D7.3 Data Management Plan (2)

WP7 – Project Management

Stavros Tekes, Apostolos Agriafiotis

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730109.

Disclaimer

Copyright message

Document Information

Grant Agreement Number	**730109**	**Acronym**		**DIANA**
Full Title	Detection and Integrated Assessment of Non-authorised water Abstractions using EO			
Horizon 2020 Call	EO-1-2016: Downstream applications			
Type of Action	Innovation Action			
Start Date	1st January 2017	**Duration**		36 Months
Project URL	-			
Document URL	-			
EU Project Officer	Iulia SIMON			
Project Coordinator	Polimachi Simeonidou, Anna Osann			
Deliverable	D7.3: Data Management Plan (2)			
Work Package	WP7 – Project Management			
Date of Delivery	**Contractual**	M18	**Actual**	M18
Nature	R – Report	**Dissemination Level**		CO - Confidential
Lead Beneficiary	AgroApps			
Lead Author	Stavros Tekes	**Email**	stekes@agroapps.gr	
	AgroApps	**Phone**	+30 2310 253810	
Other authors	Apostolos Agriafiotis			
Reviewer(s)	Reviewers (Affiliation)			
Keywords	Data Management Plan			

Document History

Version	Issue Date	Stage	Changes	Contributor
0.1	27/3/2918	Draft	-	Stavros Tekes
0.2	28/3/2018	Draft	-	Apostolos Agrafiotis
0.3	30/3/2018	Final	-	Stavros Tekes

Table of Contents

Executive summary

The present document is a deliverable of the DIANA project, funded by the European Commission's Directorate – General for Research and Innovation (DG RTD), under its Horizon 2020 Innovation Action Programme (H2020).

The deliverable presents the second version of the project Data Management Plan (DMP). This second version lists the various new datasets that will be produced by the project, the main data sharing and the major management principles the project will implement around them. Thus, the deliverable includes all the significant changes, such as changes in consortium policies and any external factors that may have influenced data management in the DIANA project. It is submitted on M18 as a Mid-Term review of the DIANA Data Management Plan.

For implementing the updates, the methodology proposed by the European Commission Guidelines was adopted. For every existing and new dataset, WP leaders were conducted to provide additional information for the datasets used in the DIANA project.

The deliverable is structured in the following chapters:

- Chapter 1 includes an introduction to the deliverable.
- Chapter 2 includes the description of the datasets along with the documented changes and additional information.

1. Introduction

This document, D7.3 – Data Management Plan (2) is a deliverable of the DIANA project, which is funded by the European Union's Horizon 2020 Programme under Grant Agreement 730109. DIANA is aimed at co-designing and openly demonstrating a commercial service platform that will empower water managers and authorities to optimize the identification and inspection of non-authorised water abstractions for irrigation as well as improve their water management policies and practices, especially in extreme conditions such as drought. The scope of the project is to be co-created and defined along with the users and stakeholders so as to be shaped according to their needs and requirements.

This deliverable aims to document all the updates on the DIANA project data management life cycle for all datasets to be collected, processed or generated. A description of how the results will be shared, including access procedures and preservation according to the guidelines in Horizon 2020 projects. It is a living document and it evolves and gains more precision and substance during the lifespan of the project.

Although the DMP is being developed by AgroApps, its implementation involves all project partners' contribution. The next version of the DMP, to be published at M36, will describe more in detail the practical data management procedures implemented by the DIANA project.

2. DMP Components in DIANA

2.1 DMP Components in WP1 – Analysis and co-creation

DMP Component	Issues to be addressed
Data Summary	The purpose of the data collection is to identify and analyse the requirements of DIANA users and stakeholders of the co-creation process of DIANA tools. The development and the evaluation is one of the central objectives of the project. The following data formats will be produced: • Excel database of extended stakeholder mapping results; • Audio and video documentation of some events (2-3 files per pilot area). Data will be collected from events (mostly during meetings, but also online) of interaction with stakeholders. The data will be used in the ongoing multi-actor stakeholder process and will be confidential, to

	remain within each pilot area stakeholder community (with sharing to be provided only with the consensus by the information providers, i.e. each stakeholder).
Making data findable, including provisions for metadata	N/A (Confidential Data)
Making data openly accessible	N/A (Confidential Data)
Making data interoperable	N/A (Confidential Data)
Increase data re-use	N/A (Confidential Data)
Allocation of resources	N/A (Confidential Data)
Data security	N/A (Confidential Data)
Ethical aspects	N/A (Confidential Data)
Other issues	N/A (Confidential Data)

2.2 DMP Components in WP2 – Earth observation data products and services

DMP Component	Satellite images
Data Summary	The purpose of this data collection is to produce time series of high resolution pre-processed imagery covering the entire crop growing season. These data represent the inputs for further EO based products (vegetation indices time series, classification, evapotranspiration, crop water requirements, etc.), Following the satellite data type collected: • **Sentinel-2 data**, provided by ESA satellite Sentinel 2A and 2B; Sensor MSI, product level 2A, revisit time 5 dd with both satellites. Source: https://scihub.copernicus.eu/ Size of the data: about 50 tiles cloud free acquisitions per year per area test: L2A= 1,2 GB; 50 acquisition x 1.2GB = 60 GB per pilot. About 180 GB per 3 Pilot per year. • **Landsat-8** provided by NASA; Sensor OLI-TIRS, product level: CDR; revisit time 16 dd; Source: https://espa.cr.usgs.gov/

	Size of the data: about 20 tiles cloud free acquisitions per year per area test: CDR= 400 MB; 20 acquisition x 400 MB = 8 GB per pilot. About 24 GB per 3 Pilot per year. **Sentinel 1 mission** (S1 L1 IW GRDH mode) from the ESA Scientific Hub of Copernicus in ESA SAFE format (GeoTiff for the measurement format). These data will be used from Project partners who will use the input data for further processing.
Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by Geoserver (Catalog Services for the Web-CSW). Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata are provided). Each data produced will be identified by unique ID name file following the rule: Sensor_TYPE_Processing_Level_YYYYMMDD For example, per Sentinel 2A atmospherically corrected: S2A_L2A_20180610.tif
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and available through Geoserver's Web Mapping Service (WMS). Only web browser and Internet access will be needed to access the data. S2_L2A and Landsat8_CDR will be available for final users by DIANA Platforms and OGC WMS standards.
Making data interoperable	Data will be provided to other project partners by SFTP protocol[1].
Increase data re-use	Data will be provided to other project partners by SFTP protocol. Original data will be available by their data providers.

[1] SFTP protocol is a network protocol that provides file access, file transfer, and file management over reliable data stream.

Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID[2] technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	NDVI Time series
Data Summary	The purpose of this data collection is to produce time series of Normalized difference vegetation Index to explore crop development and phenological trends. It is also needed for further crop classification procedures. Data type and format: • **NDVI** processed from Sentinel-2 data L2A, using R code; Format: Floating point (FLT4S), GeoTIFF. Size of the data: about 50 tiles cloud free acquisitions per year per area test: NDVI= 460 MB; 50 acquisition x 460MB = 23 GB per pilot. About 70 GB per 3 Pilot per year. • **NDVI** processed from Landsat 8 data CDR, using R code; Format: Floating point (FLT4S), GeoTIFF. Size of the data: about 20 tiles cloud free acquisitions per year per area test: L1T= 244MB; 20 acquisition x 244MB = 5 GB per pilot. About 15GB per 3 Pilot per year These data will be used from: • Project partners who will use the input data for further processing; • Final users.

[2] RAID is a data storage virtualization technology that combines multiple physical disk drive components into one or more logical units for the purposes of data redundancy, performance improvement, or both.

Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by Geoserver (Catalog Services for the Web-CSW). Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata will be provided). Each data produced will be identified by unique ID name file following the rule: Sensor_NDVI_YYYYMMDD For example, per Sentinel 2A corrected: S2A_NDVI_20180610.tif
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and will be available through Geoserver's Web Mapping Service (WMS). Only web browser and Internet access will be needed to access the data. NDVI time series will be available for final users by DIANA Platforms and OGC WMS standards[3].
Making data interoperable	Data will be provided to other project partners by SFTP protocol.
Increase data re-use	Data will be provided to other project partners by SFTP protocol.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	Crop Evapotranspiration
Data Summary	The purpose of this data collection is to produce time series of Crop Evapotranspiration (monthly and daily).

[3] A Web Map Service (WMS) is a standard protocol developed by the Open Geospatial Consortium in 1999 for serving georeferenced map images over the Internet.

	These data represent the inputs for the calculation of Crop Water Requirements Data type and format: • **ETp (crop Evapotranspiration)**, produced using R code; Format: Floating point (FLT4S), GeoTIFF. Size of the data (with best resolution of 10m): about 12 maps per year per area test: ETP= 460 MB; 12 acquisition x 460MB = 5.5 GB per pilot. About 17GB per 3 Pilot per year. In case of daily evapotranspiration map storage, the total size will be about: 500 GB per year These data will be used from: • Project partners who will use the input data for further processing; • Final users.
Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by Geoserver (Catalog Services for the Web-CSW). Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata will be provided). Each data produced will be identified by unique ID name file following the rule: ETP_YYYYMM per monthly maps ETP_YYYYMMDD per daily maps
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and will be available through Geoserver's Web Mapping Service (WMS). Only web browser and Internet access will be needed to access the data. ETP time series will be available for final users by DIANA Platforms and OGC WMS standards.
Making data interoperable	Data will be provided to other project partners by SFTP protocol.

Increase data re-use	Data will be provided to other project partners by SFTP protocol.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	Crop water requirement
Data Summary	The purpose of this data collection is to produce time series of Crop water requirements (monthly and Daily). These data represent the inputs for the calculation of water abstraction and net irrigated water volume. Data type and format: • **CWR (crop water requirements)**, produced using R code; Format: Floating point (FLT4S), GeoTIFF. Size of the data (with best resolution of 10m): about 12 maps per year per area test: ETP= 460 MB; 12 acquisition x 460MB = 5.5 GB per pilot. About 17GB per 3 Pilot per year. In case of daily evapotranspiration map storage, the total size will be about: 500 GB per year. These data will be used from: • Project partners who will use the input data for further processing; • Final users.
Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by geoserver (Catalog Services for the Web-CSW). Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata are provided). Each data produced will be identified by unique ID name file following the rule:

	CWR_YYYYMM per monthly maps CWR_YYYYMMDD per daily maps
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and will be available through Geoserver's Web Mapping Service (WMS). Only web browser and Internet access will be needed to access the data. CWR time series will be available for final users by DIANA Platforms and OGC WMS standards.
Making data interoperable	Data will be provided to other project partners by SFTP protocol.
Increase data re-use	Data will be provided to other project partners by SFTP protocol.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	Field inspections
Data Summary	The purpose of these data is to provide ground true to train classification algorithms and subsequently the accuracy assessment. Data type and format: • **GT (Ground Trues)**, collected using GPS devices during into field inspection campaign. Format: ESRI shapefile. Size of the data: less then 1MB These data will be used from: • Project Partners
Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by geoserver (Catalog Services for the Web-CSW).

	Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata will be provided). Each data produced will be identified by unique ID name file following the rule: Pilot_GT_YYYYMM per monthly maps
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and will be available through Geoserver's web mapping service (WFS). Only web browser and Internet access will be needed to access the data.
Making data interoperable	Data will be provided to other project partners by SFTP protocol.
Increase data re-use	Licence will be defined through business model that will be defined during the project implementation. Data will be provided to other project partners by SFTP protocol.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	**Irrigated Area Maps**
Data Summary	The purpose of these data is to identify irrigated plots (one or twice for year). • **Irrigated area maps**, produced using image-processing techniques performed by opensource software (OTB, R code); Format: Unsigned Integer 8 bit (INT1U), GeoTIFF. Size of the data (with best resolution of 10m): about 2 maps per year per area test: Irrigated

	Area Map= 460 MB; 2 x 10MB = 20 MB per pilot. About 60MB per 3 Pilot per year. These data will be used from: • Final users
Making data findable, including provisions for metadata	Metadata will be stored in the CSW server provided by geoserver (Catalog Services for the Web-CSW). Metadata will be produced using INSPIRE Geoportal: Metadata Editor (not all the metadata will be provided). Each data produced will be identified by unique ID name file following the rule: IRRIGATEDMAP_YYYYMM per monthly maps
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the official DIANA web server and will be available through Geoserver's Web Mapping Service (WMS). Only web browser and Internet access will be needed to access the data. Irrigated Area Maps will be available for final users by DIANA Platforms and OGC WMS standards.
Making data interoperable	Data will be provided to other project partners by SFTP protocol.
Increase data re-use	Licence will be defined through business model that will be defined during the project implementation. Data will be provided to other project partners by SFTP protocol. Quality check is applied based on ground trues collected during field campaign. Global accuracy will be calculated.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

DMP Component	Meteo Data
Data Summary	The purpose of these data is to calculate main meteo variables including Evapotranspiration and Crop Water Requirements. Data will be periodically downloaded from ERA - Interim Reanalysis data (http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/) at hourly scale (analyses - four times per day, at 00:00, 06:00, 12:00 and 18:00 and forecast - from 00:00 and 12:00, with 3, 6, 9, and 12-hour steps, and more, into the future) and they will be aggregated at daily scale to perform further calculation (ETp and CWR). GRIB DATA format with spatial resolution of 0.125° x 0.125°. The provided variables are: • 2 metre dewpoint temperature [K] • 2 metre temperature [K] • 10 metre U wind component [m.s-1] • 10 metre V wind component [m.s-1] • Surface net solar radiation [J.m-2] • Surface net thermal radiation [J.m-2] • Surface pressure [Pa] • Surface solar radiation downwards [J.m-2] • Surface thermal radiation downwards [J.m-2] • Total precipitation [m] These data will be used from: • Project Partners • Final users
Making data findable, including provisions for metadata	Metadata such as creation date, version, bounding box, projection, quality of the data will be useful for discoverability of the data. Only final user and Project partner will be able to find meteo data for their plot using DIANA web Platform.
Making data openly accessible	Project partners will have permissions to access all data. Metadata and documentation will be deposited into the

	official DIANA web server and will be available through Geoserver's Web Mapping Service (WMS).
Making data interoperable	Data will be provided to other project partners by SFTP protocol.
Increase data re-use	Licence will be defined through business model that will be defined during the project implementation. Data will be provided to other project partners by SFTP protocol.
Allocation of resources	N/A
Data security	All data are stored in backup server. All server use RAID technologies which give the necessary robustness against disk failure.
Ethical aspects	N/A
Other issues	N/A

2.3 DMP Components in WP3 – DIANA service platform design and development

DMP Component	System architecture and design
Data Summary	Integration framework, component descriptions and dependencies, API descriptions, information flow diagram, internal and external interfaces, software and hardware requirements, use case scenarios and testing procedures have been described in deliverable D3.1. Overall architecture, system design and integration framework specification have also been included. These are the basis upon which the system is built.
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. This document contains a table indicating all versions of the document, along with who contributed to each version, what the changes where as well as the date a new version was created.

Making data openly accessible	The data are available in D3.1: Overall architecture, system design and integration framework. The dissemination level is confidential. It is available through dropbox for the members of the consortium.
Making data interoperable	N/A
Increase data re-use	This deliverable could be used as an example for technical teams who are interested in building this kind of systems.
Allocation of resources	N/A
Data security	All data are securely saved in the AgroApps premises and are shared with the rest of the partners using dropbox.
Ethical aspects	There are no ethical aspects related to the described dataset.
Other issues	N/A

DMP Component	Maps produced by the DIANA services
Data Summary	The DIANA platform is heavily based on map controls that will offer to the users the possibility to have access to multiple map layers indicating useful information derived by the DIANA services. Tiff files, bearing geolocation metadata, for each type of map are produced. The size of each map for each pilot area is approximately 40 – 50 MB and the size of each drought map is approximately 10MB. DIANA uses existing data such as: • Land use-land cover maps • Maps of irrigated and non-irrigated areas • Maps of soil physical properties • Crop maps
Making data findable, including provisions for metadata	Registered users are able to discover maps that offer them useful information for the completion of their duties.

	As part of every user action we produce meaningful metadata (time and date of data creation or data amendments, owners of actions that took place, service that produced the map). Metadata assist the discoverability of the data and related information.
Making data openly accessible	Only signed in users have access to the produced maps. Given that DIANA will be a paid product, users will have to pay in order to access the maps produced. The maps and the metadata are made available for use by DIANA or third-party applications through the secure API that we created. Only the authorized technical team has access to the hard drive that hosts all the produced tiff files.
Making data interoperable	N/A
Increase data re-use	Maps produced during the project, are offered to anyone who asks for it. After the end of the project, these data will only be available to users who will buy the product. Paying users will be able to download such data and use it for their own purposes.
Allocation of resources	We plan to preserve map tiff files for long time in order to offer the users the opportunity to go back in time and compare current farm conditions with those of the past.
Data security	All data generated by the platform are saved on the DIANA server. SSL connections are used to ensure secure exchange of information. In case of necessary updates, the old data will be overwritten and all actions will be audited in detail. A log is kept, containing the changed text for security reasons. Daily backups for a period of 3 days are kept. All backups are hosted on a remote server to avoid disaster scenarios. All servers are hosted behind firewalls inspecting all incoming requests against known vulnerabilities such as SQL injection, cookie tampering and cross-site scripting.

	Finally, IP restriction enforces the secure storage of data.
Ethical aspects	There are no ethical aspects related to the described dataset.
Other issues	N/A

DMP Component	Statistical reports
Data Summary	EO products and service outputs are further processed in order to produce statistical reports in the form of charts, graphs and tables. DIANA users use this information to make better decisions.
Making data findable, including provisions for metadata	Registered users are able to discover reports, charts and graphs corresponding to their area of interest and selected time period. Users are able to export the reports and use them the way they want. Metadata containing the dates the reports/charts/graphs were created, the user for whom they were produced and number of downloads are also produced.
Making data openly accessible	Only registered users have access to reports, graphs and charts. As a private product, users have to pay in order to access the maps produced. The reports and the metadata are made available for use by DIANA and third-party applications through the secure API that we created. The database that hosts the above data and any related metadata are not discoverable to other network machines operating on the same LAN, VLAN with the DB server or other networks. Therefore, only users with access to the server (DIANA technical team members) are able to discover the database.
Making data interoperable	Statistical reports are downloadable in xml format so that it is easy to be further used for other purposes.

Increase data re-use	Reports produced during the project, are offered to anyone who asks for it. After the end of the project, these data will only be available to users who buy the product. Paying users will be able to download such data and use it for their own purposes.
Allocation of resources	We plan to preserve these data for long time in order to offer the users the opportunity to go back in time and compare current farm conditions with those of the past.
Data security	All data generated by the platform are saved on the DIANA server. SSL connections are used to ensure secure exchange of information. In case of necessary updates, the old data will be overwritten and all actions will be audited in detail. A log is kept, containing the changed text for security reasons. Daily backups for a period of 3 days are kept. All backups are hosted on a remote server to avoid disaster scenarios. All servers are hosted behind firewalls inspecting all incoming requests against known vulnerabilities such as SQL injection, cookie tampering and cross-site scripting. Finally, IP restriction enforces the secure storage of data.
Ethical aspects	There are no ethical aspects related to the described dataset.
Other issues	N/A

DMP Component	Website content
Data Summary	DIANA users generate data via the platform. These data apart from the users' personal information contain the administrative boundaries and settlements of each area, maps of irrigated and non-irrigated areas and soil maps. The data described above are saved in the DIANA central database or as tiff files on the DIANA server.

	Detailed log of user actions (login, logout, account creation, visits on specific parts of the app) are kept in the form of a text file. This log will be useful for debugging purposes. Reports containing information on user devices (which browsers and mobile phones) as well as number of mobile downloads (taken from play store for android downloads and app store for mac downloads) will be useful for marketing and exploitation purposes, as well as decisions regarding the supported browsers and operating systems.
Making data findable, including provisions for metadata	As part of every user action we produce meaningful metadata (time and date of data creation or data amendments, owners of actions that took place, service that produced the map). Metadata assist the discoverability of the data and related information. The administrator of the platform is the only one with the ability to discover all data generated by the platform.
Making data openly accessible	Data are only available to registered users and administrators. The data produced by the platform cannot be shared with others without the user's permission. No open data will be created as part of DIANA. The database is only accessible by the authorized technical team.
Making data interoperable	N/A
Increase data re-use	N/A
Allocation of resources	N/A
Data security	Platform generated data are saved on the DIANA database server. Sensitive user data (emails, passwords) are encrypted using strong algorithms. SSL connections are used in order to ensure secure exchange of information.

	In case of necessary updates, the old data will be overwritten and all actions will be audited in detail. A log is kept, containing the changed text for security reasons. Daily backups for a period of 3 days are kept. All backups are hosted on a remote server to avoid disaster scenarios. All servers are hosted behind firewalls inspecting all incoming requests against known vulnerabilities such as SQL injection, cookie tampering and cross-site scripting. Finally, IP restriction enforces the secure storage of data.
Ethical aspects	There are no ethical aspects related to the described dataset.
Other issues	N/A

2.4 DMP Components in WP4 – Pilot deployment, monitoring and co-evaluation

DMP Component	Pilot implementation
Data Summary	In situ data will be collected in order to keep track of the performance and impact of the platform and its services. The data will include time series of precipitation, map of evapotranspiration, land use map, crop phenology, etc. The data will be collected by the pilot partners. The data will be available in *.xls and *.csv formats and will be used to validate the existing models and methodologies.
Making data findable, including provisions for metadata	The data will be locally hosted at DIANA server. For each dataset, related metadata will describe data structure and methodology used to collect the data.
Making data openly accessible	All DIANA in situ data will be locally hosted at DIANA server.
Making data interoperable	The data will be provided in commonly used physical units, BBHC scale of phenological development stages of a crop.

Increase data re-use	The data will be available for re-use as soon as the quality is approved. No time limits for re-use of the data will be imposed.
Allocation of resources	N/A
Data security	N/A
Ethical aspects	No personal data will be distributed within the described datasets.
Other issues	N/A

DMP Component	Co-evaluation and validation of DIANA
Data Summary	The aim of the data collection is to effectively manage ambiguity during the pilot deployment period and lead to meaningful, demand-driven improvements. The accuracy of DIANA services will be validated based on its own validation methodology. The following data formats will be produced: • *.csv • *.xlsx These data files will be useful mainly to the DIANA project partners and DIANA Advisory Board members for evaluation purposes.
Making data findable, including provisions for metadata	Task 4.3 will produce results of a blend of quantitative and qualitative techniques for validation of DIANA services.
Making data openly accessible	The validation results will be documented in reports along with recommendations about adjustments and will be made publicly available on the DIANA website; the results of the validation will be possibly published in peer-reviewed scientific journals. Public access of the raw data files will be determined later in the project life according to the business model which will be decided. DIANA partners will have full access over the raw data by authenticating themselves into the platform.

Making data interoperable	N/A
Increase data re-use	N/A
Allocation of resources	N/A
Data security	The private data (raw data) will be placed in a password area on the website or on DIANA dropbox folder, where only the project members can have access. The reports will be public, so there is no need for security measurements.
Ethical aspects	N/A
Other issues	N/A

2.5 DMP Components in WP5 – Business planning and innovation management

DMP Component	Issues to be addressed
Data Summary	**Market Intelligence Data** The purpose of the Market Intelligent Dataset is to include, document and manage all the data that will be available for and/or come out of the market analysis exercise during the DIANA project. This information is necessary to fulfil the objectives of WP5, namely: • Objective 5.1. Analyse the targeted market of the DIANA service platform and conclude on suitable business models for its commercial rollout and uptake. • Objective 5.3. Develop a business plan for the DIANA service platform to guide its launch in the market and safeguard its long-term sustainability beyond the lifespan if the project. The data has the following formats: *.xlsx, *.doc, *pdf files. Market reports and market data related to the pilots' regions and countries are in *.xlsx, *.doc, *pdf and .pptx format. In order to make the results of the market

	analysis more visible, a series of meta-data tables and figures will be produced in similar forms to inform the drafting of Business Modelling, Business Planning and Exploitation tasks. All deliverables categorised as reports and related to WP5 are expected to include data and meta-data of this kind. These documents are in *.doc and/or *.pdf format. Existing data and reports in partners' repositories may also be used or reused. Generated meta-data will be produced only within the time duration in the context of DIANA project as described in the GA and DoA. The total size of this dataset is expected to be approximately 1 Gb. This will include text and .xlsx files. These data will be useful to anyone who would like to reflect on the market intelligence in the domains relevant to the DIANA project and especially in the three pilot countries. **Contact details of external stakeholders** The purpose of the datasets is to manage and document all the data that is produced regarding the external stakeholders that have been or will be contacted and interviewed within the context of WP5 of the DIANA project. More specifically external stakeholders include: Potential lead users / early adopters, members of the Advisory Board of the project, people contacted for the Network of Interest as well as other similar ones during the pilot operation of the DIANA service platform (based on suggestions made by partners). Interviews aim either to validate and improve DIANA business models or establish direct contacts and relationships that may result in a more pro-active adoption of DIANA value propositions at an early stage among its target groups. In general, external stakeholder data is linked to the objective 5.2 (validate

	the business models based on feedback provided by pilot users and stakeholders), but may also feed in the definition of the Innovation and IPR Management Strategy and related action plans. Moreover, external stakeholder data may be collected for the purpose of creating greater buy-in for the project. The data has the following formats: *.xlsx, *.doc, *pdf files. Reports related to the external stakeholders' interviews also are in *.doc format. In order to make the results of the stakeholders' interviews more visible, a series of *.xlsx format tables and figures has been and will be produced mainly with coded recommendations. An *.xlsx table will also be created to include the list of the contact details of the external stakeholders contacted and interviewed. All deliverables of report type related to WP5 are expected to include data and meta-data of this kind. These documents will be in *.doc and/or *.pdf format. When relevant, data will be presented in an aggregated form in PowerPoint presentations within the consortium to help partners build on it for other tasks. Presentations will be in .pptx or in .pdf format. No existing data will be used or reused. Data will be produced only within the time duration of the DIANA project as described in the GA and DoA. The total file of these datasets will be approximately 1 Gb. This will include text and .xlsx files. These data will be useful to anyone who would like to reflect on novel ideas and recommendations for business modelling and value propositions of services in domains relevant to the DIANA project.
Making data findable, including provisions for metadata	A series of market intelligence metadata tables and figures will be produced to feed the drafting of Business Modelling, Business Planning and Exploitation. Unique

	and persistent identifiers will not be used for these datasets. Keyword search will be based on geographical and typological terms. Versioning will follow the versioning approach defined by the project.
Making data openly accessible	Data will be accessible in the case of public deliverables, mostly in aggregated form. The datasets will not be publicly available. This is both for privacy reasons (especially with regard to the data of external stakeholders) and to avoid disclosing commercially relevant information, thus enabling the business planning exercises and the innovation and IPR management to achieve full maturity. Business planning activities will be carried out throughout the whole duration of the project to ensure that DIANA's commercial exploitation strategy maintains a high adaptability to the information and feedback provided by the testing phase in the pilots. Thus, the data gathered to achieve this goal cannot be used for a stand-alone exercise, but should rather be regarded as a sensitive instrument that is strictly related to the relevance and effectiveness of WP5 as a whole. The datasets will only be accessible through the DIANA Dropbox and only the members of the consortium will have access to that material. The datasets will be renewed when new data will be available. The use of these datasets at the end of the project's lifecycle will make the object of an exploitation agreement that will be signed by each partner of the consortium.
Making data interoperable	N/A
Increase data re-use	The datasets cannot be accessed and there is no permission related with these datasets. Final data will be publicly available within the related deliverables

	deposited in the project website. Any individual or third party interested to access and reuse the data can download the deliverable from the project website in *.pdf format.
Allocation of resources	All costs related to the WP5 data collection and processing are covered by the project budget with dedicated person months under the WP5.
Data security	The WP5 *market data* will be preserved and shared with the members of the consortium through the DIANA Dropbox and/or Wiki. The data is collected for internal use in the project, and not intended for long-term preservation. The WP leader is keeping a quarterly backup on a separate disk. The preservation of the *contact details of the external stakeholders* that will be contacted and interviewed within the context of WP 5 will concern the entire time of the project to facilitate communication and engagement with them, following previous consent. Also, in this case, the data will be preserved and shared with the members of the consortium through the DIANA Dropbox. The data is collected for internal use in the project, and not intended for long-term preservation. The WP leader is keeping a quarterly backup on a separate disk.
Ethical aspects	To carry out the interviews with external stakeholders, the partners will employ those informed consent procedures that abide to the ethical standards and guidelines of HORIZON 2020 and in compliance with applicable international and national law. The anonymity and confidentiality of the research participants, as well as their right to turn down or withdraw from the process at any point (i.e. volunteer participation), will be guaranteed and be made clear in written (consent form) as well as in oral form where possible. Any personal information will be handled

	according to the principles laid out by the EU General Data Protection Regulation (GDPR; in force since 25 May 2016; it applies since 25 May 2018), while no sensitive personal data will be collected. More specifically, the interviewees will be provided with an 'information sheet', which will:

	• Inform them about for the aim of the project; • Outline the aim, method and implication of the interviews, the nature of the participation and any benefits, risks or discomfort that might be involved; • Explain the reason of their invitation to be interviewed; • Explicitly state that participation is voluntary and the interviewee has the right to refuse to participate and to withdraw his/her participation, at any time, without any consequences; • Explain that the results of the interviews will be used only for the purposes of the project; • Specify the duration of data storage; • Indicate the potential benefits of participating in the project, such as receiving information about the outcomes of the project, the opportunity to influence evolutions in the field as well as references to related research if they are interested; • Explain the lack of risks involved in this project other than what they would encounter in daily life; • Declare that any personal information (e.g. name, contact details, audio/video recordings, etc.) will be handled in accordance with the EU GDPR (25 May 2018), as well as the national law on data protection; • Explain the need of permission for the interview to be audio-recorded and clarify that the interviewee can revoke this permission at any time; • Clarify that the interviewer will hold all information and data collected securely and in

	confidence and that all efforts will be made to ensure that interviewees cannot be identified as a participant in the interviews (except in those cases where it might be required by law). To ensure that the interviewee has fully understood the scope and implications of his/her participation and does not feel pressured or forced to provide any information, he/she will be asked to give his/her consent in writing by signing the 'informed consent form'. If the consent cannot be given in written, for example because of illiteracy, the non-written consent must be formally documented and independently witnessed. If requested, both the "information sheet" and the "informed consent form" will be provided to the interviewees in their national language, while emphasis will be given to ensure that the terms are clear and fully understandable to them.
Other issues	N/A

2.6 DMP Components in WP6 – Enabling environment and awareness raising

DMP Component	Issues to be addressed
Data Summary	Data collection is an essential procedure for the elaboration of the enabling environment and awareness raising. Lists of communication recipients containing stakeholders and their e-mail addresses will be created and they will be in *.xls format. Information regarding direct/indirect competitors and data regarding water management authorities and farmers.
Making data findable, including provisions for metadata	All data will be publicly available through the DIANA website.

Making data openly accessible	Data concerning e-mail addresses will not be openly available because they are personal data. Any personal information will be handled according to the principles laid out by the EU General Data Protection Regulation (GDPR; in force since 25 May 2016; it applies since 25 May 2018).
Making data interoperable	N/A
Increase data re-use	N/A
Allocation of resources	This dataset does not require specific allocation of resources for its maintenance.
Data security	Automated backup of files.
Ethical aspects	N/A
Other issues	N/A

2.7 DMP Components in WP7 – Project Management

DMP Component	Issues to be addressed
Data Summary	**Contact details of project partners and advisory board** The purpose of the Databases is to manage and document all the data that will be produced regarding the project partners and Advisory Board members during the DIANA project. Advices and quality control, from the members, ensure the development and application of state-of-the-art methodologies, algorithms and generated services. The data will have the following formats: *.xlsx, *.doc, *pdf files. Reports related to the AB will also be in *.doc format. In order to make the results of the AB more visible, a series of *.xlsx format tables will be produced mainly with coded recommendations and their relation to specific WP, deliverable or responsible partner. An *.xls table will also be created to include the list of the contact details of the members.

	Thus, all deliverables of type report related to the project's management will be produced, such as Data Management Plan, financial reports and management reports. These documents will be in *.doc and/or *.pdf format. No existing data will be used or reused. Data will be produced only within the time duration in the context of DIANA project as described in the GA and DoA. The total file of this dataset will be approximately 1 Gb. This will include text and *.xlsx files. These data would be useful to anyone who would like to reflect on the impact that the AB would have to the evolution of the project itself.
Making data findable, including provisions for metadata	The inclusion of metadata for the current dataset has not been yet decided. Unique and persistent identifiers will not be used for this dataset.
Making data openly accessible	The datasets will not be publicly available. They will only be accessible through the DIANA Dropbox and only the members of the consortium will have access to that material. The databases will be renewed when new data will be available.
Making data interoperable	N/A
Increase data re-use	The dataset cannot be accessed and there is no permission related with this dataset. Final data will be publicly available within the related deliverables deposited in the project website. Any individual or third party interested to access and reuse the data can download the deliverable from the project website in *.pdf format.
Allocation of resources	All costs related to the AB data collection and processing are covered by the project budget with dedicated person months under the WP7 "Project Management".
Data security	Preserving contact details of the project partners and advisory board members for the entire time of the

	project will facilitate the internal communication. The data will be preserved and shared with the members of the consortium through the DIANA Dropbox and/ or Wiki. The data is collected for internal use in the project, and not intended for long-term preservation. The WP leader is keeping a quarterly backup on a separate disk.
Ethical aspects	N/A
Other issues	N/A

3. Conclusion

The DMP reflects the data management strategy and the procedure that DIANA will follow in order to identify issues and missing information related to data management that can be further clarified until the submission of the 3rd DMP. The DMP is not a fixed document but it will be updated once more during project lifespan (M36).

Abbreviations

AB	Advisory Board
API	Application Programming Interface
BBHC	Biologische Bundesastalt, Bundessortenamt und Chemische Industrie
CSV	Comma-seperated Values
CSW	Catalog Service for the Web
CWR	Crop Water Requirements
DoA	Description of Action
DMP	Data Management Plan
EC	European Commission
EO	Earth Observation
ESA	European Space Agency
EODC	Earth Observation Data Centre
ETp	Crop Evapotranspiration
EU	European Union
FTP	File Transfer Protocol
GA	Grant Agreement
GT	Ground Trues
IP	Internet Provider
LAN	Local Area Network
NDVI	Normalised Difference Vegetation Index
PDF	Portable Document Format
RAID	Redundant Array of Independent Disks
SQL	Structured Query Language
SSL	Secure Sockets Layers
TIFF	Tagged Image File Format
VLAN	Virtual LAN
WMS	Web Mapping Servie
WP	Work Package
XML	Extensible Markup Language