Checks by Monitoring
Quality Assurance Framework

Based on JRC 119733 and TG CbM QA

ANNEX IV

Ground data, version 1.1

December 2020

Authors: Wim Devos, Pavel Milenov, Rafal Zielinski, Blanka Vajsova

Table of Content

1. Release notes (changes/updates from version 1.0)

This is the second draft version of a document on the collection and use of ground data in the CbM and CbM QA context. It is neither final nor complete.

The changes from the first draft (from August 2020) are the following:

- Point 3.2.2: Paragraph added on the importance of proper metadata on the ground data collection methods and the local conceptualizations (definitions) applied
- Tables 1 to 4 updated
- Point 4.2.: Revised to account for the experience gained in ground data collection in the course of CwRS implementation.
- There are also a number of editorial changes to enhance the readability and clarity of the content

2. Annex IV - Rationale

2.1.1. The CbM QA performs both qualitative and quantitative external inspection of the Sentinel processing of the agricultural land cover and associated land use on the FOI representation (individual GSAA parcels or dedicated spatial aggregates). Both the automated processing in CbM and the subsequent inspection in CbM QA, observe either a state or, more likely in monitoring, a change of state of the land phenomenon. Whereas CbM processes make that observation mostly by automated methods, the inspection mostly relies on human judgement.

2.1.2. Ground truth is required for both, because it allows to understand, to quantify and to analyse how a state or a behaviour of the land can be optimally observed and interpreted. The ground truth collection does not directly look for a verdict of compliance, regularity or impact, but it is an key component in the design of the relevant methods that provide the verdict.

2.1.3. As a result, the ground truth management is, although obviously related to processing and inspection, essentially independent from both. Ground truth data collected for the design of the CbM algorithms can be reused for the elaboration of the inspection processes used in the CbM QA; a correctly designed and implemented inspection process doesn't need separate validation by newly collected ground truth.

2.1.4. The above elements summarize a part of the mentioning of ground truth in the previous documents: The TG on the decision to go for substitution of OTSC by monitoring (DS/CDP/2018/17), introduces farmer collected ground truth as independent data for comparison and error prediction of the automated machine learning analysis of the discriminatory performance of Sentinels, in local context. The discussion document DS/CDP/2018/18) calls for a "*systematic, but very reduced collection of ground truth allows to demonstrate both achieving the target errors as well as to assess the performance improvement over the years*". The discussion document continues: "[ground truth] *links field observations to the monitoring process outcomes. … Although the numbers involved can be very low, it is important that the entire system and the whole territory are covered and analysed. This … is NOT directly linked to the processing of applications… it is a small scale, independent system wide collection … . The primary purpose is to keep the monitoring system in touch with the reality on the ground by collecting representative ground truth that is used for the*

performance/tuning of the monitoring". Finally, the document mentions: "*Ground truth for marker and marker parameters has a clear expiry date*".

2.1.5. In these document, the term "ground truth" was extensively used. However, this term has some connotation and it's probably better to use the neutral term "ground data". Hence, **ground data** will be the term used henceforth.

3. Technical concept for the ground data

3.1. Nature of the ground data

3.1.1. In the context of CbM or CbM QA ground data relates to observations of the state or the changing state of a given land phenomenon, collected through an observation procedure applied in its natural surroundings (in-situ). This implies the ground data involved represent point observations, identified by a place and a time.

3.1.2. Many ground surveys and capture methods collect aggregated ground data (e.g. abundance observed while walking a transect). Although such aggregated ground data are useful for a number of application and uses, they are not the subject of this document. Also area measurement, known from the OTSC procedures is a form of ground data, but area measurement is not a default part of the Sentinel monitoring procedures and therefore not considered here.

3.2. Properties of the ground data

3.2.1. In the context of CbM or CbM QA described above, desirable ground data properties could be:

- **Systematic**: relevant to the monitoring processes and covering all territory to avoid bias and/or gaps. i.e. easy to use for **sampling**.
- **Affordable**: avoiding heavy investments, costly acquisitions or expensive processing to keep a **low cost**.
- **Timely**: observing the phenomena at the appropriate time and having low **latency** between the event or phenomenon on the field and the timing of the corresponding observation.

3.2.2. As the world in not ideal, neither will any particular ground data collection method. This is however not a major obstacle as long as a particular weakness of the method is understood and appropriate mitigation is introduced. This requires the availability of sufficient metadata to

comprehend the technical aspect of data creation, as well as its epistemological context and the implicit conceptualizations involved. For example, a field observation that report an agricultural land being "abandoned" should be accompanied with the appropriate definition of what an abandoned land is in the given local/cultural context. This is particularly important when the ground data represents already information (convey a fact to resolve given uncertainty) and not just a "raw" instrument measurement.

3.3. Time latency of ground data

3.3.1. In principle, each process flow of the items in the CbM should rely on one or more observable phenomena because that is essentially, the only thing that is being observed through the Sentinels. Such phenomena or behaviours, could be either of the type, by nature of detection process from the Sentinel data:

- an occurrence of an abrupt land cover change;
- an evidence of a gradual land cover transition;
- an observation of a tell-tale event;
- the identification of a crop;
- the detection of a geometric anomaly or pattern.

3.3.2. Latency plays a much more important role for the dynamic phenomena: the impact of a land cover change lasts much longer than the impact of mowing on a grassland. If the CbM process doesn't need a fast verdict, latency will not be an issue for the irreversible events.

3.3.3. Crop identification is good example to show how latency is method dependent: in the field, many crops can be identified throughout the growing season, however remote sensing techniques can often only deliver an identification by the end of the season.

3.4. Ground data sources

3.4.1. The most intuitive observation of the land phenomenon comes from your eyes, i.e. being there in time or having someone/something documenting the view as if you were there. Indeed, the well-known rapid field visit are currently often cooperating with more effective capturing methods (similar to Google Streetview) or UAV-drones, being complementary source of data. These terrestrial methods (called as such since they require presence of

operators or technicians on the field) share a need for serious logistic preparations, critical dependence on timing and relatively high costs.

3.4.2. The Computer-Assisted Photointerpretation (CAPI) can be sufficiently intuitive to derive ground data for many phenomena directly from the observations made on remote sensing imagery. Indeed, there's no actual ground excursion involved, but an experience photo-interpreter with supported with supported visual tools can make a lot of relevant calls, with the same certainty as if applies an observation procedure its natural surroundings. There is a variety of remote sensing sources available, such as the VHR imagery. Their suitability varies by coverage (from wall2wall to zoned) and by information content (spatial resolution, refresh cycles, electromagnetic bands) and these specifications affect the purchase costs (from free to commercial).

3.4.3. Terrestrial methods and remote sensing techniques are usually rigidly organized by the collecting organization with rigorous data capturing procedures. However, with the leaps on information and communication technology (ICT) of the recent years, less orchestrated elements like, crowd capture, voluntary land cover surveys or farmer's geotagged imagery are becoming prominent source of abundant data. With the use of artificial intelligence tools (self-organizing maps, convolution neural networks), these big data sets can be mined for relevant ground data (otherwise not possible manually), in particular for complementing Sentinel imagery that has a 3-day revisit cycle. Furthermore, farmer's geotagged imagery can theoretically be set up to mimic a rapid field visit.

3.4.4. Table 1 illustrates certain aspects of the possible ground data collection methods in support of a CbM system. It classifies the (generic) performance of each method regarding latency, sampling and cost in three classes. The higher ranked classes are the most desirable.

Performance parameters	Controlled and driven by collecting organization							Voluntary, sporadic, citizen-driven
	Terrestrial methods and data sources			*CAPI methods and data sources*				
	RFV	Drone	StreetView	Sen1&2	HHR	VHR	Aerial	Geotagged
LATENCY								
Short								
Medium								
Long								
SAMPLING METHOD								
Random								
Clustered								
Convenience								
COST								
Low								
Medium								
High								

Table1: Performance of some common ground data collection methods from the experience collected from the OTSC, LPIS and CbM. Dark green= very good, green=acceptable, light green=acceptable under conditions, white = not acceptable/not applicable; yellow=acceptable, but potentially flawed.

3.4.5. Table 1 also shows clearly that, provided appropriate methods are used, the Sentinels are an overall winner with no time restraints, a complete coverage allowing random sampling and a low cost. These traits explain why Sentinel is the source of the CbM QA inspection itself. Table 2 elaborates and complements the table.

Source/method	Pros	cons
RFV observation	Intuitive and Universal (applicable for any land phenomena)	High cost, resource demanding, time dependent
StreetView-technology	Low or no cost	Location dependent
Sentinel 1+2 / L8 CAPI	Free of charge	Limited spatial resolution
HHR CAPI	Higher spatial resolution	Acquisition cost
VHR CAPI	High detail, familiar	High cost of dedicated acquisition, timing
Aerial CAPI	High detail, familiar	Expensive, timing
Geotagged imagery	Intuitive and Universal (applicable for any land phenomena)	Organisation, validity
Third party data (national thematic data, Copernicus Land, LUCAS)	Existing data, easy to use	Case dependent,

Table2: The advantages and disadvantages of the various ground data collection methods and related data sources

3.4.6. Tables 1 and 2 are quite generic and not intended to be exhaustive. Technologies and methodology continuously develop and any emerging tool that provides reliable ground data can in theory be involved.

4. Practical implementation

4.1. Selecting ground data collection methods

4.1.1. For each targeted marker/output/observation[1], the EU Member State (MS) should investigate which ground data collection method(s) could be applicable. It is clear that there will be often more than one ground data collection method that can be validated, i.e. proven to correctly provide the particular ground data intended. It is merely a logistic and financial decision which source to use; either "off the shelf" (i.e. as a well-known technology) or with a set of flanking/innovative implementation methods (e.g. instructions to farmers for geotagged imagery uptake, investment in AI methods for data mining).

[1] See CbM discussion document and Annex III of TG on CbM QA

4.1.2. Table 3 presents a tentative evaluation of the ground data collection methods for a number of information extraction types (as defined in Annex III of the present TG), which form the basis of the tests and the inspection lots in the CbM QA Note that this table 3 acts as a mere illustration and cannot be considered as directly applicable template, in a particular CbM or CbM QA context; MS could use knowledge of local conditions, studies and expert judgement to compile a template that is suitable for their application.

Type of information extraction	**Controlled and driven by collecting organization**						**Voluntary, sporadic, citizen-driven**	**Ready-to-use third part datasets**
	Terrestrial methods and data sources		***CAPI methods and data sources***					
	RFV	StreetView	Sen1&2	HHR	VHR	Aerial	Geotagged	Copernicus, LUCAS
G1. Spatial cardinality								
T1. Abrupt land cover change								
T2. Land cover transition								
T3. Tell-tale soil roughness change								
T3. Tell-tale vegetation drop								
T4. Winter barley								
T4. Winter Wheat								

Table 3: provisional, incomplete matrix of ground data to validate the CbM QA inspection instructions per lots with: RFV= rapid field visit, SEN1&2= 10m resolution, HHR= 3-5m resolution VHR= 0.5-1m resolution. Dark green= very good, green=acceptable, light green=acceptable under conditions, white = not acceptable/not applicable.

4.2. Possible best practices

4.2.1. The current regulation 2013/1306 and its successor have provision for a acquisition of satellite imagery by the European Commission (COM). Consequently, VHR imager collected for the LPIS QA exercise and residual CwRS activities can be reused to collect ground data. In future, complementary HHR image sets could also be considered to target particular ground data collection zones in selected well documented case. An appropriate specification for such HHR imagery needs to be developed, trading off spatial and temporal properties.

4.2.2. Where the amount of data flow/number of images is low, human operators can interpret most of the capture methods, including the Streetview-like data streams. But the surge in the use of geotagged imagery offers an unseen potential to re-use those data for automated mining of ground data from these source. The Commission will remain available to explore the potential of other ground data sources, such as drones, to target particular zones.

4.2.3. As ground data is essentially a set of point observations, the Commission proposes to use a common template (profiled from and ISO 19156 and ISO 19157) to describe the basic observations. Similar templates are also used for the LPIS QA ETS-annex I.

4.2.4. Ground data collection was always an essential component of the CwRS and there were reporting templates developed in the scope of the CwRS QC. Unfortunately, for many years already, JRC does not collect systematically such QC data, neither follows the development of the ground data collection protocols of the EU MSs. The availability of the Sentinels and the emergence of the new technologies, require a revisit of the way EU MS organize this process, towards its adaptation in CbM context. This revision should provide the EU MSs with the a general guidance or template on how MS:

- plan the collection of ground data,
- organize the collection ground data,
- asses the suitability of a particular observation for a process or marker,
- document their findings,
- apply it in an inspection methodology,
- feed all this back in the CbM design.

4.3. Compatibility

4.3.1. Despite its appearance, Table 3 is not much different from the approach currently used for the LPIS QA inspections. The logic behind the selection of method is identical; LPIS QA does an area measurement (rather than the mere occurrence check required for CbM) on land cover and allows a combination of data capture methods for a single sample. Table 4 illustrates how the LPIS QA inspection is represented in the mind-set of table 3.

Type of information extraction	**Controlled and driven by collecting organization**						**Voluntary, sporadic,**	**Ready-to-use third**
	Terrestrial methods		***CAPI methods and data sources***					
	RFV	StreetView	Sen1&2	HHR	VHR	Aerial	Geotagged	Copernicus, LUCAS
G1. Spatial cardinality								
T1. Abrupt land cover change								
T2. Land cover transition								
T3. Tell-tale soil roughness change								
T3. Tell-tale vegetation drop								
T4. Winter barley								
T4. Winter Wheat								

Table 4: Equivalent of Table 1, illustrating the LPIS QA inspection options of ground data: Either CAP on high resolution imagery or a combination of field observation with geotagged imagery (ref: LPIS_TG_ETS). *Dark green= very good, green=acceptable, light green=acceptable under conditions, white = not acceptable/not applicable.*

4.3.2. This evidences that LPIS QA acknowledges the potential for equivalence of terrestrial and remote sensing methods to collect data (either ground data or inspection data). CAPI is however expected to be less costly, due to the simpler logistics implied.