Ref. Ares(2019)2896504 - 30/04/2019

BIOPLAT-EU

Promoting **S**ustainable **U**se of **U**nderutilised **L**ands for **B**ioenergy **P**roduction through a **W**eb-**B**ased Platform for Europe

D2.1

Report on Existing Geospatial Data for the Use in the Project

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 818083.

Project no.	818083
Project acronym:	BIOPLAT-EU
Project title:	Promoting sustainable use of underutilised lands for bioenergy production through a web-based Platform for Europe
Call:	H2020-LC-SC3-2018-RES
Start date of project:	01.11.2018
Duration:	36 months
Deliverable title:	D2.1 (Existing Geospatial Data)
Due date of deliverable:	April 2019

Organisation name of lead contractor for this deliverable: *JR* (*04*)

1. table

Name	Organization
Manuela Hirschmugl, Carina Sobe	*JR*
Marco Colangeli, Lorenzo Traverso	*FAO*
Giuseppe Pulighe	*CREA*
Alfonso Calera	*UCLM*

Dissemination level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	

History			
Version	Date	Reason	Revised by
01	24/04/2019	Draft version	All consortium
02	30/04/2019	Final version	WIP

Table of Contents

Executive Summary 5
1 Introduction 6
2 Existing Geospatial Datasets 6
2.1 LUCAS Data 6
2.2 Protected Areas 7
2.2.1 European Protected Areas 7
2.2.2 World Database of Protected Areas 8
2.3 Copernicus Land Monitoring Services 9
2.3.1 Pan-European Component: CORINE Land Cover 9
2.3.2 Pan-European Component: High Resolution Layers (HRLs) 12
2.3.3 Local Component: Urban Atlas 13
2.4 Land Parcel Identification System (LPIS) 14
2.5 European Layers of Contaminated Soils 14
2.5.1 Maps of Heavy Metals in the Soils of the EU 15
2.5.2 Heavy Metals in Topsoils 15
2.6 Marginal Lands as Mapped by SEEMLA 15
2.7 Shuttle Radar Topography Mission Digital Terrain Model (SRTM-DTM) 16
2.8 Open Street Map (OSM) 16
2.9 Climate Variables (GAEZ) 17
2.10 Soil Maps 22
2.10.1 SEEMLA SQR map 22
2.10.2 Soil maps from the European Soil Data Centre (ESDAC) 22
2.10.3 GSOCmap 23
3 Evaluation of Existing Data for Use in BIOPLAT-EU 24
3.1 LUCAS data usability 24
3.2 European protected areas usability 26
3.3 COPERNICUS Layers 27
3.3.1 CORINE Land Cover Data Usability 27
3.3.2 HRL data usability 27
3.3.3 Urban Atlas Data Usability 28
3.4 LPIS data usability 29
3.5 Usability of European Layer of Contaminated Soils 30

3.6 Usability of "SEEMLA" Marginal Land Layer 31
3.7 Usability of SRTM 31
3.8 Usability of OSM 32
3.9 Usability of GAEZ Climate Variables 33
3.10 Usability of Soil Maps 33
4 Integration of Existing Layers 34
4.1 Layers to be Directly Included 34
4.2 Layers to be Used for Exclusion 35
4.3 Layers to be Used for Training to Classify Underutilized Lands (Indirect Inclusion) . 35
4.4 Layers to be Used for STEN Modelling 35
5 Conclusions 36
6 References 37

Figures

Figure 1: GAEZ protected areas layer and European protected areas layer in comparison 9
Figure 2: Detailed Corine Land Cover Nomenclature 11
Figure 3: Seemla MagL layer superimposed on a Sentinel-2 satellite image 16
Figure 4: Thermal suitability screening procedure 21
Figure 5: Problem with location of LUCAS points 26
Figure 6: Preliminary Overall Workflow 34

Tables

Table 1. List of crops relevant for bioenergy purposes in Europe contained in GAEZ.............. 19
Table 2. List of additional crops relevant for bioenergy purposes in Europe not included in GAEZ and to be modelized in BIOPLAT-EU.. 19
Table 3: Heavy metals in soils: existing EU-wide thresholds 30

Executive Summary

This document presents a review of existing geospatial data sets available for Europe and their potential use within BIOPLAT-EU. This document includes existing EU-wide or partly even global geospatial data sets to be used within BIOPLAT-EU. As there is a large amount of different data sets available, we constrict our analysis on those, which are officially available through EU institutions such as European Environmental Agency (EEA), Joint Research Centre (JRC) or similar. Data sets on project basis are only considered as far as they are directly connected to BIOPLAT-EU content. Data sets, which are collected at a national basis, but used under a common legislation, such as the Land Parcel Information System (LPIS) or the database of protected areas are also covered in this deliverable.

In the first part of the document, we list all potential data sets and explain their main properties. These are:

- LUCAS point information data
- European and global layers of protected areas
- Copernicus layers including
    - Corine Land Cover (CLC)
    - High Resolution Layers (HRL)
    - Urban Atlas (UA)
- Land Parcel Identification System (LPIS)
- European Layers of Contaminated Soils
- Marginal Lands as Mapped by SEEMLA
- Digital Terrain Model
- Open Street Map (OSM)
- Climate Variables
- Soil quality maps

In the second part, we individually evaluate their use for BIOPLAT-EU and discuss the advantages and disadvantages of individual data sets. We evaluate individual classes against the definitions within BIOPLAT-EU and select the most appropriate data sets. This is done separately for the calculation of the TIER1 map for whole Europe and for the TIER2 maps on Hot Spot areas.

Finally, in Section 4, we evaluate, in which step of the process the individual data sets can be included. National data sets and national thresholds, which are not uniform throughout Europe by some European legislation, are considered in Deliverable D2.2. Therefore, this deliverable and Deliverable D2.2 contain complementary information.

1 Introduction

This document is a status report on existing EU-wide or partly even global geospatial data sets to be used within BIOPLAT-EU. As there is a large amount of different data sets available, we constrict our analysis on those, which are officially available through EU institutions such as European Environmental Agency (EEA), Joint Research Centre (JRC) or similar. Data sets on project basis are only considered as far as they are directly connected to BIOPLAT-EU content. Data sets, which are collected at a national basis, but used under a common legislation, such as the Land Parcel Information System (LPIS) or the database of protected areas are also covered in this deliverable.

In the first part of the document, we list all potential data sets and explain their main properties. In the second part, we evaluate individually their use for BIOPLAT-EU. Finally, in Section 4, we evaluate, in which step of the process the individual data sets can be included. There are two basic usages of existing data in the project:

1) Use of existing data to generate the map of MUC lands
   Data sets can be used for
   a) Training (see also Deliverable D2.2)
   b) Direct inclusion in the MUC map
   c) Exclusion from the MUC map
2) Use of existing data for the modelling in the STEN tool.

National data sets, which are not uniform throughout Europe by some European legislation, are considered in Deliverable D2.2. Similarly, national thresholds to be applied for contaminated lands are also collected and discussed in D2.2.

2 Existing Geospatial Datasets

2.1 LUCAS Data

LUCAS means 'Land Use/Cover Area frame statistical Survey'. It is a harmonised *in situ* (terrestrial) land cover and land use data collection procedure based on systematic regular sample with points spaced 2 km apart in the four cardinal directions. The sample grid extends over the whole European Union (EU). Each point is classified by visual image interpretation in the first phase. In the second phase, a sub-sample is drawn and the selected points are classified again in field campaigns. LUCAS is done every three years. The first one was conducted in 2011 and the last one in 2018. All details can be found on the website: https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey

The LUCAS classification has separate classification systems for land cover (LUCAS SU LC) and land use (LUCAS SU LU). The difference between land cover and land use is that land cover is the physical cover of the Earth's surface, whereas land use refers to the socio-economic function of the land. It also allows comparisons in time. The land cover nomenclature consists of 8 main categories with 29 classes and 76 subclasses. The land use nomenclature includes 4 main categories (three economic sectors and one category including unused and abandoned areas) with 16 classes and 31 subclasses.

2.2 Protected Areas

2.2.1 European Protected Areas

Currently, 21% of Europe's territory is protected. This is a significant area, which includes both marine and terrestrial areas. Considering only the terrestrial part, this area represents 25 % of the terrestrial lands in Europe, when Natura 2000 sites are included. These data sets and related descriptions are available through European Environmental Agency (EEA) at http://ec.europa.eu/environment/nature/natura2000/index_en.htm.

Within the BIOPLAT-EU consortium, there was an intensive discussion on how these areas have to be evaluated. Many of these areas can potentially be used for agricultural and/or bioenergy, as energy crop cultivation would not be forbidden. However, there is no spatial differentiation in terms of a map layer, which can be used to separate areas under strict protection from areas under "mild" protection with many usages still allowable. This kind of information is only available in the management plans. To this end, regulations foresee:

"Article 6 is one of the most important articles in the Habitats Directive as it defines how Natura 2000 sites are managed and protected. Within Natura 2000, Member States:

- *Take appropriate conservation measures to maintain and restore the habitats and species for which the site has been designated to a favorable conservation status;*
- *Avoid damaging activities that could significantly disturb these species or deteriorate the habitats of the protected species or habitat types."*

Source*:* https://natura2000.gov.mk/en/management-plans/

Management plans are typically documents of planned use and planned protection measures, some including spatial reference in terms of maps or cadastral information, many including spatial reference only in textual description. It would go far beyond the feasibility to differentiate areas according to the management plans. In addition, management plans are not even ready for all countries. In some cases, such a differentiation would even have to be done on a crop-by-crop basis, as some crops would have a damaging effect on protected species, while others would not.

2.2.2 World Database of Protected Areas

The World Database of Protected Areas Annual Release 2009 (henceforth WDPA 2009) and for the territory of the European Union the NATURA 2000 network, were applied to identify broad categories of protected areas, which are distinguished in the Global Agro-Ecological Zones (GAEZ) analysis:

1. Protected areas, where restricted agricultural use is permitted
2. Strictly protected areas, where agricultural use is not permitted.

The WDPA 2009 includes both point and polygon data. The global polygon database was used to delineate 30 arc-second grid cells (approximately 900x900 m) of protected areas in GAEZ. WDPA 2009 identifies 80,142 different mapping units (termed “Site-ids”) with associated attribute data for over 450,000 polygons. The majority of mapping units (51,556) refers to either an international or national convention. The remaining mapping units record the type of protected area, e.g. national park, natural monument, etc. (item DESIG_ENG in WDPA 2009). From these units, 77 designations were considered to be ‘strictly protected’ and therefore these categories are considered not available for agriculture. The most important designations include ‘National Parks’, ‘Forest Reserves’, ‘Zapovednik’ (a protected area in Russia which is kept "forever wild"), ‘Wildlife Management Area’, ‘Nature Park’, ‘Resource Reserve’, ‘Nature Reserve’, and ‘Game Reserve’.

The European part of the WDPA inventory does not include important protected areas for the EU 27, which are however part of the NATURA 2000 network. WPDA 2009 grid and the NATURA 2000 network information were combined to form the GAEZ protected area layer.

In an attempt to compare the two different layers, we downloaded them and showed an example in Figure 1. The blue (strictly protected) and purple (restricted use possible) areas are the GAEZ protected areas. The overlaid yellow striped polygons are the European protected areas.

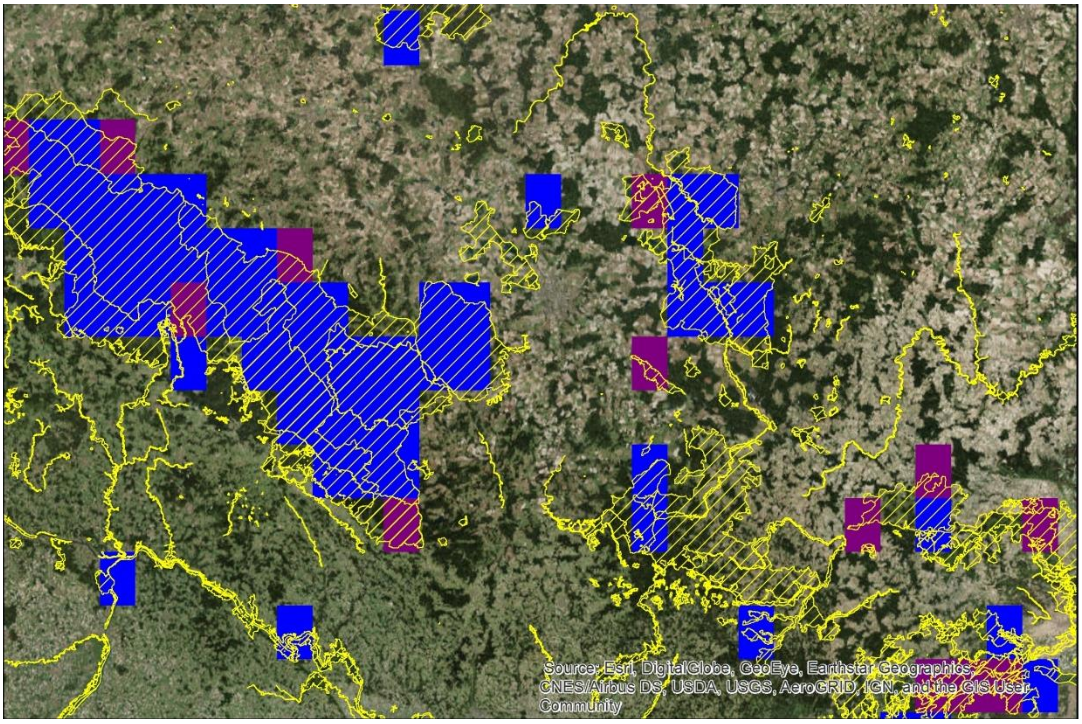

Figure 1: GAEZ protected areas layer and European protected areas layer in comparison

2.3 Copernicus Land Monitoring Services

The Copernicus Earth Observation (EO) Programme was initiated by the European Union (EU) in order to set up a system for monitoring the earth. It provides information services based on data from EO satellites and in situ data. These services are freely accessible to all users. They are divided into six thematic categories: land, marine, atmosphere, climate change, emergency management and security. Regarding BIOPLAT-EU, the Copernicus Land Monitoring Services provide valuable data sets for the generation of a database of maps of marginal, underutilized and contaminated (MUC) lands in Europe. The Land Monitoring Services are categorized in four components: Global, Pan-European, Local and Imagery and Reference Data. In the context of BIOPLAT-EU, mainly the Pan-European and the Local components are considered.

2.3.1 Pan-European Component: CORINE Land Cover

CORINE stands for 'coordination of information on the environment'. The CORINE Land Cover (CLC) inventory was initiated in 1985 (reference year 1990) to generate a standardised data collection framework on land cover information. It was a prototype project working on many different environmental issues. The project is managed and coordinated by the European Environment Agency (EEA). The CORINE database is operationally available for most areas of Europe (39 countries). The first status layer from 1990 was produced using photo-interpretation only. Updates of the inventory have been produced 2000, 2006 and 2016 by visual interpretation of high resolution satellite imagery as well as semi-automated image processing methods and integration of additional national data sets. The update for 2018 was completed in early 2019. CLC uses a Minimum Mapping Unit (MMU) of 25 hectares (ha) for areal phenomena and a minimum width of 100m for linear phenomena. The data is presented as a cartographic product, at a scale of 1:250 000. Additionally, change layers are provided which contain land cover changes greater than 5ha.

The land cover information in CLC inventory is divided into 5 main groups that are further divided into 15 classes and 44 subclasses (see Figure 2). The 5 main groups are:

- 100 artificial surfaces
- 200 agricultural areas
- 300 forests and semi-natural areas
- 400 wetlands
- 500 water bodies

All details as well as the maps are available at https://land.copernicus.eu/pan-european/corine-land-cover.

CLC Code

111: Continuous urban fabric
112: Discontinuous urban fabric
121: Industrial or commercial units
122: Road and rail networks and associated land
123: Port areas
124: Airports
131: Mineral extraction sites
132: Dump sites
133: Construction sites
141: Green urban areas
142: Sport and leisure facilities
211: Non-irrigated arable land
212: Permanently irrigated land
213: Rice fields
221: Vineyards
222: Fruit trees and berry plantations
223: Olive groves
231: Pastures
241: Annual crops associated with permanent crops
242: Complex cultivation patterns
243: Land principally occupied by agriculture, with significant areas of natural vegetation
244: Agro-forestry areas
311: Broad-leaved forest
312: Coniferous forest
313: Mixed forest
321: Natural grasslands
322: Moors and heathland
323: Sclerophyllous vegetation
324: Transitional woodland-shrub
331: Beaches, dunes, sands
332: Bare rocks
333: Sparsely vegetated areas
334: Burnt areas
335: Glaciers and perpetual snow
411: Inland marshes
412: Peat bogs
421: Salt marshes
422: Salines
423: Intertidal flats
511: Water courses
512: Water bodies
521: Coastal lagoons
522: Estuaries
523: Sea and ocean

Figure 2: Detailed Corine Land Cover Nomenclature

2.3.2 Pan-European Component: High Resolution Layers (HRLs)

The High Resolution Layers (HRLs) provide information on important types of land cover, available in raster format with a spatial resolution of 20m. Compared to CLC, the spatial detail is therefore much better. The first maps were produced for the reference year 2012. These maps have been updated for the year 2015 by using optical and radar satellite image time series, mainly from Sentinel-1 and Sentinel-2 missions. EEA is responsible for project management. The HRLs cover the whole EEA-39 area (EEA member states) and are available through Copernicus website: https://land.copernicus.eu/pan-european/high-resolution-layers.

Currently HRLs for the following five land cover classes are available:

- Impervious (sealed) surfaces (e.g. roads and built up areas)
- Forest areas
- Grassland [Langanke et al., 2018]
- Water & Wetlands
- Data on the class "small woody features".

While the Imperviousness, Forest and Water & Wetlands layers can be used to identify areas to be excluded from the MUC maps, the HRL Grassland layer could be valuable within the process of remote sensing based detection of underutilized lands within the BIOPLAT-EU project. This applies to both processing levels: the generation of a Pan-European map of MUC lands (TIER 1) and the selected hot spot areas (TIER 2). Therefore, we analyse the HRL Grassland layer in more detail below. Small woody features are currently not considered in detail, as they do not seem to have a potential use in BIOPLAT-EU.

HRL Grassland

Three different data sets are provided within the HRL Grassland product (https://land.copernicus.eu/user-corner/technical-library/hrl-grassland-technical-document-prod-2015):

1. Grassland (GRA)
2. Grassland Vegetation Probability Index (GRAVIP)
3. Ploughing Indicator (PLOUGH)

The GRA layer is a binary mask providing information on presence and absence of grassland. In addition to data from Sentinel missions, also data from the U.S Geological Survey (USGS) Landsat 8 mission was used to generate the grassland mask. The layer is provided with a MMU of 1ha. According to the product specification, the following elements are included in the grassland layer:

- Natural, semi-natural, agricultural/managed grass-covered surfaces
- Grassland with scattered trees and shrubs covering a maximum 10%
- Heathland with high grass cover, maximum of 10% non-grass cover

- Costal grassland, such as grey dunes and salt meadows located in intertidal flat areas with at least 30% graminoid species of vegetation cover
- Sparsely vegetated grasslands (>30% vegetation cover)
- Grassland in urban areas: parks, urban green spaces in residential and industrial areas
- Semi-arid steppes with scattered Artemisia scrub
- Meadows: grassland which is not regularly grazed by domestic livestock, but rather allowed to grow unchecked in order to produce hay
- Grassland in urban areas: sport fields, golf courses
- Grassland on land without use
- Natural grasslands on military sites

Elements that are not considered in the grassland layer are:

- Peat forming ecosystems dominated by sedges
- Reed beds and helophytes dominated system
- Tall forbs, fern, shrub dominated vegetation
- Grasslands that have been observed as tilled (in the reference year or a certain period before, in that case they are considered as arable fields)
- Rice fields
- Vineyards, orchards, olive groves (if more than 10% shrubs or trees)
- Tundras dominated by shrubs and lichens
- Grassland on fresh (and older) clear-cuts in the woods

The Grass Vegetation Probability Index (GRAVIP) describes the reliability of the multi-seasonal optical grassland classification for the reference year 2015 (EO data from plus/minus 1 year). It is a measure for the reliability of the grassland class assignment and indicates to which degree grassland could be separated from other vegetated land cover types. The probability index is directly related to the EO data situation: while dense time series of meaningful scenes will lead to high classification reliabilities, poor data situations will provide low probability index rates. An optimum data scenario is used to calibrate the index to the range of 1- 100%.

The PLOUGH layer provides information on the latest ploughing activities (up to 6 years). For its generation, primarily optical time series data from Landsat 8 Oli and Landsat 5 TM were used. Areas which have been ploughed within the last 6 years are removed from the grassland layer. Related data and details can be found at https://land.copernicus.eu/pan-european/high-resolution-layers/grassland/expert-products/ploughing-indicator/2015.

2.3.3 Local Component: Urban Atlas

The Urban Atlas is an initiative of the European Commission with the support of the European Space Agency (ESA) and EEA. As the local component of the Land Monitoring Services focuses on different hot spots (areas with a high disposition to specific environmental challenges and problems), the Urban Atlas does not cover all European cities, but provides harmonized land cover and land use information for a selection of 785 Functional Urban Areas (FUA). FUAs are cities and their surroundings. Data is available for the EU-28, member states of the European Free Trade Agency (EFTA) as well as Turkey and West Balkans. The first Urban Atlas was generated for the reference year 2006, an update was produced for 2012. A further update for 2018 is expected to be published in the course of 2019.

The nomenclature consists of 17 urban classes and 10 rural classes. Urban classes are mapped with an MMU of 0.25ha and rural classes with 1ha and a minimum mapping width of 10m. The maps are generated by combining digital image classification and visual interpretation of VHR satellite imagery from SPOT 5&6 and Formosat-2 pan-sharpened imagery with a spatial resolution from 2 to 2.5m. All details are available through the Copernicus geo portal: https://land.copernicus.eu/local/urban-atlas/.

2.4 Land Parcel Identification System (LPIS)

The Land Parcel Identification System (LPIS) was initiated by the EU as a key control mechanism under the Common Agricultural Policy (CAP). It is core element of the Integrated Administration and Control System (IACS) which is used for subsidy payments. LPIS is a spatial register of agricultural parcels maintaining the exact information on position, size, and unique identifier for each parcel. One of the main functionalities of the LPIS register is to prevent subsidy payments to the farmer for non-eligible areas and to avoid double payments on the same areas. LPIS is also a tool for efficient management of the utilized agricultural areas used by many different users, not only for the country administration [Pocivavsek and Ljusa, 2013]. This IT system serves to support the verification of the eligibility for area-based subsidies, monitoring farmer's cross-compliance with selected environmental rules and rural development programs [Kocur-Bera, 2019]. Based on aerial and/or satellite imagery, all member states have to map agricultural parcels and provide information about their land cover/land use characteristics. Unfortunately, not all member states provide the LPIS data for free and open to public use. Furthermore, the updating procedures are different in different countries. While some countries provide the data continuously (i.e. updates every few weeks) other countries deliver only one data set per year. Typical information for each parcel is the current (or dominant) agricultural use.

2.5 European Layers of Contaminated Soils

There are two different European-wide maps of contaminated soils available. Both are – with acquisition years 2008 and 2009 - roughly 10 years old. The two data sets are i) Maps of Heavy Metals in the Soils of the EU based on LUCAS points and ii) Heavy Metals in Topsoils based on FOREGS database (http://weppi.gtk.fi/publ/foregsatlas/).

2.5.1 Maps of Heavy Metals in the Soils of the EU

This map is based on LUCAS 2009 heavy metal (HM) data and is available at https://esdac.jrc.ec.europa.eu/content/maps-heavy-metals-soils-eu-based-lucas-2009-hm-data-0#tabs-0-description=1. It has a spatial resolution of 1x1 km and covers 27 EU member states (not including Croatia). Nine heavy metals are included: Arsenic, Cadmium, Chromium, Cobalt, Copper, Mercury, Nickel, Lead, Manganese and Antimony.

For Copper, a new map is available based on 21,682 soil samples from the LUCAS topsoil survey of 2009. It is a more detailed map than the general map of all heavy metals with a spatial resolution of 500 m. It is also available through JRC (https://esdac.jrc.ec.europa.eu/content/copper-distribution-topsoils).

2.5.2 Heavy Metals in Topsoils

This map is based on 1588 georeferenced topsoil samples from the FOREGS Geochemical database (http://weppi.gtk.fi/publ/foregsatlas/), which were acquired in 2008. The concentrations were interpolated using block regression-kriging over the 26 European countries that contributed to the database.The maps are available at a 5x5 km resolution for 8 heavy metal concentrations (Arsenic, Cadmium, Chromium, Copper, Mercury, Nickel, Lead and Zinc). The map data is available at: https://esdac.jrc.ec.europa.eu/content/heavy-metals-topsoils#tabs-0-description=1

2.6 Marginal Lands as Mapped by SEEMLA

The aim of the Horizon 2020-funded “Sustainable exploitation of biomass for bioenergy from marginal lands in Europe” (SEEMLA) project is the reliable and sustainable exploitation of biomass from marginal lands (MagL), which are used neither for food nor for feed production and are not posing an environmental threat. The SEEMLA GIS toolset includes geoprocessing models to automate and document the data management and spatial analysis for the identification and assessment of marginal lands in Europe. The SEEMLA GIS application was developed as a toolbox for ESRI ArcGIS desktop (v.10.2.2 or newer), comprising the following toolsets:

1. Identification of MagL using the Muencheberg Soil Quality Rating index (SQRI) (Mueller et al. 2007).)
2. Selection of MagL available for biomass production for bioenergy (Bioenergy Suitability)
3. MagL availability for specific bioenergy crops (Crops4MagLs).

All basics on the derivation of the products within SEEMLA are explained in detail in [Dimitriadis and Gournaris, 2017]. The basic data set to be used directly within BIOPLAT-EU is the output of point 2): Marginal land available for biomass production. An example of this map is shown in Figure 3. It is clearly visible, that the current land cover is only partly considered (i.e. cut out). Forests and settlements have been removed, but agricultural areas remain in the layer, as marginal lands can of course be used for agricultural purposes.

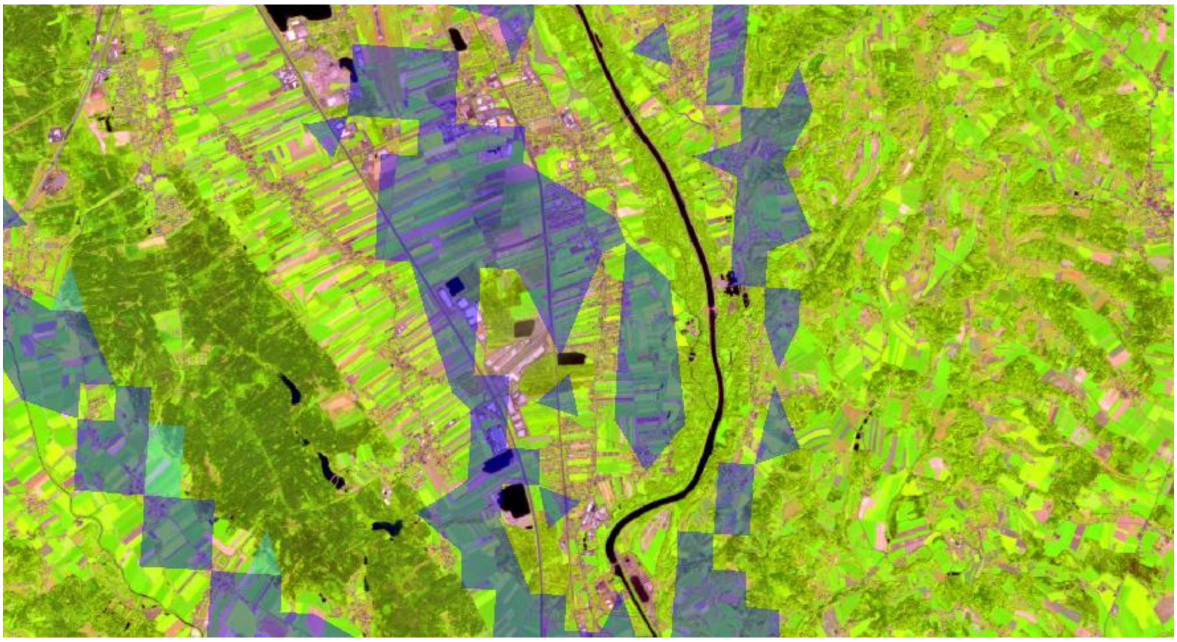

Figure 3: Seemla MagL layer superimposed on a Sentinel-2 satellite image

2.7 Shuttle Radar Topography Mission Digital Terrain Model (SRTM-DTM)

The Shuttle Radar Topography Mission (SRTM) is an international research effort that obtained digital elevation models on a near-global scale from 56°S to 60°N to generate the most complete high-resolution digital topographic database of Earth prior to the release of the ASTER GDEM in 2009. SRTM consisted of a specially modified radar system that flew on board the Space Shuttle Endeavour during the 11-day mission in February 2000. Originally, SRTM data for regions outside the United States were sampled for public release at 3 arc-seconds, which is 1/1200th of a degree of latitude and longitude, or about 90 meters. The current data have been released with a 1 arc-second, or about 30 meters sampling that reveals the full resolution of the original measurements. Data is available through various websites, e.g. through the United States Geological Survey (USGS) at https://earthexplorer.usgs.gov/.

2.8 Open Street Map (OSM)

OSM is a collaborative project, similar to Wikipedia, which was started in England in 2004 by Steve Coast. The project aims at creating and providing free geographic data. This is because many geographic data, even those available at no cost, are provided with licenses restricting the use of information. OSM's data are distributed under the license "Creative Commons Attribution-ShareAlike 2.0 license". This license allows freedom of data use, provided that any derived data are distributed under the same license [Girres and Touya, 2010]. Despite the fact that the provided data are not error free, increasing use of this data set is expected to improve their quality.

2.9 Climate Variables (GAEZ)

The Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA) have developed the Global Agro-Ecological Zones (GAEZ) model and its methodology over the past 30 years for assessing agricultural resources and potentials. The GAEZ programme (http://gaez.fao.org) aims at developing global assessments on the world's agriculture and natural resources potentials to support strategy, management, planning, rational use and sustainable development goals addressing food security and providing easy access to data and information. Therefore, since the model is based on agro-climatic variables relevant to plant production, plausibly, in the context of BIOPLAT-EU, GAEZ's climatic information could represent a useful tool for the assessment of bioenergy crop suitability and yields on marginal, underutilised and contaminated lands in Europe. Particularly, the Module I of GAEZ develops climate-related variables and indicators at grid-cell level starting from climatic data such as precipitation, temperature, wind speed, sunshine hours and relative humidity, etc. Firstly, available monthly climate data are read and converted to variables required for subsequent calculations. Temporal interpolations are used to transform monthly data to daily estimates required for characterization of thermal and soil moisture regimes. The latter includes calculation of reference potential and actual evapotranspiration through daily soil water balances (IIASA/FAO, 2012. Global Agro-ecological Zones: [Fischer et al., 2012]).

Thermal regime characterization generated in Module I includes thermal growing periods, accumulated temperature sums for average daily temperature (respectively above 0°C, 5°C and 10°C), delineation of permafrost zones and quantification of annual temperature profiles. Soil water balance calculations determine potential and actual evapotranspiration for a reference crop, length of growing period (LGP, expressed in number of days) including characterization of LGP quality, dormancy periods and cold brakes, and begin and end dates of one or more LGPs. Based on a sub- set of these indicators, a multiple-cropping zones classification is produced for rain-fed and irrigated conditions [Fischer et al., 2012].

The following eight climatic variables are converted by spline interpolation from monthly to pseudo-daily values and are measured by GAEZ:

- Minimum temperature (Tn)
- Daytime temperature (Tday)
- Maximum temperature, (Tx)
- Reference evapotranspiration (ETo)
- Wind speed at 2 m height (U2)

- Relative humidity (RH)
- Sunshine hours (n)

Other five crop specific variables are then produced and used for further computations in GAEZ modules. These are:

- Maximum evapotranspiration of 'reference' crop (ETm)
- Actual evapotranspiration of 'reference' crop (ETa)
- Water balance for 'reference' crop (Wb)
- Snow balance (Sb)
- Excess water of 'reference' crop water balance (We)

According to GAEZ's methodologies, data and information related to farm management are considered for each crop through the Land Utilization Types (LUTs). LUTs are essential parameters for any environmental, social and economic sustainability assessment of agricultural production. Differences in crop types and production systems are empirically characterized by the concept of Land Utilization Types. Therefore, a LUT consists of a set of technical specifications for crop production within a given socioeconomic setting. Attributes specific to a particular LUT include agronomic information, nature of main produce, water supply type, cultivation practices, utilization of produced, and associated crop residues and by-products. The GAEZ v3.0 framework distinguishes nearly 900 crop/LUT and management combinations, which are separately assessed for rain-fed with and without moisture conservation and irrigated conditions. These LUTs are made-up of 49 different food, feed, fiber, and bio-energy crops.

Table 1 provides a list of 9 crops currently used as feedstock for bioenergy production in Europe. These crops are included in the 49 crops (900 crop/LUTs) of GAEZ's model and their suitability and potential yields performances could be possibly exploited in the context of the BIOPLAT-EU project. Together with other project partners, FAO is exploring the process to link the existing GAEZ database and model to the Web platform that will be created in BIOPLAT-EU. Therefore, in the event that BIOPLAT-EU and GAEZ's methodologies can be aligned, the Web platform will gain from the already available information including these selected crops in the list of BIOPLAT-EU's sustainability tool (STEN).

Table 1. List of crops relevant for bioenergy purposes in Europe contained in GAEZ.

N°	GAEZ/BIOPLAT
1	Wheat
2	Maize
3	Sunflower
4	Soybean
5	Sugar beet
6	Rapeseed
7	Sorghum
8	Miscanthus
9	Switchgrass

On the other hand, there are crops (mainly lignocellulosic crops) potentially of high interest for bioenergy purposes in the EU, which are not included in the GAEZ's database as of today. These crops will be subsequently integrated in the BIOPLAT-EU web platform.

Table 2. List of additional crops relevant for bioenergy purposes in Europe not included in GAEZ and to be modelized in BIOPLAT-EU

N°	BIOPLAT
1	Smilograss
2	Camelina
3	Giant reed
4	Willow
5	Poplar
6	Eucalyptus
7	Black locust

The calculated yield of each crop/LUT is affected by water source and the intensity of management assumed to be applied. In GAEZ, three generic levels of management are defined: (i) low, intermediate, and high input level [Fischer et al., 2012].

Low level inputs

Under a low level of inputs (traditional management assumption), the farming system is largely subsistence based. Production is based on the use of traditional cultivars (if improved cultivars are used, they are treated in the same way as local cultivars), labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures.

Intermediate level inputs

Under an intermediate level of input (improved management assumption), the farming system is partly market oriented. Production for subsistence plus commercial sale is a management objective. Production is based on improved varieties, on manual labor with hand tools and/or animal traction and some mechanization. It is medium labor intensive, uses some fertilizer application and chemical pest disease and weed control, adequate fallows and some conservation measures.

High level inputs

Under a high level of input (advanced management assumption), the farming system is mainly market oriented. Commercial production is a management objective. Production is based on improved or high yielding varieties, is fully mechanized with low labor intensity and uses optimum applications of nutrients and chemical pest, disease and weed control.

In GAEZ, this variety in management and input levels is translated into yield differences by assigning different parameters for LUTs depending on the input/management level, e.g. such as harvest index and maximum leaf area index.

LUTs are parameterized to reflect environmental and eco-physiological requirements for growth and development of different crop types. Numerical values of crop parameters are varied depending on the assumed input/management level to which LUTs are subjected [Fischer et al., 2012].

There are several steps applied to test the match between thermal conditions and LUT temperature (and relative humidity) requirements: (i) Thermal (latitudinal) climatic conditions; (ii) permafrost conditions; (iii) length of temperature growing period (LGPt=5); (iv) length of frost free period (LGPt=10); (v) temperature sums (Tsumt); (vi) temperature profiles; (vii) vernalisation conditions; (viii) diurnal temperature ranges (for selected tropical perennials); and (ix) relative humidity conditions (for selected tropical perennials). LUT specific requirements are individually matched with temperature regimes (and relative humidity) prevailing in individual grid-cells. Matching is tested for the full range of possible starting dates and resulting in optimum match, sub-optimum match and not suitable conditions. The “optimum and suboptimum match categories” are considered for further biomass and yield calculations [Fischer et al., 2012]. The thermal suitability screening procedure is sketched in the following figure.

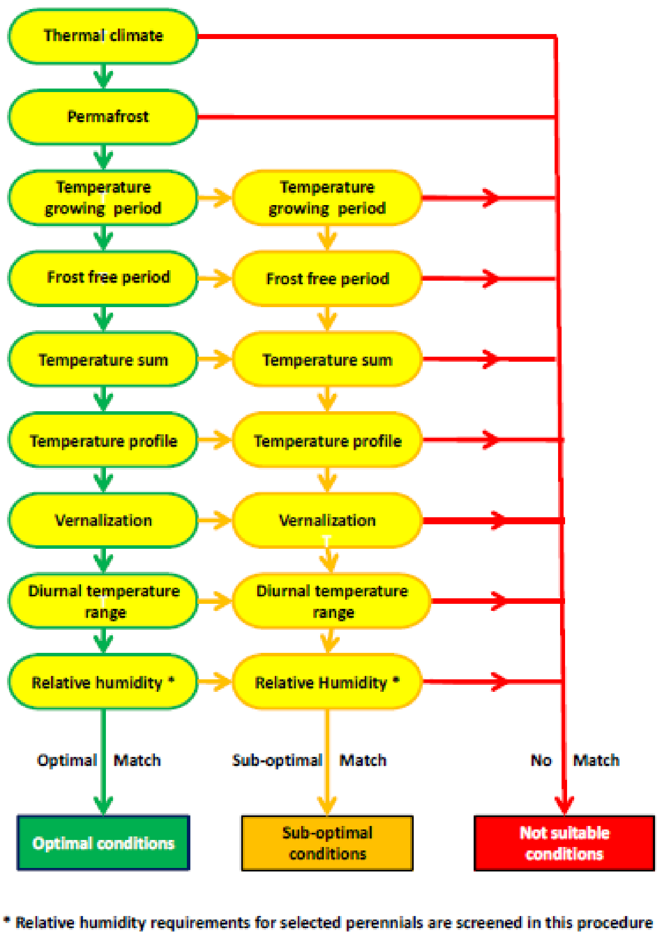

Figure 4: Thermal suitability screening procedure

2.10 Soil Maps

2.10.1 SEEMLA SQR map

The SEEMLA SQR (soil quality ratio) toolset calculates the values of the 8 basic indicators and 10 out of 12 hazard indicators based on the Muencheberg Soil Quality Rating field manual [Mueller et al., 2007]. Furthermore, the toolset includes two ancillary models (S_basic_indicator, min_hazard_indic) that are used for intermediate calculations before the final calculation of the SQR index using the SQR model. The SQR data is available as indicator for the soil quality within the given limits of the method.

2.10.2 Soil maps from the European Soil Data Centre (ESDAC)

The basic data, which also led to the SQRI, stems from JRC's European Soil Data Centre (ESDAC). There are different data sets available for Europe, which are mainly based on LUCAS 2009 soil survey with around 20,000 points for EU-25. Topsoil sampling locations were selected as to be representative of European landscape using a Latin hypercube stratified random sampling, taking into account CORINE land cover 2000, the Shuttle Radar Topography Mission (SRTM) DEM and its derived slope, aspect and curvature. Several soil properties were predicted using hybrid approaches like regression kriging. For those datasets, we predicted topsoil texture and related derived physical properties. Regression models were fitted using, along other variables, remotely sensed data coming from the MODIS sensor. The high temporal resolution of MODIS allowed detecting changes in the vegetative response due to soil properties, which can then be used to map soil features distribution.We will also discuss the prediction of intrinsically collinear variables like soil texture which required the use of models capable of dealing with multivariate constrained dependent variables like Multivariate Adaptive Regression Splines (MARS). Cross validation of the fitted models proved that the LUCAS dataset constitutes a good sample for mapping purposes leading to cross-validation R^2 between 0.47 and 0.50 for soil texture and normalized errors between 4 and 10% ([Ballabio et al., 2016], https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data).

Maps at 500m resolution include: clay, silt and salt content; coarse fragments; bulk density; USDA soil textural class; and available water capacity.

2.10.3 GSOCmap

The Global Soil Partnership at FAO has defined GSOCmap (Global Soil Organic Carbon Map) a consultative and participatory process involving 110 countries. The global soil carbon map consists of national SOC maps, developed as 1 km soil grids, covering a depth of 0-30 cm. A generic GSOC mapping guideline has been developed, which provides definitions and methodological options.

Single GSC national maps can be found at this link:

http://www.fao.org/global-soil-partnership/pillars-action/4-information-and-data-new/global-soil-organic-carbon-gsoc-map/gsocmap-contributors/en/

3 Evaluation of Existing Data for Use in BIOPLAT-EU

3.1 LUCAS data usability

As LUCAS data is a point-wise information, it can only be used for training, not for direct inclusion or exclusion in the map. The classes particularly interesting are U410 and U420, following description from LUCAS Technical reference document C3 Classification:

U410 "Abandoned Areas"

This class consists of abandoned areas with signs or structures of previous use of any kind. Areas belonging to the abandoned class are not in use and can't anymore be used for the original purpose without major reparation/renovation work.

This class includes:

- abandoned industrial areas
- abandoned transport areas (e.g. railways)
- abandoned residential areas
- abandoned residential gardens
- abandoned agricultural areas, mainly permanent crops (e.g. orchards, vineyards)
- abandoned construction sites
- abandoned mining areas

This class excludes:

- Fallow land (U112)
- Construction sites (U330)

U420 "Semi-Natural and Natural Areas not in Use"

This class includes areas which are in natural / semi-natural state and no signs of any use are visible.

This class excludes:

- Unused land formerly developed, now vacant or derelict (U410)
- Fallow land, Agricultural land which is temporarily not used (U112)
- Grazed areas (U111)

In LUCAS2015, from the interpreted 340,143 points, 60,802 were labelled as either 410 or 420. However, among these are 2,223 points, which's land cover is various types of water. Another 1,482 points are wetlands and 17,555 are points in wooded or forested areas, which also cannot be considered for BIOPLAT-EU, as clearing forest for energy crops would need to include calculation of land cover change into the assessment. The remaining points are still almost 4% of all LUCAS points. These points have to be combined with the points classified as

410 or 420 in 2018 in order to be sure to have areas not used for at least 3 years. A further combination with data from 2012 would extend this time frame to 6 years, which is similar to the definition applied within BIOPLAT-EU (five years for underutilized lands). The land cover on these points can vary significantly: from abandoned buildings and other urban structures, croplands, bare soils to shrublands and grassland.

U112 fallow land

This class is also interesting, although it is difficult to assess the time span of fallow. Often, land is left to fallow for one year or even only part of the year in the normal frame of agricultural usage. However, points, which are repeatedly (2012, 2015 and 2018) classified as fallow, have a high potential of being underutilized.

The main limitation of LUCAS is its limited spatial extent per observation. *The "point" (or basic unit of observation) is in fact a circle with a radius of 1.5m corresponding to an identifiable point on an orthophoto. As we have not only homogeneous classes that we would like to observe, for example forests (forest definition requires observing a certain area to define the crown coverage or canopy of the trees) or orchards (which may consist in more than one tree species etc.), the LUCAS observation framework also specifies an observation area, the "extended window of observation" which is the area defined by a 20m radius around the point, for specific classes"* (Source: LUCAS Technical reference document C3 Classification).

The potential use of LUCAS data for the assessment of MUC lands can be divided into two parts: First, the use of LUCAS plots for training of remote sensing data for the classification of underutilized lands; second, the use of the plots for verification purposes. Both are not possible without additional evaluation of the potential plots. If the area related to the point is very small, a direct use for training purposes is not possible. However, points, whose land use was set to 410, 420 or 112 consecutively in 2012, 2015 and 2018 and having a land cover of agriculture, shrubland, bare soil or grassland can give a hint on where to look for training areas.

The use for verification is also possible with these points, although sometimes the location accuracy is questionable. A shift of 215m was for example observed between the same point measured in 2012 and 2015 (Figure 5).

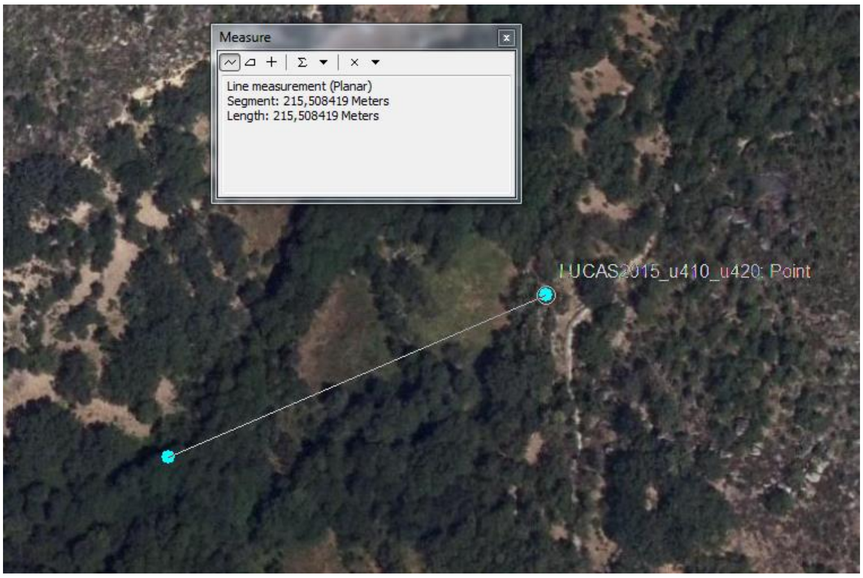

Figure 5: Problem with location of LUCAS points

In this case, it is unclear, which location is thus correct and should be used. Furthermore, in this specific case, also the small spatial extent limits the application for evaluation, because very small areas cannot be captured by remote sensing analysis.

In conclusion, the LUCAS data can be used within BIOPLAT-EU with limited applicability. The limitation is with regards to an automated input to classification or validation, which is not feasible due to various reasons explained above. However, the points can still be used within the frame of interpreted validation and help in selecting training areas, specifically in the differentiation of used and unused land. More details on how training data can be derived are given in Deliverable D2.2.

3.2 European protected areas usability

The two “protected areas” datasets are very different. While the European map has high spatial details, it lacks a subdivision in terms of what use is possible (strict protection versus restricted agricultural use). The global dataset on the other hand, provides such information at the cost of spatial detail. Based on the information given in Section 2.2, it can be stated, that for the spatially detailed European dataset, only all or none of the protected areas can be considered within BIOPLAT-EU, as a subdivision is not feasible or not even possible. Therefore, it was commonly agreed within the consortium to use this layer for exclusion. This means, we will not consider potential areas within protected sites. We are aware of the entailed error; however, we have decided that this error is less problematic for potential investors than the error of promising too much land and all related problems in setting up a project within a protected area.

3.3 COPERNICUS Layers

3.3.1 CORINE Land Cover Data Usability

The only limitation of CLC is its large MMU. However, for a European-wide approach such as the envisaged TIER1 map, land cover can be used for excluding specific areas such as settlement, water, wetlands and (potentially) forests, annual or permanent crops. Further, this information can be used in the STEN tool.

In terms of using CLC for identifying underutilized land, the classes do not directly show the existence or absence of human use of a specific parcel of land. Nonetheless, there are several CLC classes, which can be considered as "potentially underutilized". These are:

- 243: Land principally occupied by agriculture, with significant areas of natural vegetation
- 321: Natural grasslands
- 323 Sclerophyllous vegetation
- 324: Transitional woodland/shrubs
- (332: Bare rocks)
- 333: Sparsely vegetated areas

Especially the "sparsely vegetation areas" cover a significant part of Europe: for 2012, more than 228,000 km² were assigned to this class. In 2018, about 142,000 km² were assigned to this class.

Concluding, CLC data can be used for TIER1 in terms of excluding areas and in the generation of data for the STEN tool. CLC cannot be used as input data for classification or for the purpose of training process directly, however, its use can limit the geographical search area for training data generation.

3.3.2 HRL data usability

Imperviousness, Forest and Water & Wetlands layers can be used to exclude areas from the potential MUC land layer. This is necessary, if for example marginal lands are calculated based on soil quality or similar not taking into account the actual land cover. Since the STEN tool considers land use change only in specific cases, these areas have to be eliminated beforehand. The HRL layers are more detailed in terms of spatial resolution and MMU than the CLC, thus they should be preferred over CLC when possible, specifically for the TIER2 maps.

The HRL Grassland layer can be valuable within the process of remote sensing based detection of underutilized lands in the BIOPLAT-EU project. However, the fact that it is a binary mask not providing any detailed information on type of grassland limits its direct usability for the purpose of training and validation within the remote sensing based detection of underutilized areas. For BIOPLAT-EU, the PLOUGH layer, which is part of the HRL grassland layer, can be useful within the process of identifying (potentially) underutilized lands. According to the definition, underutilized land refers to agricultural land that has had no signs of human activity (including grazing) over the last five years. Thus, areas, which have been ploughed during the last five years are to be removed from the underutilized land layer.

3.3.3 Urban Atlas Data Usability

There are several classes in the UA, which are of interest for BIOPLAT-EU. Firstly, the class “13400 land without current use, which, according to the Urban Atlas Mapping Guide (2012) includes:

- *Areas in the vicinity of artificial surfaces still waiting to be used or re-used. The area is obviously in a transitional position, “waiting to be used”*
- *Waste land, removed former industry areas, (“brown fields”) gaps in between new construction areas or leftover land in the urban context (“green fields”)*
- *No actual agricultural or recreational use.*
- *No construction is visible, without maintenance, but no undisturbed fully natural or semi-natural vegetation (secondary ruderal vegetation).*
- *Also areas where the street network is already finished, but actual erection of buildings is still not visible.*
- *Excluded are “leftover areas and areas too small/narrow for any construction with regard to the MMU size”.*

The class is considered helpful to identify potentially underutilized areas for training. The direct usability of polygons for training may be limited due to the small MMU and their location in the core areas of the FUAs. These factors limit their availability due to mixed occurrence with other urban classes and the spectral influence of the surrounding urban areas, which results in an untypical spectral signal not comparable to underutilized lands located further away from human settlements.

Other classes, which are interesting for BIOPLAT-EU, are:

- 32000 Herbaceous vegetation association, which includes:
    - Vegetation cover more than 50%, ground coverage of trees with height >5 m: <30%, areas with minor / without artificial or agricultural influence;
    - Sclerophyllous vegetation;
    - Bushy sclerophyllous vegetation (e.g. maquis, garrigue);
    - Abandoned arable land with bushes;
    - Woodland degradation: storm, snow, insects or air pollution;
    - Areas under power transmission lines inside forest;
    - Fire breaks;

  - Steep bushy slopes of eroded areas;
  - Abandoned vineyards or orchards, arable land and pasture land under natural colonisation;
  - Dehesas with bush proliferation indicating no agricultural or farming use for a rather long time;
  - Bushy areas along creeks.

  Bushes, shrubs and herbaceous plants, dwarf forest in alpine or coastal regions (Pinus Mugo forests). Height is maximum 3 m in climax stage

- 33000 Open spaces with little or no vegetation, which includes:
  - Beaches, dunes, sand:
  - Bare rocks
  - Sparsely vegetated areas
  - Burnt areas
  - Snow and ice

In general, there is a time gap between the reference year of the last available Urban Atlas (2012) and the reference year of the EO data that will be used to generate the maps of underutilized lands (last five years, i.e. 2013-2019) within BIOPLAT-EU. This time gap needs to be considered. The limitation will become obsolete as soon as the UA 2018 update will be available, which is planned for 2019-2020. Concerning the direct inclusion in detailed maps for selected hot spot areas (TIER 2), Urban Atlas data will be useable, if the mapped FUAs are located within the TIER2 study area. Since UA is a local data set, no full area-coverage is possible. However, all three classes can be used, with some consideration, for training of the remote sensing approach, as soon as the 2018 UA data set is available.

3.4 LPIS data usability

Theoretically, LPIS data has a high potential to be used within BIOPLAT-EU. If all agricultural data sets are available, the areas, which have been declared as “fallow land” for the past five years form the “underutilized land”. However, there are several limitations:

1) Not all areas are within LPIS: if a land owner is not claiming subsidies for a specific parcel, it will not be included
2) If a parcel is not used, it will in many cases not qualify for subsidies and thus not be included either
3) In many countries, LPIS data is not openly and freely available
4) In those countries, where it is available, not all data for the past five years is available, e.g. for Czech Republic, the data is only dating back to 2015.

Based on these limitations, we can draw two conclusions:

First, available LPIS data can be used for generating training data for underutilized land classification. Second, if LPIS data is freely available in the case study countries (Tier2 maps) it can be used for direct inclusion into the Tier2 maps.

3.5 Usability of European Layer of Contaminated Soils

There are two issues to be considered, when using contaminated soil maps for BIOPLAT-EU:

1) The thresholds, above which soils cannot be used for food and fodder production, are difficult to obtain.
2) The existing land use on these contaminated soils.

In terms of thresholds, there is a European Directive in place (https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1986:181:0006:0012:EN:PDF), which defines limit value ranges, which are shown in Table 3. In the EU-wide mapping attempt of JRC [Tóth et al., 2016], similar values have been used (see also Table 3). In addition, we also obtained national thresholds, which are given in Deliverable D2.2.

Table 3: Heavy metals in soils: existing EU-wide thresholds

Heavy metals in soil	Tóth et al. (2016)	EU directive (range)
	threshold (mg/kg)	
Arsenic (As)	5	n/a
Cadmium (Ca)	1	1 - 3
Chromium (Cr)	100	n/a
Copper (Cu)	100	50 - 140
Mercury (Hg)	0,5	1 - 1,5
Nickel (Ni)	50	30 - 75
Lead (Pb)	60	50 - 300
Zinc (Zn)	200	150 - 300
Cobalt (Co)	20	n/a
Manganese (Mn)	n/a	n/a
Antimony (Sb)	2	n/a
Vanadium (V)	100	n/a
Molybdenum	n/a	n/a

In terms of existing land use, the issue is similar as for the marginal lands. Current land use information will be used to cut settlement, water, wetlands and forest areas from the contaminated lands layer. For agricultural use, the current land use will be stored as an attribute to be considered in the STEN tool.

3.6 Usability of "SEEMLA" Marginal Land Layer

The layer of marginal lands as generated within SEEMLA is a well-known data set of European-wide scale with homogeneous properties. It is therefore a logical dataset for the inclusion in BIOPLAT-EU. The spatial detail is limited (see Figure 3) and might therefore not be sufficient for the Hot Spot areas, but it is certainly usable for the EU-wide map (TIER1). The evaluation of the SQRI values for further use in BIOPLAT-EU is still open.

In addition to underutilized and contaminated lands, BIOPLAT-EU will focus on marginal lands that are not employed for the production of food crops. From a preliminary screening of the SEEMLA Marginal Land Layer (MagL), several lands included in the map as showing aspects of marginality, according to the SEEMLA methodology, present a cover type for food purposes. These marginal lands shall not be considered in BIOPLAT-EU. The aim of BIOPLAT-EU is to build upon experiences from similar EU projects and adapt these methodologies to the scope of our project. As a result, it is expected that the SEEMLA MagL will be used as a reference only on those marginal lands where agricultural activities for the production of food or fodder do not take place. The CORINE Land Cover Maps allows BIOPLAT-EU researchers to investigate which areas identified by SEEMLA as having characteristics of *marginality* are de facto cultivated with crops. In BIOPLAT-EU these lands will be subtracted from the layer to allow the user to run a simulation concerning a bioenergy value chain without the substitution of food crops. It remains to be calculated, how this reduction will affect the remaining areas left for bioenergy or whether this procedure will be too strict.

3.7 Usability of SRTM

A digital elevation model like the SRTM allows the calculation of height above sea level, slope inclination and aspect (slope direction). Such data is important for the calculation of suitability for different types of bioenergy crops. In fact, altitudinal ranges vary widely between EU countries. As an example, while poplar (P. tremula; P. alba) in norther Spain and Portugal does not exceed 1300 m a.s.l., in Italy could vary between 0 (sea level), 800-900 m a.s.l. in the northern part of the country (Alps), and 1500 m a.s.l. in other more suitable central zones (Appennini). Furthermore, in the context of agricultural crops, perennial lignocellulosic energy crops (e.g. Miscanthus[1]) in general are less sensitive to slope than annual crops. In fact, on high slope terrain the main feasible activity is forestry for bioenergy use or cultivation of perennial crops.

[1] The extensive root system of Miscanthus makes it suitable for stabilizing slopes or soils

In the FORBIO project we have considered the elevation criteria for land suitability mapping in the Sulcis area. The issue is related to the optimal temperatures and water availability. We followed recommendations from the literature [Angelini et al., 2009] for perennial rhizomatous grasses (i.e. giant reed, miscanthus) in the Mediterranean areas suggesting minimum precipitation at 300 mm and at least 1800 growing degree days (GDD). We used WorldClim for these constraints, as they are directly related to elevation and slope direction. For the same area, we have considered a (conservative) threshold of soil slope > 10% as unsuitable for mechanized harvesting for preventing soil erosion. The detailed technical inclusion of the SRTM remains to be specified in the STEN tool. Direct use of mean slope, aspect and height per polygon might be insufficient in case of large areas. This has to be tested within the implementation phase of the STEN tool.

3.8 Usability of OSM

OpenStreetMap is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license. It can be used as a base layer of what the visor shows, in other words, the background, so it provides a context for the visualised information.

OSM also provides some services. The routing service might be useful for the project purposes, as it would allow to calculate distances between two points, e.g. between a potential field to be used for bioenergy production and the next processing plant for the harvested material.

3.9 Usability of GAEZ Climate Variables

Provided that the coding language and the amount of data included in GAEZ is compatible with the standard of BIOPLAT-EU's Web Platform, the information available from GAEZ will be possibly integrated in the BIOPLAT-EU WebGIS application of the latter database. All the suitable methodologies concerning climate and soil variables as well as potential crop yield assessments will be selected from GAEZ, adapted and integrated into BIOPLAT-EU's web platform with the aim to expand and improve BIOPLAT-EU's platform's features. However, the statistical natures of the tools need to be aligned. In addition, in order to maximise the usability of the information in GAEZ to do so, the BIOPLAT-EU platform is being planned in a way to communicate seamlessly with the existing code of GAEZ's code. At the time of the writing of this Deliverable, the compatibility among the two platforms is to be obtained and the best strategy for the integration of the algorithms and data belonging to GAEZ to be defined.

3.10 Usability of Soil Maps

There are several soil indicators necessary for the assessment of the bioenergy crop types in STEN. The Muenchenberg's soil quality ratio index (SQRI) is a combination of different attributes. It remains to be analysed, whether the SQRI can be used directly. Definitely the attributes of the ESDAC topsoil maps such as clay, sand, etc. are needed. Although the 500m resolution is quite coarse, it should be sufficient for the estimations needed within STEN.

4 Integration of Existing Layers

The following Figure shows a simplified and preliminary workflow, on how the different layers can be used in the overall system. For readability's sake, we have not divided the workflow between TIER1 and TIER2. This system has been tested on a small test site so far and will have to be adapted during the realization in the next months.

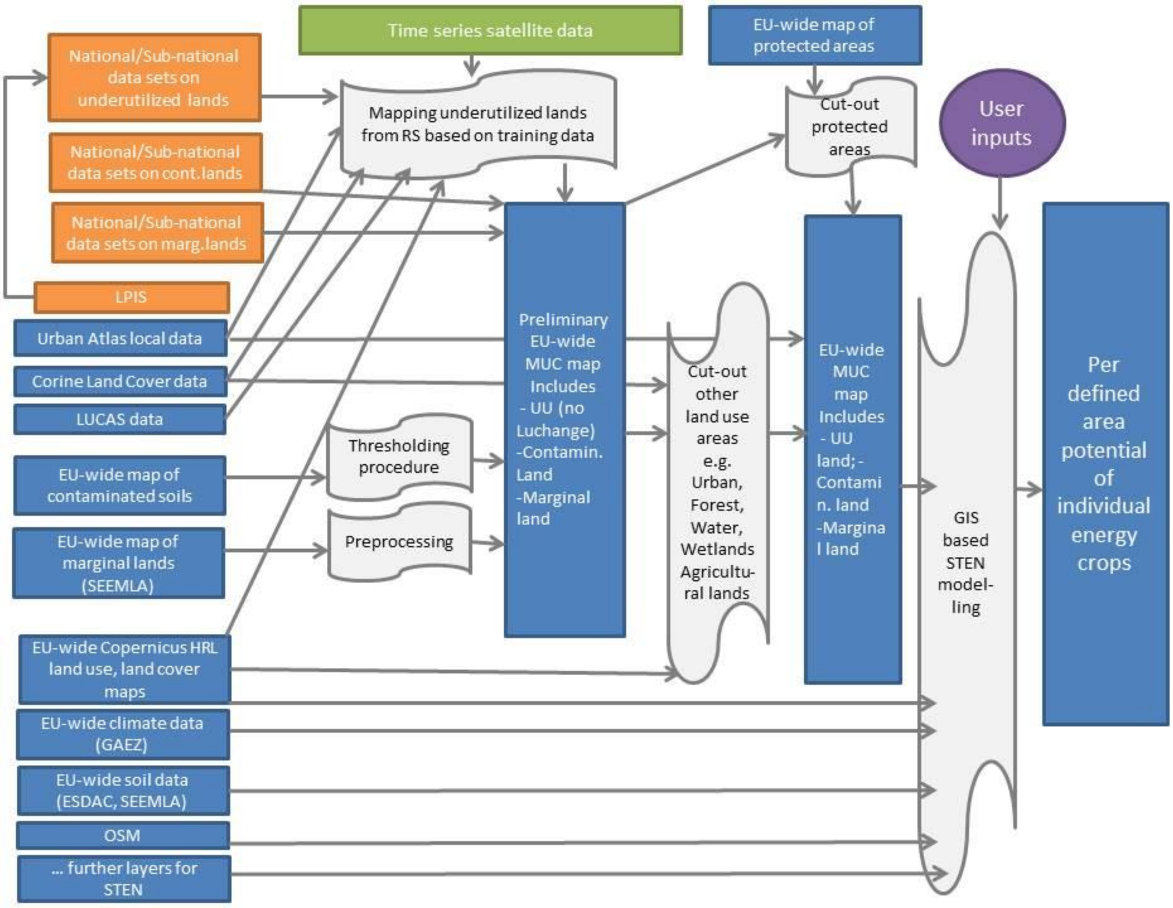

Figure 6: Preliminary Overall Workflow

4.1 Layers to be Directly Included

The following layers will be directly included in the MUC database (in TIER1 and/or TIER2 maps), partly of course with pre-processing operations. These data sets are:

- SEEMLA marginal land layer – for marginal lands

The SEEMLA map will be post-processed in order to remove strange shapes, small holes, etc.

- Contaminated soils map – for contaminated lands

In order to generate a map of contaminated lands, for each of the heavy metal contaminations, a threshold will be applied. This threshold can be an EU-wide threshold (see Table 3) or national thresholds (see Deliverable D2.2). All "thresholded" maps (together with available national or regional maps) will then be combined into one map of contaminated lands.

- Urban Atlas selected classes – for underutilized lands

Selected classes can be directly integrated into the MUC land map as soon as the UA 2018 is available.

- National and sub-national maps of underutilized, marginal and contaminated lands mainly for TIER2 maps.

4.2 Layers to be Used for Exclusion

Since the STEN tool is not foreseen to include land cover change calculations in general, areas covered by settlement, forests, water, wetlands and agricultural lands have to be excluded from the potential MUC areas. The same is true for areas under protection.

Thus, for exclusion in Tier1, we will use:

- CLC
- HRL settlements, forests, water and wetlands, grassland plough layer
- Protected areas

4.3 Layers to be Used for Training to Classify Underutilized Lands (Indirect Inclusion)

The following layers can be used for training, but only after cross-checking the data with remote sensing time series data. An example, how such a cross-checking can be undertaken, is given in Deliverable D2.2.

- HRL grassland layer including information of the plough layer (no ploughing)
- LUCAS data
- CORINE land cover data
- UA data

4.4 Layers to be Used for STEN Modelling

The following layers will be used in the modelling approach:

- CORINE Land Cover data for the assessment of existing land cover on marginal and contaminated lands
- SRTM for the assessment of elevation (height) and slope
- GAEZ variables for the modelling of the climate suitability of different bioenergy crops
- Soil maps for the modelling of the soil suitability of different bioenergy crops

There are basically two different ways to integrate data into the STEN tool: first, as permanent attributes and second, as flexible/on-demand attributes. Permanent attributes should be generated, if the content of the attribute is expected to be stable over a longer period of time, one example would be the elevation or slope of an area, which typically does not change. Such attribute can be stored in the shapefile directly and can be accessed fast. On the other hand, updates are difficult to implement. The second category consists of flexible attributes, which are extracted on-demand from an existing (on-line) layer. An example for this would be e.g. meteorological data, which change rapidly. The ratio between permanent and on-demand attributes has to be balanced, as too many on-demand attributes would slow down the overall processing. Therefore, this ratio has to be taken into account for the final design.

5 Conclusions

In conclusion, there is already a lot of data available, which can be built upon. This deliverable forms the basis for generating the MUC map on the one and for setting up the GIS based web-mapping STEN tool on the other hand. The chapter on usability shall be considered as theoretical use potential. Only with the practical implementation, the practical usability of the different data sets can be finally assessed.

6 References

[Angelini et al., 2009] Angelini, L. G., Ceccarini, L., o Di Nasso, N. N., and Bonari, E. (2009). Comparison of arundo donax l. and miscanthus x giganteus in a long-term field experiment in central italy: Analysis of productive characteristics and energy balance. *Biomass and Bioenergy*, 33(4):635–643.

[Ballabio et al., 2016] Ballabio, C., Panagos, P., and Monatanarella, L. (2016). Mapping topsoil physical properties at european scale using the LUCAS database. *Geoderma*, 261:110–123.

[Dimitriadis and Gournaris, 2017] Dimitriadis, E. and Gournaris, N. (2017). SEEMLA - Sustainable exploitation of biomass for bioenergy from marginal lands in Europe.

[Fischer et al., 2012] Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Tóth, G., van Velthuizen, H., Verelst, L., and Wiberg, D. (2012). *Global Agro-Ecological Zones (GAEZ v3.0) - Model Documentation*.

[Girres and Touya, 2010] Girres, J.-F. and Touya, G. (2010). Quality assessment of the french OpenStreetMap dataset. *Transactions in GIS*, 14(4):435–459.

[Kocur-Bera, 2019] Kocur-Bera, K. (2019). Data compatibility between the land and building cadaster (LBC) and the land parcel identification system (LPIS) in the context of area-based payments: A case study in the polish region of warmia and mazury. *Land Use Policy*, 80:370–379.

[Langanke et al., 2018] Langanke, T., Richter, R., Sandow, C., Storch, C., Desclée, B., and Moran, A. (2018). Copernicus Land Monitoring Service –High Resolution Layer Grassland: Product Specifications Document.

[Mueller et al., 2007] Mueller, L., Schindler, U., Behrendt, A., Eulenstein, F., and Dannowski, R. (2007). *The Muencheberg Soil Quality Rating (SQR) - FIELD MANUAL FOR DETECTING AND ASSESSING PROPERTIES AND LIMITATIONS OF SOILS FOR CROPPING AND GRAZING*.

[Pocivavsek and Ljusa, 2013] Pocivavsek, G. and Ljusa, M. (2013). Characteristics of the land parcel identification system (LPIS) as the main subcomponent of the agriculture information system.

[Tóth et al., 2016] Tóth, G., Hermann, T., Silva, M. D., and Montanarella, L. (2016). Heavy metals in agricultural soils of the european union with implications for food safety. *Environment International*, 88:299–309.