Deliverable 3.1

Data Management Plan

Project acronym:	EnviroLENS
Project title:	Copernicus for environmental law enforcement support
Project number:	821918
Instrument:	Horizon 2020
Call identifier:	H2020-SPACE-2018
Topic	DT-SPACE-01-EO-2018-2020
Type of action	Innovation action

Start date of project:	01-12-2018
Duration:	24 months

Deliverable number	D3.1
Deliverable title	Data Management Plan
Deliverable due date	31-03-2019
Lead beneficiary	GeoVille
Work package	3
Deliverable type	ORDP
Submission date:	31-03-2019
Revision:	Version 1.0

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Title:	3
Data Management Plan	
Author(s)/Organisation(s):	
Magdalena Steidl, GeoVille	
Contributor(s):	
Klemen Kenda, JSI Giorgos Mallinis, AUTH Florian Girtler, GeoVille David Kolitzus, GeoVille	

Short Description:
D3.1 Data management plan (DMP) is intended to ensure improved access and re-use of research data generated by Horizon 2020 projects. This version in the first project phase provides the initial draft which will be regularly updated over the course of the project whenever significant changes arise. As no periodic update is foreseen in the grant agreement, the DMP will be updated for the final review at the latest. In the project after-life the DMP will become a part of EnviroLENS policy.
Keywords:
Data management, Open Research Data Pilot, "FAIR" data

History:				
Version	Author(s)	Status	Comment	Date
0.1	Magdalena Steidl	Draft	First version	22 March 2019
0.2	Klemen Kenda	Draft	Input on collected and generated data of Semantic data and on ontologies	22 March 2019
0.3	Giorgos Mallinis	Draft	Input on collected and generated data of both the EO and Semantic data and on data security	28 March 2019
1.0	Magdalena Steidl David Kolitzus	Interim Document	Initial version of the DMP	29 March 2019

Review:			
Version	Reviewer	Comment	Date
1.0	Franziska Albrecht	Final deliverable ready for submission	02 April 2019

Table of Contents

List of Tables

Abbreviations

DMP	Data management plan
EC	European Commission
EO	Earth Observation
EODC	Earth Observation Data Centre
ESA	European Space Agency
EV	Essential variable
FAIR	Findable, Accessible, Interoperable and Re-usable
HR	High resolution
INSPIRE	Infrastructure for Spatial Information in Europe
IUCN	International Union for Conservation of Nature
MEA	Multilateral Environmental Agreements
OGC	Open Geospatial Consortium
ORDP	Open Research Data Pilot
SAR	Synthetic Aperture Radar
VHR	Very high resolution
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service

1 Introduction

The Data Management Plan (DMP) is considered as a key element of good data management within the EnviroLENS project. The present document provides a throughout description of the data management life cycle for the data to be collected, processed and/or generated within the project. By doing so, it follows the guidelines of the EC and applies the so called FAIR (findable, accessible, interoperable and re-usable) principle. As such it includes in information on:

- Data handling during as well as after the end of the EnviroLENS project
- Provides an overview of the data that is collected, processed and/or generated within the course of the project
- Lists the respective methodologies and standards applied
- Details whether research data will be shared and/or made openly accessible
- Explains how the data will be curated and preserved (including after the end of the project).

This version in the first project phase provides the initial draft, which will be regularly updated over the course of the project lifetime whenever significant changes arise. As no periodic update is foreseen in the grant agreement, the DMP will be updated for the final review at the latest. In the project after-life the DMP will become a part of EnviroLENS policy.

2 Data summary

2.1 Purpose of data collection and generation

"Copernicus for environmental law enforcement support" (EnviroLENS) aims to widely open the doors towards regular usage of EO technologies in the legal sector. Earth Observation (EO) systems offer significant capacities to monitor, verify and enforce compliance with legal and administrative requirements in this field. Particularly violations of environmental laws or contractual obligations, which result in physical, chemical or biological alterations of landscapes or ecosystems can be documented through EO in spatially explicit and quantitative means.

Linkage of satellite monitoring procedures with legal specifications would allow both for identification of historic violations and identification of ongoing violations along standardized routines. In addition, in some cases, the EO information can be treated as direct evidence, probative of a fact in issue, whereas under other circumstances it can provide circumstantial or other indirect evidence of facts.

2.1.1 Collected and generated data of both the EO and Semantic data

This part describes the main characteristics of the data and their provenance. The following macro-information will be managed for each data source and data set: (1) dataset reference and name, (2) data-set description, (3) data-set scope and goal.

The **Environmental legislative/law** knowledge base will consist of:

- Textual data including environmental legislative data (multi-lingual; the majority in English) to support semantic analysis and construction of eLENS Miner.

- Custom textual data, which will be provided by IUCN and DLA Piper.

- Essential and Environmental Variables to support information extraction for the use cases of the project

- Validation environmental legislative data to support testing and validation activities

- Showcase data including data regarding the specific environmental cases, which will be used for (1) validation of demonstration use cases and as (2) showcase dissemination material (exposed through the eLENS portal).

More details are shown in Table 1.

Table 1: Textual and EO data used

Data-set reference and name	Data-set description	Use of data-set
Showcase data	Data regarding specific environmental cases as defined in D7.1.	Used for (1) validation of demonstration use cases and as (2) showcase dissemination material (exposed through eLENS portal)
EUR-LEX	Legislation and judicial data. It provides EU and national EU legislation in multiple languages, annotated with EUR-VOC vocabulary. https://eur-lex.europa.eu	Used as input for the construction of eLENS Miner to support semantic analysis
Essential Variables (EV)	List of a minimal set of EO based variables related to environmental legislation enforcement over the USE-CASEs	Will be used for estimating spatial explicit information over the use cases of the project.
InforMEA	Access to information on Multilateral Environmental Agreements (MEA)	InforMEA ontology is used to extract keywords from legislation data

The **Compilation of EO datasets** to provide development, validation and demonstration data-sets includes the following information:

- EO data set for a selected project region(s) to be used during the use case development activities
- EO data set for Essential Variables modelling during the selection of the Essential Variables (EVs) list
- A set of EO data where one can experiment with eLENS services or can be used during 3rd party integration.
- Showcase EO dataset will be compiled based on the particular showcases as delivered during demonstration activities
- Auxiliary data that could be used for relevant information extraction over the EnviroLENS use cases will also be considered
- Spatial explicit information about hard infrastructures related to EnviroLENS use cases content

EO datasets will be constantly maintained and upgraded during the whole lifetime of the project. An overview of the EO data used within the project is given in Table 2.

Table 2: EO data used

Data-set reference and name	Data-set description	Use of data-set
VHR data	EO data at a very high resolution (VHR) which are provided by the EC within EnviroLENS via ESA Data Warehouse[1]	Required to provide reliable legal evidence in terms of either visually validating violations or detecting small scale violations (such as illegal dumping sites, spatially refined sea pollution, violations in agricultural practices, small scale illegal logging,…)
HR data	EO data at high resolution (HR) including optical and Synthetic Aperture Radar (SAR) data as for example Sentinel-1 / 2 / 3. For the initial service development, it is foreseen to link to the Earth Observation Data Centre (EODC)[2] and SentinelHub[3] as both the project leader GeoVille and partner Sinergise have operational access to these.	Provision of a pool of EO services as for instance: temporal historic imagery, mapping of landscape changes, detection of unusual objects, etc. Will be used also for the candidate EVs modelling and prior to the selection of the final EV list.
Protected area boundaries	Spatial data-sets showing protected areas of relevance for the use cases.	Information relevant in connection with the respective environmental law for protected areas.
Country / administrative boundaries	Spatial data-sets showing country/administrative boundaries of the relevant use cases.	Information relevant in connection with the respective environmental law for specific countries/territories
Infrastructures	Oil and gas pipeline networks, energy sector facilities, buildings for the relevant EIA use cases, road networks etc.	Required to identify the areal extent (and focus) of the EO analysis over eLENS Demonstration and show case implementation.
Auxiliary geospatial datasets	Spatial data providing auxiliary information to the analysis	Spatial explicit Information necessary for the analysis over EnviroLENS use cases (i.e. Digital Elevation Models, soil maps etc.)

---

[1] For more information, please refer to https://spacedata.copernicus.eu/

[2] https://www.eodc.eu/

[3] https://www.sentinel-hub.com/

3 FAIR data

One of the key concepts in good data management is to make data FAIR (Findable, Accessible, Interoperable and Reusable) by following the FAIR principles published in 2016[4].

3.1 Making data findable

3.1.1 Metadata provision - Catalogue data

Metadata on all available datasets (historical, current and future) and available EO services will be provided through data discovery service, which will inform users on availability of data and EO services in their target areas. Standard metadata formats will be privileged (Open Geospatial Consortium (OGC®) Catalogue Services[5], Infrastructure for Spatial Information in Europe (INSPIRE)[6] where relevant), while ad-hoc ones will be considered only if strictly necessary.

A list of metadata created will be made available in this document once the use cases and the corresponding outputs/services are defined.

3.2 Making data accessible (Open data access)

3.2.1 Data sharing policy

The data sharing policies will follow the rules of Open Research Data Pilot (ORDP), which "aims to improve and maximise the access to and re-use of research data generated by Horizon 2020 projects, taking into account the need to balance openness and protection of scientific information, commercialisation and IPR, privacy concerns, security and data management and preservation questions"[7]. Hence, it is foreseen to provide open access to research data generated through the EnvironLENS project, while at the same time considering the privacy and data ownership.

In more detail, the following principles will apply:

(1) data generated through project research activities will be openly available

(2) all the data which is available free of cost (for instance Sentinel data) will be available at the same terms also through eLENS and

(3) all of the other data will be available on the terms set by the data owner.

---

[4] https://www.force11.org/group/fairgroup/fairprinciples

[5] https://www.opengeospatial.org/standards/cat

[6] https://inspire.ec.europa.eu/

[7] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

The only exception of this rule is represented in EnviroLENS's demo region, where we will provide all data free-of-charge (following special agreements with data owners). The described principles will assure that all data, required to validate project deliveries, will be openly available.

3.2.2 Showcase repository

This will consist of showcase data demonstrating practical usage of eLENS services. DLA PIPER will contribute actual data on historical judicial processes. All data will be exposed through showcase repository. Note that, detailed information on showcase data will be provided as soon as showcases are finally defined.

3.2.3 Archiving and preservation

Rolling archive infrastructure will be implemented on cloud infrastructure. Data-driven models will be implemented on the data-input side - automatic download of Sentinel data will be provided for EnviroLENS's demonstration regions. The data will be stored for the whole subscription period. 3 months after the subscription expires, the data will be archived and then erased from the production database. Presentation data, meant to be used as a background data layer will be refreshed regularly.

Archiving and storage of additional data besides Sentinel as VHR data and in-situ data will be evaluated during the final use case definition and set-up and reported in this document as soon as further information is available.

3.3 Making data interoperable

3.3.1 Commonly used ontologies

Available legal and environmental ontologies will be analysed by JSI in order to merge them (using eLENS InfoMiner) into initial eLENS Ontology. In addition, the project will attempt to establish an EO ontology for use together with legislative texts.

We will build our work upon EUR-VOC and InforMEA ontologies, which are specific for law/environmental law. Some other potentially useful existing ontologies are: ISO-Metadata.owl, OGC: Ontology for Geography markup Language (GML3.0) of Open GIS Consortium (OGC), SWEET Ontologies (A Semantic Web for Earth and Environment Terminology)[8].

In the project, we will enrich the existing law ontologies with additional concepts (if needed). We will integrate EO related ontologies, however, the main challenge of the project will be to develop a functional mapping between law and EO ontologies, which will serve as a basis for semi-automatic rule extraction from legislative texts.

---

[8] https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library

3.3.2 Data protocols

The data protocols will be specified to be used for data exchange within the eLENS system, and between the eLENS system and the external world. Worldwide spatial standards (OGC WMS, Web Feature Service (WFS), Web Coverage Service (WCS), GeoJSON) and non-spatial (XML) standards will be used.

3.4 Making data re-usable

This section will be elaborated during the course of the project when we have more information about the data that will be made available for EnviroLENS.

4 Data security

Data security is of major importance in the EnviroLENS project. All data collected and information extracted from the analysis will be archived in a secure manner, via regular database backups of the storage facilities from each partner (local repositories backup). Database backups will be also foreseen in the project's collaborative platforms (servers).

Research data will be made available for verification and re-use unless special attention for security and confidentiality is critical as in the following cases:

- Commercial and private contracts as part of procurement documents
- Personal data that might have been collected and sharing them is not allowed by the national and European legislation.
- Confidential information as part of the negotiation of agreements

Furthermore, EnviroLENS partners will use procedures and appropriate protocols for the encryption of all internet transactions.

All partners will be encouraged to disseminate and promote the achievements of EnviroLENS project through among others scientific journal publications that allow open access, in order to make the findings accessible to others beyond the consortium.

5 Conclusion and further steps

The above sections describe the current status of the data used and generated. As mentioned in the introduction, the DMP is a living document and further details can only be added at a later stage of the project. As it is not defined in the Grant Agreement, the next update will take place at the latest for the final review of EnviroLENS.