31 March 20156**	
Status		**Final**	
File Name:		**[Scan4Reco] Deliverable 1.2.doc**	
Project start date, duration		**01 October 2015, 36 Months**	

Authors List

Leading Author (Editor)			
Surname	*Initials*	*Beneficiary Name*	*Contact email*
Apostolidis	GA	OF-ADC	g.apostolidis@artdiagnosis.gr
Co-authors (in alphabetic order)			
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*

#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	Drosou	AD	CERTH	drosou@iti.gr
2	Dimitriou	ND	CERTH	nikdim@iti.gr
3	Karagiannis	GK	OF-ADC	g.karagiannis@artdiagnosis.gr
4	Krukowski	AK	RFSAT	Artur.Krukowski@rfsat.com
5	Vogiatzaki	EV	RFSAT	emmanouela@rfsat.com
6	Giachetti	AG	UNIV	andrea.giachetti@univr.it

Reviewers List

List of Reviewers (in alphabetic order)				
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	Taka	ET	CERTH	evitaka@iti.gr
2	Krukowski	AK	RFSAT	Artur.Krukowski@rfsat.com

Document history			
Version	Date	Status	Modifications made by
1.0	16/2/2016	deliverable structure preparation	OF-ADC
2.0	28/3/2016	1st draft version of the full deliverable	OF-ADC
3.0	31/3/2016	Final version of the full deliverable	OF-ADC

List of definitions & abbreviations

Abbreviation	Definition
3D	Three Dimensional
APD	Avalanche Photo-Diode
CCD	Charged-Coupled Device
CH	Cultural Heritage
DMP	Data Management Plan
ID	Identifier
IR	Infrared
LED	Light-Emitting Diode
RAID	Redundant Array of Independent Disks
UV/VIS	Ultraviolet/Visible
TBD	To Be Defined

Executive Summary

This document presents a first version of the project Data Management Plan (DMP). DMP provides an analysis of the various datasets that will be produced by the project and the main elements of the data management policy that will be used by the beneficiaries with regard to all these datasets that will be generated by the project. This version of the document reflects the current state of the datasets paving the way for further updates during the lifecycle of the project.

Table of Contents

List of Tables

1. Introduction

1.1   *Project overview*

Scan4Reco will develop a novel portable, integrated and modular solution for customized and thus cost-effective, automatic digitization and analysis of cultural heritage objects (CHOs), even in situ. A multi-sensorial 3D scanning - facilitated by a mechanical arm - will collect multi-spectra data and then, a hierarchical approach for 3D reconstruction of CHOs will be applied, enabling multi-layered rendering, advancing both analysis and 3D printing procedures. The goal will be to create highly accurate digital surrogates of CHOs, providing also detailed insight over their surface and also the volumetric structure, material composition and shape/structure of underlying materials, enabling rendering either via visualization techniques or via multi-material 3D printing.

Material analyses will be applied, to understand the heterogeneous nature and complex structures of CHOs, to identify the broad and varied classes of materials and to understand their degradation mechanisms over time, deriving context-dependent ageing models per material. Uni-material models will be spatiotemporally simulated, based on environmental phenomena modeling, so as to collectively render imminent degradation effects on the multi-material CHOs, enabling prediction and recreation of their future appearance, as well as automatic restoration, reaching even back to their original shape. Scan4Reco will further facilitate conservation, by indicating spots/segments of cultural objects that are in eminent conservation need and require special care, while suggestions will be provided by a dedicated Decision Support System (DSS), over conservation methods that should be followed.

All the above will be validated on real case scenarios involving heterogeneous objects of various sizes and materials, in 2 pilot real-world use cases. To enhance the accessibility of the digitized cultural objects to the scientific community, field experts and the general public, a virtual model of a museum will be launched.

1.2   *Data management plan purpose*

The purpose of this deliverable (D1.2 Data Management Plan) is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated and/or collected by the project consortium.

Scan4Reco is funded by Horizon2020 EC programme, so the Data Management Plan (DMP) must at least cover the specific aspects about project's datasets. Particularly, from [3], a DMP describes the data management life cycle for all datasets to be collected, processed or generated by a research project. It must cover:

- the handling of research data during & after the project

- what data will be collected, processed or generated

- what methodology & standards will be applied

- whether data will be shared /made open access & how

- how data will be curated & preserved

The DMP plays a crucial role in project's success for two main reasons. On the one hand, it ensures the availability and the quality of the datasets, which will be used/produced in the technical Work Packages of the project (WP3, WP4, WP5, WP6, WP7), serving the achievement of project's goals. On the other hand, it provides a thorough elaboration of the state of all datasets, and corresponding software, proving that they are assessable and useable by third parties.

1.3   *Data management roadmap*

The DMP is not a fixed document, but will evolve during the lifecycle of the project. This first version of the DMP overviews the current status of datasets and the specific conditions attached to them. Updates of the DMP will be published to revise the current status of data management, if needed, at months M12, M24 and M36. Particularly, the activities related to data management along the Scan4Reco project are planned as follows:

- M6: Preliminary analysis and production of the first version of the Data Management Plan (contained in this document);

- M12: Refined analysis based on the progress in the development of the tools and the definition of the case studies, described in the second version of the Data Management Plan;

- M24: Third version of the Data Management Plan, describing actual, proven procedures implemented by the project in the pilot demonstrators, and preparing the sustainability of the data storage after the end of the project;

- M36: Final Data Management Plan, reflecting on the lessons learnt through the project, and describing the plans implemented by Scan4Reco for sustainable storage and accessibility of the data.

2. General Principles

This DMP consists of five basic elements, namely, 1) *data collection/development methods*, 2) *standards and metadata*, 3) *legal, rights and ethical issues*, 4) *access, data sharing and reuse* and 5) *archiving and preservation*. In the remainder of this section, the general properties of these elements are discussed as applied to all members of project's consortium. While, in the next section (section 3.) these elements are discussed dataset by dataset.

2.1   *Data collection/development methods*

The data that will be managed within the Scan4Reco project mainly regard the collected data from the various sensors and the generated outcomes of the Scan4Reco algorithms. Briefly, the sensory data consist of non-invasive/non-destructive acquisitions of structural as well as chemical information of a tested CH object. Furthermore, the generated data from Scan4Reco algorithms are data mostly related to the Decision Support System and the spatiotemporal simulation procedures.

2.2   *Data Standards and metadata*

The standards used in the project are strongly related to the specific dataset, so it will be described in more detail in the Section 3.

Regarding metadata, a general condition is that they cannot be fully defined in this very initial stage of the project. However, for a few of datasets, there are abstract indications about the basic metadata elements that are to be adopted, described also in more detail in Section 3.

2.3   *Legal, rights and ethical issues*

The access and the manipulation of a cultural heritage object aiming its study, preservation or restoration is crucial the data management to comply with specific ethical guidelines and take into account intellectual property rights (IPR). As described in *D1.3 Ethical Issues Manual* [2], the main key points of IPR on CH digitation are:

- Permission from right-holders to acquire data, digitise and publish must be obtained. This permission rule is referred to as rights clearance.

- The rights of right-holders and of the data holders must be secured during the process of publication of the achieved results and the digital material.
- Digitisation is a form of reproduction that is subject to copyright restrictions.
- Online and traditional publication entails a reproduction and subsequent diffusion of digitised copyrighted content that copyright law reserves to the copyright owner.
- The permission of the rights owner must be secured before such material is digitised or made available by publication.
- The consortium will look primarily at the national legal codes to find out the copyright rules applicable to acts that they intend to perform with copyrighted works.

2.4   *Access, data sharing and reuse*

Scan4Reco participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 programme. The consortium believes firmly in the concepts of open science, and the large potential benefits the European innovation and economy can draw from allowing reusing data at a larger scale. Therefore, all data produced by the project are to be published with open access. Particularly, access is given through Creative Commons CC0 license [7] for all datasets at this project stage, unless it is defined otherwise.

All datasets will be available for sharing and re-use via the management/exploitation portals of the project. These portals will be the common means for exchanging data either among the partners of the project or between the consortium and third parties. The specifications of how and which of the portals will be used in the context of data sharing and reuse will be defined in the next months by the consortium. These specifications have to be referred to any new plugins that have to be installed to the current portal infrastructure for dealing with the different data manipulation needs.

2.5   *Data archiving and preservation*

Every dataset will be stored in hard disks provided by the partner in charge with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.).

All datasets will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion. After the project ends, all datasets will be stored in a centralized facility in order to minimize maintenance costs. Moreover, some selected representative examples shall be available via Scan4Reco community on the management/exploitation portals.

Finally, the total volume of the generated data including the redundancy storage will be about 1.2 TB.

3. Scan4Reco Datasets

All Scan4Reco partners have identified the datasets that will be collected and/or generated in the different project activities. The list is provided below, while the nature and details for each dataset are given in the next sections.

This list is indicative and allows estimating the data that Scan4Reco will produce. It may be adapted (addition/removal/modification of datasets) in the next versions of the DMP to account for the progress of the project activities.

Table 1 Dataset List

#	Dataset name
1	DS.CERTH.CHI-3DM
2	DS.CERTH.CP
3	DS.UNIVR.SAMPLES-PROFI
4	DS.UNIVR_CRS4.SAMPLES-RTI
5	DS.RFSAT.SURFACE-3D-SCANS
6	DS.OF-ADC.ACOUSTIC-MICR
7	DS.OF-ADC.IR-MAP-SPEC
8	DS.OF-ADC.RAMAN_MAP-SPEC
9	DS.OF-ADC.UV-VIS-MAP-SPEC
10	DS.OF-ADC.XRF
11	DS.OF-ADC.INFRARED-CAM

The following tables consist of the data management analysis for every identified dataset.

DS.CERTH.CHI-3DM	
Dataset Description	The data is collected using a depth/stereo sensor. Indicatively, our initial approach is using a stereo pair of RealSense F200 sensors. Each 3D model will consist of a 3D point cloud and a textured mesh model.
	These data are useful for the spatiotemporal simulation of an item's lifespan as they are necessary for the initialization of the simulation process. They will also significantly facilitate the access of the general public to the digitized versions of the items in the form of a virtual museum.
Standards & Metadata	3D models are stored using the PLY (Polygon File Format) and PCD (Point Cloud Data) formats for the textured mesh and the 3D point cloud representation respectively.
	Metadata have not been finalized yet, but will include information regarding the materials and the stratigraphy of the item. Additionally, the age and status of the item will be recorded as well as the time of the 3D model's acquisition.
Partners' Activities and Responsibilities	CERTH is the owner of the capturing device.
	CERTH is in charge of data collection/analysis/storage
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access. Additionally, a virtual museum with the digitized items will be available to the general public via a website.

Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by CERTH with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that a 30x40 cm item requires approximately 60 MB of storage, broken down to 30 MB for its point cloud, 20 MB for its textured mesh and approximately 10 MB for its metadata. Assuming that we process 100 items, the total volume of data is 120 GB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.CERTH.CP	
Dataset Description	The dataset will consist of practices for the conservation of the cultural heritage items. It will be generated from expert's knowledge and will be integrated on the decision support system in order to assist the curators/restorers.
Standards & Metadata	Data will be stored in text format and will contain a qualitative description of the cases where each conservation practice is applicable. This description will be accompanied by the cultural heritage item's type and respective measurement ranges.
Partners' Activities and Responsibilities	CERTH is in charge of data collection/analysis/storage
Access, Data Sharing and Reuse	The dataset, specific portions of it, e.g. metadata, and necessary software description are decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by CERTH with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	The volume of the generated data including the redundancy storage will be negligible (less than 50 MB) as it will consist of plain text.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.UNIVR.SAMPLES-PROFI	
Dataset Description	Data is collected using an optical scanning microprofilometer.
Standards & Metadata	Original data consists of matrices of depth values with associated pointwise measurement Signal to Noise Ratio. A specific file format has been defined for easy data exchange with a header including metadata and matrices saving uncompressed binary encodings of the measurement matrices. Metadata include information regarding • the sample (unique ID, name, materials and treatments, creators, date, ageing and treatments) • the acquisition setup (optical probe and scanning system components) • the acquisition protocol (probe CCD frequency and laser power, spot size, scanning resolution steps) • acquisition date, time and operator
Partners' Activities and Responsibilities	UNIVR is the owner of capturing devices. UNIVR is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. In addition, selected datasets could be made available to generic users for dissemination purposes.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by UNIVR with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.) Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion. On average we expect that each sample requires about 5MB. Assuming that we process 100 items, the total volume of data is 1 GB including a backup copy for each item. Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.UNIVR_CRS4.SAMPLES-RTI	
Dataset Description	The dataset is collected using a High Resolution calibrated camera, (full band or with selectable frequency filters) and multiple illumination sources with controlled types and

	directions. The imaged object is a material sample or mock-up.
	These datasets are useful for the characterization of material appearance and for modelling ageing and treatments, capturing appearance variations and micro-structural changes.
Standards & Metadata	Original data consists of stacks of RAW photographic images of the imaged samples and (eventually) of calibration targets. Derived data will be stored, as well, as appearance profiles data, including per-pixel information about illumination direction and intensity, estimated reflectance and colour components.
	A specific folder structure is designed to store raw data as well as metadata and processed information.
	Metadata include information regarding
	• the sample (unique ID, name, materials and treatments, creators, date, ageing and treatments)
	• the acquisition setup (camera and light parameters)
	• the acquisition protocol
	• acquisition date, time and operator
Partners' Activities and Responsibilities	UNIVR and CRS4 are the owners of capturing devices.
	UNIVR and CRS4 are in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access.
	Derived information in the form of relightable images (e.g., Polynomial Texture Maps) could be made available to generic users as well as other reflectance functions encodings.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by CRS4 and UNIVR with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that each sample requires up to 2 GB of storage, mainly for the storage of raw data (up to 40 MB for each photo with the NikonD810 initially chosen for the project setup). Assuming that we process 100 items, the total volume of data is 800 GB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in

	order to minimize maintenance costs.

DS.RFSAT.SURFACE-3D-SCANS	
Dataset Description	Data is to be collected using innovative time-of-flight distance measurements, as well as Pulsed Light based signal processing with various combinations of the emitter and detector pairings, LED/PIN, Laser/PIN or Laser APD. State of Art technologies and/or new developments shall be made in the project.
Standards & Metadata	Two standard 3D file formats shall be used: STL [4] and OBJ [5]. 3MF [6] printing standard shall be potentially used as a future common format The metadata content is TBD.
Partners' Activities and Responsibilities	RFSAT shall own the scanning system(s). RFSAT shall be in charge of relevant data collection, storage and analysis related to integration of partial 3D scans into a single one, including determining possible degradation of the model subject to environmental conditions.
Access, Data Sharing and Reuse	Example datasets that are to be produced shall NOT represent any specific CH object and as such shall NOT be subject to any access restrictions. As such all models, including metadata shall be subject to Open Access through Creative Commons CC0 license [7].
Archiving & Preservation (including storage/backup)	All data shall be stored on RFSAT and/or public portal of Scan4Reco project Furthermore, selected representative examples shall be available via Scan4Reco community on the ZENODO portal [8]. This will ensure that datasets will remain available to the public not only for the duration of the project, but also after its completion.

DS.OF-ADC.ACOUSTIC-MICR	

Dataset Description	The dataset is collected using a piezoelectric transducer which converts an ultrasonic wave into an electrical signal. Each collection contains the structural characteristics of a region of a few μm^2. A raster scanning from a region of interest (ROI) (a few cm^2) results in the final dataset which contains structural information from the ROI.
	The device, the acoustic microscope, consists of a transducer, electronics which excite and "listens" the transducer, an analog-to-digital converter (ADC) and computer unit.
	These data are useful to those who work on non-destructive testing applications.
	To our knowledge there are not similar data of such kind of ultrasonic wavelengths in the case of cultural heritage applications.
Standards & Metadata	The standard currently used is nearly raw raster data (NRRD data file format), but other standards will eventually be adopted.
	Metadata consist of information about the measurement parameters and about the sample
	• Dimensions of the ROI
	• ADC parameters (sampling rate, input range, time gate)
	• Time/date of the collection
	• Sample parameters (unique ID, name, materials and treatments, creators, date, ageing and treatments)
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices.
	OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that a 10x10 mm ROI requires approximately 500 MB of storage. Assuming that we process 100 ROIs, the total volume of data is 100 GB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in

	order to minimize maintenance costs.

DS.OF-ADC.IR-MAP-SPEC	
Dataset Description	The dataset is collected using basic Fourier Transform IR (FTIR) spectrometer in reflectance mode. The data is the Fourier Transforms (spectra) of the reflected infrared electromagnetic wave. Moreover, raster scanning collects a set of spectra from a ROI (mapping measurement). These datasets are useful for the characterization of material, as infrared radiation can excite the chemical bonds of the molecules.
Standards & Metadata	The standard used for the FTIR spectra is the data point table (DPT). A mapping measurement is stored in a text file which indexes the respective dpt files. Metadata consist of information about the measurement parameters and about the sample - Dimensions of the mapping measurement - Device parametrization - Time/date of the collection - Sample parameters (unique ID, name, materials and treatments, creators, date, ageing and treatments)
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices. OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.

Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that a 10x10 mm ROI requires approximately 10 MB of storage. Assuming that we process 100 ROIs, the total volume of data is 2 GB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.OF-ADC.RAMAN-MAP-SPEC	
Dataset Description	The dataset is collected using Raman spectrometer. A monochromatic laser with a nominal frequency excites an object, the portion of the electromagnetic wave that captured back on the spectrometer contains components attributed to Raman scattering, a physical phenomenon related to the type of bonds of the object. Moreover, raster scanning collects a set of spectra from a ROI (mapping measurement).
	These datasets are useful for the characterization of material, as infrared radiation can excite the chemical bonds of the molecules.
Standards & Metadata	The standard used for the Raman spectra is the data point table (DPT). A mapping measurement is stored in a text file which indexes the respective dpt files.
	Metadata consist of information about the measurement parameters and about the sample
	• Dimensions of the mapping measurement
	• Device parametrization
	• Time/date of the collection
	• Sample parameters (unique ID, name, materials and treatments, creators, date, ageing and treatments)
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices.
	OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.

Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that a 10x10 mm ROI requires approximately 10 MB of storage. Assuming that we process 100 ROIs, the total volume of data is 2 GB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.OF-ADC.UV-VIS-MAP-SPEC	
Dataset Description	The dataset is collected using UV/Vis spectrometer in reflectance mode. The data is the spectra of the reflected ultraviolet/visible electromagnetic wave. Moreover, raster scanning collects a set of spectra from a ROI (mapping measurement).
	These datasets are useful for the colour properties of a material.
Standards & Metadata	The standard used is for the UV/Vis spectra is the data point table (DPT). A mapping measurement is stored in a text file which indexes the respective dpt files.
	Metadata consist of information about the measurement parameters and about the sample
	- Dimensions of the mapping measurement - Device parametrization - Time/date of the collection - Sample parameters (unique ID, name, materials and treatments, creators, date, ageing and treatments)
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices.
	OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)

	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion. On average we expect that a 10x10 mm ROI requires approximately 10 MB of storage. Assuming that we process 100 ROIs, the total volume of data is 2 GB including a backup copy for each item. Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.OF-ADC.XRF	
Dataset Description	The dataset is collected using X-ray fluorescence (XRF) analysis. An XRF elemental analyser includes an X-ray source, the sample holder and the detector (solid state detector), which collects the secondary fluorescence radiation. The elemental analysis is based on the characteristic spectral lines of the elements that appear in specific energy values . These datasets are useful for elemental analysis of a material complementary to the IR-MAP-SPEC and RAMAN-MAP-SPEC.
Standards & Metadata	The standard used is for the XRF spectra is the text files (txt). Metadata consist of information about the measurement parameters and about the sample • Device parametrization • Time/date of the collection • Sample parameters (unique ID, name, materials and treatments, creators, date, ageing and treatments)
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices. OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.

Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that one acquisition requires approximately 0.25 MB of storage. Assuming that we process 500 acquisitions, the total volume of data is 125 MB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in order to minimize maintenance costs.

DS.OF-ADC.INFRARED-CAM	
Dataset Description	Infrared reflectographic multispectral images from 1000 up to 5000 nm using Focal Plane Array of InSb with 640x512 elements and with 90 frames per second (full frame) are collected. The area of the image depends on the lens used.
	In the case that a mosaicking procedure will be followed the dataset will be an array of images.
Standards & Metadata	The standards are used are the JPEG and PNG.
	Metadata are not finalized yet, but they are to contain information about the scanning protocol parameters as well as data/time and samples descriptions.
Partners' Activities and Responsibilities	OF-ADC is the owner of capturing devices.
	OF-ADC is in charge of data collection/analysis/storage.
Access, Data Sharing and Reuse	The portion of the dataset that is not restricted by intellectual property rights is decided to become of open access. Data management/exploitation portals are to be designed and developed to provide this access.
Archiving & Preservation (including storage/backup)	Data will be stored in hard disks provided by OF-ADC with links to the data management/exploitation portals. To avoid any data losses and ensure fast and reliable access to the data, common storage/redundancy mechanisms will be utilized (e.g. RAID-1, RAID-5 etc.)
	Data will be maintained for the entire duration of the project as well as for 2 additional years after its conclusion.
	On average we expect that a macro image requires approximately 1 MB of storage. Assuming that we process 10 x 10 images for mosaicking, the total volume of data is 100 MB including a backup copy for each item.
	Finally, after the project ends, the dataset will be stored in a centralized facility with the rest of project's datasets in

| order to minimize maintenance costs. |

4. Conclusion

The first data management analysis contained in this report allows anticipating the procedures and infrastructures to be implemented by Scan4Reco project to efficiently manage the generated and/or collected data.

This document points to present the general principles of DMP and to summarize the current state of each and every identified dataset, containing information regarding brief descriptions, standards and metadata specifications, activities and responsibilities of each partner, access, data sharing and reuse policies and archiving/preservation strategies over the respective datasets from each and every related partner.

Finally, open issues have been recognized, i.e., the role and the functionality of the management/exploitation portals as a common means for data sharing have to be specified, and the metadata specifications which cannot be explicitly defined yet. These open issues are taken into sincere consideration and they are to be addressed in the next period.

References

[1] Scan4Reco Grant Agreement Annex I – "Description of Action" (DoA)

[2] Scan4Reco D1.3 Ethical Issues Manual.

[3] European Commission, Guidelines on Data Management in Horizon 2020, v.2, 30/10/2015. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[4] The STL file format: http://www.3dsystems.com/quickparts/learning-center/what-is-stl-file

[5] Wave Front OBJ file format: https://en.wikipedia.org/wiki/Wavefront_.obj_file

[6] The 3MF consortium: http://3mf.io

[7] Creative Commons licensing: https://creativecommons.org/examples/

[8] Scan4Reco on ZENODO: https://zenodo.org/collection/user-horizon2020-reflective7-scan4reco