NEW METHODS IN ACQUISITION, UPDATE AND DISSEMINATION OF NATURE CONSERVATION GEODATA – IMPLEMENTATION OF AN INTEGRATED FRAMEWORK

G. Tintrup gen. Suntrup [a, *], T. Jalke [a], L. Streib [a], N. Keck [a], S. Nieland [b], N. Moran [b], B. Kleinschmit [b], M. Trapp [a]

[a] Dept. Environmental Systems, RLP AgroScience - Institute for Agroecology, Breitenweg 71 67435 Neustadt, Germany
gregor.tintrup@agroscience.rlp.de

[b] Geoinformation in Environmental Planning Lab, Technical University of Berlin, 10623 Berlin, Germany

KEY WORDS: automated processing, interoperability, nature conservation, INSPIRE, remote sensing, land cover classification, biodiversity monitoring, ontology

ABSTRACT:

Within the framework of this project methods are being tested and implemented a) to introduce remote sensing based approaches into the existing process of biotope mapping and b) to develop a framework serving the multiple requirements arising from different users' backgrounds and thus the need for comprehensive data interoperability. Therefore state-wide high resolution land cover vector-data have been generated in an automated object oriented workflow based on aerial imagery and a normalised digital surface models.These data have been enriched by an extensive characterisation of the individual objects by e.g. site specific, contextual or spectral parameters utilising multitemporal satellite images, DEM-derivatives and multiple relevant geo-data. Parameters are tested on relevance in regard to the classification process using different data mining approaches and have been used to formalise categories of the European nature information system (EUNIS) in a semantic framework. The Classification will be realised by ontology-based reasoning. Dissemination and storage of data is developed fully INSPIRE-compatible and facilitated via a web portal. Main objectives of the project are a) maximum exploitation of existing "standard" data provided by state authorities, b) combination of these data with satellite imagery (Copernicus), c) create land cover objects and achieve data interoperability through low number of classes but comprehensive characterisation and d) implement algorithms and methods suitable for automated processing on large scales.

1. INTRODUCTION AND BACKGROUND

1.1 Situation

A key task of the authorities in the German federal state of Rhineland-Palatinate is the provision and regular update of state-wide geo-data on ecologically valuable areas. These data serve for multiple purposes ranging from local to EU level, e.g. local and regional administration, biotope management, biodiversity monitoring or EU report obligations (NATURA 2000 ((Council Directive) 92/43/EEC, 1992), CAP-Cross Compliance (Eu, 2013)). Regular state-wide field recordings are expensive and time consuming. Remote sensing and innovative technologies in data management and -analysis offer new opportunities for an increased efficiency (Corbane et al., 2015). The development in other countries show (Banko et al., 2012) that such technologies can help to increase the usability of the data produced and introduce geo-data to new fields of administration technology. Overall guideline is to step by step introduce automated processes into traditional ways of geo-data acquisition, e.g. the field-mapping-based workflows of habitat mapping.

1.2 Project Overview

Since mid 2014 the project NATFLO (Landscape Objects from remotely sensed data for Nature Conservancy) is setting up a system of data production for nature conservation purposes based on remote sensing methods and the extended use of automated workflows. The project makes use of the experiences gained in previous studies and the results of expert meetings held regularly from 2010 until present.
The purpose of the meetings was to develop the basic methodological background for the production of multifunctional geo-data with special regard to biotope mapping. Some major conclusions were drawn guiding the following project work.

1.2.1 High resolution vector data
Since the data were to be used in administrational contexts an object based mapping approach (vector geometries) was preferred allowing the attachment of additional information in attributes and data bases. A very high spatial resolution was regarded to be adequate being able to capture the landscape in detail (e.g. individual trees or small scale gradual differences in vegetation cover) and to facilitate the combination with cadastral data.

1.2.2 Multifunctionality
Although biotope mapping was to be the main purpose of the envisaged process further fields of application like land use mapping or landscape planning were identified. On top of this both the local habitat classification scheme OSIRIS and EU-schemes (e.g. EUNIS habitat types) had to be addressed. Therefore a high degree of interoperability, i.e. the usability of one vector object in different semantical contexts was identified as a major aim. To achieve semantic interoperability ontology based reasoning using object properties was recognized to be

---

* Corresponding author.

the appropriate means rather than the direct implementation of classification schemes and a subsequent translation to other nomenclatures. This approach adopts the ideas published in the concepts of Einonet Group on Land Monitoting in Europe (EAGLE) (Arnold et al., 2013).

1.2.3 State wide approach vs. test regions

As mentioned before the project can take advantage of the experiences made in previous studies. It is due to these experiences that it was decided to run a comprehensive state wide approach instead of developing a method in test regions. For example automated workflows for state wide object based image analysis did already exist and could be adjusted to a new subject. In an iterative way, the project is addressing one question after the other. The advantage of such an approach is that although a system is not yet fulfilling the full range of tasks useful (geo-) information is dripping from it and can already be transferred to official processes because the data exist for the entire country.

### 1.3 Requirements to be met by data and framework

In short, the data and their generation process described below will have to fulfil different requirements and serve for quite a number of tasks:

- Deliver a comprehensive state-wide coverage of data
- Run regular updates
- Offer objects at a high spatial resolution able to be used in administrational processes, e.g. in combination with cadastral information.
- Provide objects suitable for different nomenclatures and thematical backgrounds in a standardised way (semantic and syntactic interoperability).
- Fulfil requirements of EU-related contexts (INSPIRE, EEA, NATURA2000)
- Serve as input data for eGovernmental adminstration
- Make use of official data base in Rhineland-Palatinate
- High degree of automation in the involved steps of data pre-processing and generation, data exchange, semantic categorisation and the dissemination of information (e.g. maps) and physical data.

1.4 Study area

The German federal state Rhineland Palatinate (RLP) is situated in the south-west of Germany, Central Europe, and covers an area of 19,854 km². In the West, RLP shares borders with Belgium, Luxembourg and France while the rest of the adjacent territories are those of German federal states (North-Rhine-Westphalia, Hessen, Baden-Württemberg). Except in the mainly flat regions of the Upper Rhine Valley, a rift zone in its south-west, Rhineland-Palatinate is a very hilly area. The mountain ranges of Eifel, Westerwald, Hunsrück and Pfälzerwald/ Pfälzer Bergland are subdivided by an extensive, deeply incised network of permanent water bodies, especially the valleys of several greater rivers like Mosel, Saar and Nahe. Altitude ranges from about 100 to 816 masl. In general RLP is characterised by a western european atlantic climate (Cfb) with mild winters and temperate summers with high precipitation values. However, due to its diverse topography there are great differences in the climatic conditions on local as well as

regional scale, e.g. leading to great regional shifts in the onset of the vegetation period. 42 % of the area is covered by forest (beech, oak, spruce, pine), 42 % is under agricultural use. Agriculture in the higher altitudes is dominated by meadowland with arable land in some places, whereby arable land dominates in the mid to lower altitudes down to River Rhine. Viticulture with an area of about 63,000 ha and hot-spots along the rivers Mosel and Rhine is a strong economic and cultural factor. Due to its geographical diversity, the significant areas of forest and meadowland and the dense network of permanent water bodies Rhineland-Palatinate is rich in ecologically valuable areas.
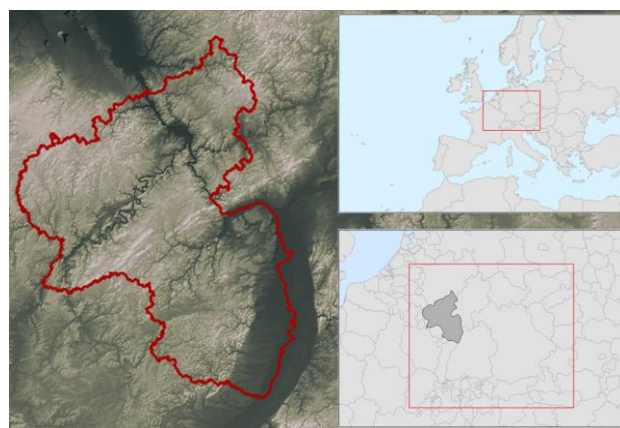
Figure 1: Rhineland-Palatinate, terrain and situation (Illustr. by author)

2. MATERIAL AND METHODS

2.1 Major Data

To keep costs low one of the preconditions of the project was to derive as much information as possible from standard geo-data provided by the RLP authorities.

2.1.1 Aerial Images

A full coverage of multispectral orthophotos (B, G, R, NIR) have been provided by the RLP Ordnance Survey (Landesamt für Vermessung und Geobasisinformation RLP) with a ground resolution of 0.2 m in tiles of 2 by 2 km. Aerial images in RLP are updated every two years.

2.1.2 Stereo Matching Based DSM

DSM as ASCII point data with a resolution of 0.5 m from automated stereo matching have been provided by the RLP Ordnance Survey. Orthophotos and DSM are derived from the same aerial imagery. The point data was rasterised via IDW-algorithms in automated workflows.

2.1.3 LiDAR DTM

LiDAR DEM as ASCII point clouds (first and last pulse) have been provided by RLP Ordnance Survey. The data have been acquired during flight campaigns between 2003 and 2009. Last pulse data with 4 points (average) per square meter were rasterised by IDW-algorithms in automated workflows. These data served for the calculation of normalised DSM with a ground resolution as well as for the derivation of DTM of lower resolution e.g. for terrain analyses.

2.2 Methods - General approach

2.2.1 From Raw Data to Landcover to Habitats

In order to accomplish the challenge of developing a multifunctional data infrastructure and corresponding data model the project uses existing concepts of categorising and describing data like the EAGLE data model (Arnold 2013). Main advantages of this approach are a comprehensive interoperability to other nomenclatures (CORINE Land Cover, INSPIRE LCUS, LCCS etc.) and the ability to derive land cover and nature conservation products in different stages of the workflow. In the first stage a pure land cover product will be generated using the descriptive attributes of the EAGLE data model whereas in the second stage these objects can be enriched by indicators that are necessary to derive the actual biotope map.

2.2.2 Combined use of imagery from different sources

The workflow developed is based on the combined analysis of different data sources including aerial images (biennial update), VHR-DEM (LiDAR, stereo-matching) and satellite imagery. Auxiliary thematic geo-data (DEM derivatives, soil-data) is used to support characterisation and classification purposes. The set of image data aims at benefitting from both high spatial resolution of aerial images and high spectral and temporal resolution of satellite data (Copernicus: TSX, RapidEye, in future: Sentinels).

2.3 Image analysis

2.3.1 Object generation and validation of geometries

Vector geometries (polygons) are generated in an object based image analysis developed in eCognition Developer and run tile-wise in batch mode in eCognition Server. Besides the generation of vector objects ruleset manages a pre-classification of the objects and attaches attribute information to each of them, mainly statistical parameters on their spectral and height characteristics.

2.3.1.1 Segmentation

The basic approach of the segmentation process is rather pragmatic. Basically it aims at delivering geometries usable for a field mapper in land cover- and respective biotope-mapping. This means that every boundary possibly relevant in landscape is supposed to be mapped. Since even gradual changes in vegetation cover, e.g. in its texture, may depict important habitat boundaries this may lead to some over-segmentation which has some implications on the further processing of the data, e.g. concerning object aggregation.

Image segmentation is run as an iterative process making use of threshold-based and multiresolution approaches (Baatz & Schäpe 2000). All steps currently carried out during segmentation are based on information derived from aerial images, i.e. spectral information and height (including a LiDAR based DTM for the normalisation of the DSM). Tests are currently carried out involving satellite imagery (single scenes and time series, multispectral and SAR) in order to possibly include multitemporal information into the process of object generation. Also DTM derivatives are tested concerning their usability for segmentation in order to better represent geomorphological conditions by object shapes and pattern.

Multi-threshold segmentations based on spectral (Combination of NDVI and Bare Area Index) and DSM-derived (height above

ground) parameters are carried out to geometrically separate the main landscape components "Abiotic"/ "Water" and "Biotic" and to further split up these components themselves. Fixed threshold segmentation (instead of values individually adjusted to each image) is used to mitigate negative effects at the tile boundaries and to apply repeatable rules especially when dividing the landscape by the height above ground, a comparably stable parameter. The threshold based segmentation already delivers a clear picture of the landscape. Nevertheless the objects produced are not yet suitable for detailed mapping.

Figure 2: Threshold based segmentation of aerial image/ nDSM. High vegetation already subdivided by multiresolution segmentation, Mannebach, Saar-Mosel region (NATFLO)

Multiresolution segmentation is then applied to further subdivide the components into smaller objects capturing meaningful subunits e.g. in meadowland or forest. Forested areas and the open landscape are treated with different parameterisations due to different structure and requirements.


Figure 3: Fine-grained subdivision of the image after threshold-based and multiresolution segmentation (NATFLO)

2.3.1.2 Verification of Object Suitability

Sets of test segmentations are carried out for pilot areas with different parameterisations and under changing image conditions. In cooperation with an expert team of biotope mappers the results of the most promising algorithms are tested on applicability in field mapping campaigns. The most suitable algorithm is then chosen to be applied in the full workflow. Example: Test segmentations for forested areas have been carried out for an area of 1000 km² in Pfälzerwald. The aim was

to delineate forest sub-stands of homogeneous height structure and leaf type (broad-leaved, coniferous).
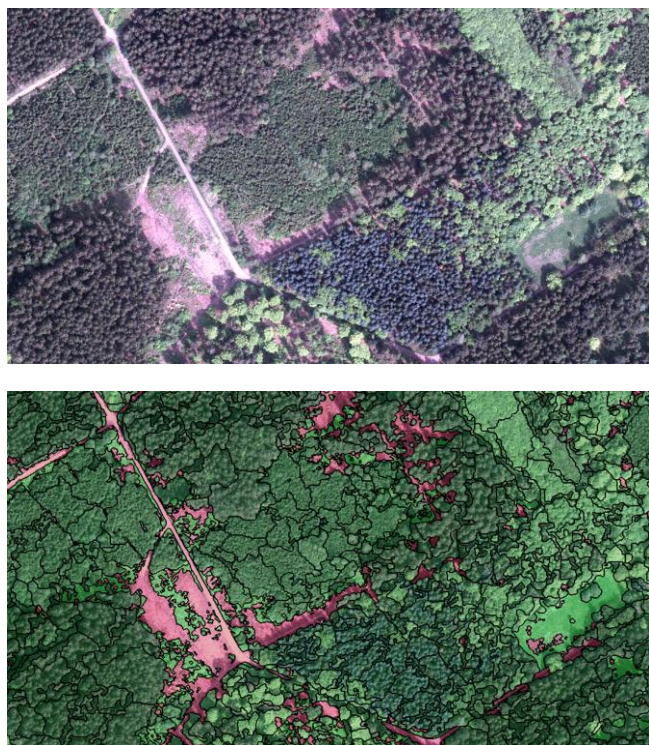
Figure 4: Multiresolution segmentation in a heterogeneous forested area (NATFLO)

2.4 Object characterisation

The classification approach described below (see 2.5.1) makes use of indicators. They describe every single habitat class of the classification scheme. The use of such indicators requires a comprehensive knowledge on the environmental conditions and characteristics of objects. Value based parameters stored as object attributes in data bases fulfil this requirement. These parameters are supposed to be as diverse as possible and describe an object concerning site (soil, macro- and topo-climate, geomorphology and terrain etc.), the characteristics of the cover type (e.g. vegetation height and –intensity) and temporal dynamics within the object like phenological development or management measures.

A data base with a collection of relevant geo-data for the calculation of zonal statistics has been built up. It consists of already available data from official authorities as well as from analyses especially run for the project purposes. For example a large number of parameters on terrain dependent site conditions have been calculated on a state wide LiDAR-based DTM with 5 m resolution (Esri ArcGIS, SAGA Gis). Examples here are simple parameters like slope and aspect. More specific information is offered by derivatives like topographical wetness indices (potential soil moisture), incident solar radiation (topo-climate) or topographic position (terrain/ morphology).

In an automated workflow (Python) all object geometries produced in the object based image analysis have been enriched with attributes on numerous value based parameters. A set of standard statistical parameters has been calculated for each parameter and been attached to each object.

2.4.1 Analysis of multitemporal satellite data

The first step of data production, the derivation of object geometries from the multispectral aerial images and nDSM took advantage from the high spatial resolution of the data. Satellite imagery offers high spectral and temporal resolution.

Further remote sensing technology can be applied to the data base through satellite imagery by methods making use of their higher spectral and temporal resolution.

Methods are being tested analysing optical (RapidEye) and SAR (TerraSAR-X) satellite imagery concerning temporal patterns in time series. Currently the detection of permanent grassland and its characterisation concerning use intensity and pattern is in the focus of interest. Grassland is very important in nature conservancy due to its high biodiversity in semi-natural or low intensity areas while at the same highly threatened by being ploughed up for the production of fuel for renewable energies.

Within this project the analysis of intra-annual multispectral Rapid Eye time series with six acquisition dates, carried out in a test region of 200 km² in western RLP separated grassland from cropland and detected the number of mowing events as well as areas managed by grazing through the use of support vector machine classification.

TerraSAR-X backscatter time series are analysed implementing methods developed and successfully applied by (Schuster et al., 2011) aiming at the detection of cutting events in semi-natural grassland. On top of this methods are being developed for the detection and quantification of significant changes by indices combining backscatter and coherence in SAR-time series.

Both optical and SAR-based approaches are supposed to be run in operational workflows for the derivation of comprehensive information for RLP. Data on management patterns and use intensities is crucial for the characterisation of ecologically valuable areas. Temporal metrics, once produced comprehensively for RLP, are expected to significantly enhance the characterisation of objects and therefore the classification results in the current process.

2.5 Classification approach

2.5.1 Formalisation of habitat classes

Basis of the classification approach is an ontological system that formalises habitat types (e.g. included in the EUNIS nomenclature) in an OWL2/XML ontology (Nieland, Kleinschmit, Förster, & Kleinschmitt, 2015). The basic concepts in this ontology are stored in a shared vocabulary and are used to describe important indicators of habitat classes which have been adopted from the classification schemes. In order to generate comprehensive interoperability this shared vocabulary is built on concepts developed by EAGLE (Arnold et al 2014). Since the EAGLE concepts are, until now, mainly focused on land cover and land use additional concepts had to be included in the system. The development of descriptive indicators is a crucial part of this work and is done in cooperation with surveying experts with great care in an iterative procedure. To generate a meaningful and comprehensive set of indicators for habitat classification is the basis for the classification process and the therefore methodological backbone of this work. In the next step the developed indicators can be used to describe habitat classes according to the subsequent nomenclatures using Description Logics (DL).

2.5.2 Feature selection and data mining

In order to derive descriptive indicators (see 3.4.1) from value based parameters (see 3.3) automated methods are needed. That

includes the selection of features that are most appropriate for the derivation of the indicators to reduce the calculation effort as well as the classification process itself. Since we want to benefit from the computation power and accuracy of supervised classification algorithms on the one hand and the transferability and reproducibility of knowledge-based approaches on the other hand, the Separability and Threshold (SeATH) (Nussbaum 2006) approach has been taken into account. This algorithm statistically identifies characteristic features and their thresholds on the basis of present training data (see 3.4.4) and can therefore be used to generate reproducible and transferable rulesets in a simple statistical approach (see Figure 5).

2.5.3   Ontology-based classification

The derived rulesets (see 3.4.2) can now be imported to the OWL ontology to produce computer readable, reproducible and transferable classification rules. In the next step the segmented objects (see 3.2.1) and subsequent value based (see 3.3) parameters can be imported as OWL individuals. The classification will be achieved by the fact++ Reasoner[i], which is able to perform efficient A-Box reasoning over large ontologies. That means that, in the first step, the reasoner assigns OWL individuals (segmented objects) to indicator concepts based on the rules derived by SeATH (see 3.4.2) and, in the second step, allocates the individuals to habitat classes formalized by taking into account the class descriptions and expert knowledge (see 3.4.1) (see Figure 5).
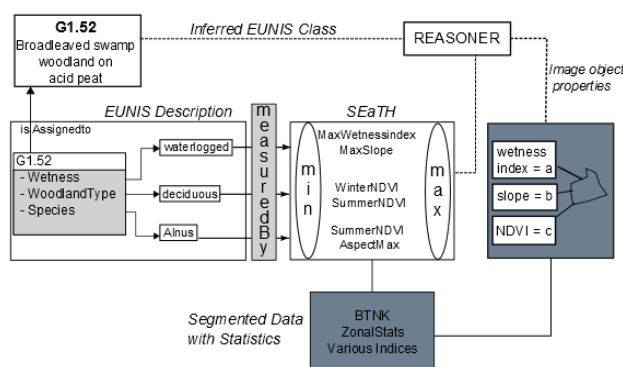
Figure 5: Overview of ontology-based classification. Adapted from Arvor et. al 2013

2.5.4   Reference Data: Training and Validation

Correct and meaningful state wide reference data is needed to provide a basis for the data mining approaches and to validate the classification results. Developing a concept for the generation of a nation-wide set of reference areas for a great number of different habitat classes and their associated indicators is an enormous challenge.
Therefore the reference data will be generated in two steps.
Reference data will be generated making use of a subset of objects geometries produced in this process. The subset is a selection based on auxiliary data taken from the recent biotope field mapping campaign, land survey and other reliable sources. The selection is supposed to cover the full range of habitats and indicator combinations of RLP. The selected objects are to be verified by performing field checks and taking into account recent orthoimagery.

2.6 Dissemination of Data and Methodology

In   Rhineland-Palatinate   an   increasing   demand   for comprehensive geo-data on biotopes resp. the ecological value of areas can be stated. Governmental authorities need such data

for planning matters. Many administrational tasks have to take into account ecological issues, e.g. local development plans or the granting of building permits. Decision making on a political level needs to be informed by reliable, standardised data, and last but not least the requirements arising from EU-law have to be fulfilled. This requires common standards, granted by European directives and institutions (NATURA2000, INSPIRE, EC, EEA) and activities (EAGLE).

Besides the conceptual backgrounds offered by NATURA2000/ EUNIS and EAGLE) it is the INSPIRE directive setting common technical standards for data exchange. The technical implementation of INSPIRE is still in progress and there are new developments regularly. One project aim is to stay up-to-date with current developments in this context. New improvements are continuously discussed in expert meetings and implemented in the data models. This affects especially the data handling and the possibilities concerning formalised exchange of workflows and methodology.

2.6.1   Data access via internet

A project website has been implemented for the dissemination of the geoinformation. A web-GIS enables the user to explore the latest state of data production. A download area offers information on project state and developments. INSPIRE-compatible metadata is stored for all data in the system. OGC-Services offer access to web-maps (WMS) as well as physical data. The latter are currently available via Atom-Feeds but will in future be implemented as Web Feature Service.

2.6.2   Exchange of data and methodology through "linked data"

One big advantage of the presented approach is the possibility to produce computer readable and fully reproducible and transferable classification workflows. That means, on the one hand they correspond to INSPIRE codelists and can be used to fulfil its technical requirements, on the other hand the whole classification logic will be made available via the web to give other projects and authorities the possibility to re-use and enhance the developed methodology. This offers further opportunities   for   a   continuously   increasing   degree   of standardisation of geo-data in Europe.

3. STATE OF PROJECT, CONCLUSIONS AND PERSPECTIVES

In its first year the project has proven that the combined use of workflows developed in previous activities enables the project partners to provide the authorities with full coverages of high resolution land cover objects from official standard data bases. These data are iteratively further developed concerning the quality of the geometric representation of the landscape and the characterisation of each object. The second version of a full data set has been finished beginning of 2015 providing highly suitable objects in all areas with high vegetation. Segmentation algorithms delivering object geometries for areas of the open landscape   are   currently   being   developed.   Skills   and methodology in the analysis of time series of image data are extended, tests are delivering promising results. The project partners are looking forward to implementing operational workflows for state wide analyses in the context of Copernicus and the Sentinel missions.
Besides data production the infrastructure and conceptual background for an ontology based classification environment has been developed. Tests are delivering promising results. The crucial point in this part of the work is the setup of a valid

network of training data representing the full range of habitat types and resp. indicators.

REFERENCES

(Council Directive) 92/43/EEC, C. The conservation of natural habitats and of wild fauna and flora (1992).

Arnold, S., Kosztra, B., Banko, G., Smith, G., Hazeu, G., Bock, M., & Valcarcel-Sanz, N. (2013). The EAGLE concept – A vision of a future European Land Monitoring Framework. In EARSeL Symposium proceedings 2013, "Towards Horizon 2020."

Banko, G., Mansberger, R., Gallaun, H., Grillmayer, R., Prüller, R., Riedl, M., … Walli, A. (2012). Land Use & land cover mapping in Europe: Current practice, trends and future. Examples from national approaches: LISA – the Austrian approach. In Land Use & land cover mapping in Europe: Current practice, trends and future.

Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., … Förster, M. (2015). Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. International Journal of Applied Earth Observation and Geoinformation, 37, 7–16. doi:10.1016/j.jag.2014.11.005

Eu. (2013). Regulation Financing, Management and Monitoring of the CAP. Official Journal of the European Union, (1306), 549–607.

Nieland, S., Kleinschmit, B., Förster, M., & Kleinschmitt, B. (2015). Using ontological inference and hierarchical matchmaking to overcome semantic heterogeneity in remote sensing-based biodiversity monitoring. International Journal of Applied Earth Observation and Geoinformation, 37, 133–141. doi:10.1016/j.jag.2014.09.018

Schuster, C., Ali, I., Lohmann, P., Frick, A., Förster, M., & Kleinschmit, B. (2011). Towards Detecting Swath Events in TerraSAR-X Time Series to Establish NATURA 2000 Grassland Habitat Swath Management as Monitoring Parameter. Remote Sensing, 3(12), 1308–1322.

---

[i] http://owl.man.ac.uk/factplusplus/