A REVISED PROCESSING LEVEL SCHEME FOR EARTH OBSERVATION DATA

Peter A. Strobl

European Commission, DG Joint Research Centre, 21027 Ispra, Italy

ABSTRACT

The simplistic, linear, and largely interpretable depiction of the Earth Observation value chain that is enshrined in the currently used scheme of "Processing Levels" should be retired and replaced by a well-defined, versatile, harmonized, and easily understandable matrix. The proposed approach of disentangling measurand and geometric processing steps has the potential to not only booster interoperability in multi-source production processes, but also to allow targeted standardization and to optimise workflows with respect to uncertainty tracing and data quality control.

Index Terms Processing Levels, ARD, Interoperability

1. INTRODUCTION

Soon after larger quantities of digital satellite imagery became available in the 1970's, mission scientist and software engineers started to organise the processing chain for respective data into several sequential steps. The rationale behind was to distinguish raw or less processed data from refined or more highly processed data products which usually would make it easier for users to access and analyse them.
An attempt of a generic definition of 'Processing Level' can be found in Weaver, 2014 [1]:

> *Processing levels*
> A means of describing the way remote sensing digital data are processed from raw or engineering units to informational geophysical products. These are typically differentiated by numeric (or sometimes alphabetical) hierarchies, from Level Zero, indicating engineering units or the data from the actual sensor system, to Level Four indicating geophysical products. Typically processing levels are applied to satellite data streams, but they have been utilized in other forms of remote sensing data.

The same paper gives a good overview of the history and different developments of processing levels over the past decades. The most influential implementation of processing levels is the one originating from NASA Earth Observing System (EOS) and endorsed by the Committee on Earth Observing Satellites (CEOS) around 1996 (Table 1). It is the basis of most level definitions still in use by major space agencies and foresees a classification of data products by the sort of processing used in their generation and, to a lesser extent, the sort of uses to which these products might be put.

Table 1. NASA Processing Level Definitions as in EOS Reference Handbook 1993 [2]

Level	Process description
(Raw)	Data in their original packets, as received from a satellite.
0	Reconstructed unprocessed instrument data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g., ephemeris, health and safety) appended.
1	Unpacked, reformatted level 0 data, with all supplemental information to be used in subsequent processing appended. Optional radiometric and geometric correction applied to produce parameters in physical units. Data generally presented as full time/space resolution. A wide variety of sub-level products are possible.
2	Retrieved environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the level 1 source data.
3	Data or retrieved environmental variables which have been spatially and/or temporally resampled (i.e., derived from level 1 or 2 products). Such resampling may include averaging and compositing.
4	Model output or results from analyses of lower level data (i.e., variables that are not directly measured by the instruments, but are derived from these measurements).

While these level definitions primarily focus on removal of sensor and acquisition artifacts and subsequent refinements of the measurand, they provide only marginal guidance with respect to the geometric improvements and how these are tied to the different levels.
The need for interoperability and the availability of reliable and automatic geometric correction methods made the production of regularly gridded and orthorectified data increasingly mandatory. Additional intermediate steps in which these operations are addressed were introduced. Since the early 2000's for example Level 1C (L1C) is frequently described as "data orthorectified and re-sampled to a specified grid". However, these intermediate levels are not harmonized across agencies and important differences between agencies persist up to today [1].
This way geometric refinement and resampling operations became part of the processing workflow. Consequently, the refinement of the measurand to higher levels in many current

doi:10.2760/46796

processing chains inherently is tied to spatial resampling steps of which the user will not always take appropriate notice. As long as users are building their products with single mission data, as it was often the case in the past, this is less problematic as data will usually not be resampled more than twice before arriving at Level 4.
However, increasing multi-mission applications and technologies such as imaging spectroscopy, which is highly sensitive to any resampling, require to open processing pathways which not only strictly keep track of performed resampling, but also allow to reduce it to a minimum in the interest of preserving data quality.

2. DISENTANGLING PRE-PROCESSING CONCERNS

In order to understand how the processing in ground segments and downstream value chains can be organised it is beneficial to distinguish between operations refining the measurand, such as radiometric calibration, or atmospheric correction in case of optical sensors, and those which alter the spatiotemporal sampling structure of the single observations (or 'pixel') as for example orthorectification, but more generally any other type of 're-sampling' in space (or time).

2.1. The measurand dimension M

As explained in the introduction, the refinement of the measured quantity is by far the most visible and noticed alteration which occurs along the value chain. Nevertheless, has it been only vaguely characterized by the CEOS processing levels, which left a wide room for interpretation of what exactly would happen at which stage, and in which way data at each level can be expected to be interoperable [3]. One of the improvements of the scheme, as laid out in Table 2, is the direct relation of the step with the main alteration in the refinement. This allows to give each step an intuitive label that clarifies its main purpose.

Table 2. Proposed steps along the Measurand dimension

Step	Process description
(Raw)	The complete and unaltered/unprocessed set of *data* acquired by one or several *sensors* on a platform
M/**0**	**uncalibrated** Unaltered/unprocessed Level 0 *(main) sensor data* annotated with processed *ancillary data* and supplemented by *auxiliary data* (including radiometric and geometric calibration coefficients and geo-referencing parameters) allowing further processing to higher Levels.
M/**1**	**sensor-calibrated** Level M/0 sensor data which have been calibrated (ideally traceable to SI) and spatially aligned (co-located, eventually co-gridded) to represent **at-sensor measurements** (value and uncertainty) in sensor nominal spatiotemporal sampling, supplemented by appropriate ancillary and auxiliary data for further processing.
M/**2**	**target-calibrated** Level M/1 data processed to represent **geophysical property values (and uncertainties) for a specified target** (object, feature of interest, e.g. surface reflectance, apparent temperature) derived from M/1 sensor data, as much as possible maintaining the sensors nominal spatial and temporal sampling (observation preserving).
M/**3**	**homogenised (for a definition see [4])** Level M/1 or M/2 data which have been generalised and integrated across one or several platforms and acquisitions to achieve an increased, more regular or in any other form enhanced spatial or temporal coverage in which states **geophysical values agnostic of the originally acquiring sensor and observation condition** and thus directly comparable. This homogenisation and fusion may include measurand re-calibration to external standards and references including use of modelling, aggregation and interpolation.
M/**4**	**derived/infered** Model output or results from analyses of Level M/3 (or lower level) data i.e., attributes that might not be directly observable by the sensor(s) but are **derived from observations in combination with other external incl. non-observational data** using techniques like modelling or machine learning (AI).

2.2. The spatiotemporal dimension G

Even more intuitive is the sequence of the spatio-temporal refinement, which for simplicity is reduced here to the spatial aspect. In principle the same consideration would apply to a temporal dimension but as long as most 'datacubes' do not discretize time to regular intervals [5], and therefore don't require temporal resampling, this is less relevant here.

Table 3. Proposed steps along the Geometry dimension

Step	Process description
G/**A**	**raw** individual observations (samples) which are not geolocated.
G/**B**	**geolocated/georeferenced** Each observation is geolocated with documented uncertainty in a (traceable) Geodetic Reference System. At this stage the individual observations can be considered forming an irregular 'point cloud' which might also be pseudo-regularised to enhance storage efficiency ('sensor grid').
G/**C**	**georectified/gridded** Observations have been spatially re-sampled to fall within a specified, usually regular, geodetic grid.
G/**D**	**regridded1** Observations have been re-sampled from the original geodetic grid into another specified (geodetic) grid.
G/**E**	**regridded2** Observations or derived values have been again re-sampled from the second geodetic grid into a third one. This should under no circumstances be equal to their Stage G/C geodetic grid.

doi:10.2760/46796

3. THE PROCESSING MATRIX

Once the steps along the two dimensions (or types) of refinement are identified, it is straightforward to combine them to span a matrix in which either type retains its independence. The results is shown in the upper half of Figure 1. Each field now yields a possible processing level as a combination of the two types of refinements.

3.1. Mapping the traditional CEOS processing levels

To illustrate the versatility of such a system it is useful to first map existing 'traditional' levels into that scheme. As can be clearly seen they by and large occupy the diagonal of the matrix. And even though it might be debatable to what extent Level 3 and Level 4 processing entails re-gridding it is evident that the coupling of both types of processing yields a significant danger of repeated resampled which might in the end not even be clear to the user, as the definition of these Levels leaves this point explicitly open.

3.2. Proposal of a new processing level scheme

So far the labelling of the different steps along each type of refinement was not of particular relevance, and in principle any kind of encoding would work. However, to maintain a certain continuity with the widely familiar CEOS labels the numbering along the measurand refinement combined with letters along the geometric axis make most sense.

The matrix of level labels that is created in this way is largely similar to the traditional CEOS labels [3] with the exception of the of course widely used L2A which depicts e.g. surface reflectance data which according to the proposed scheme would now be called L2C. This however should be considered a minor caveat with respect to the clarity and flexibility which the new scheme offers.

However, to illustrate the envisaged use of the scheme, not all the possible Levels have been populated. The refinement of raw or 'Level 0' data along any of the other axis is unlikely useful. Similarly, is a refinement of only georeferenced data up to the level of modelling and inference is hard to imagine. Possible but certainly not recommendable is the double re-gridding of low level (1&2) data as these would essentially blocked them from further refinements likely requiring another re-gridding.

Considering what orthorectification and re-gridding will entail with respect to the impact on uncertainty of the result, recommendations can be given on which path of processing through the matrix will be best to optimize the products. This has been attempted by colouring the field in the lower part of Fig. 1 accordingly.

Last but not least, could the sequence through which each single observation is run in the course of its value adding be recorded for traceability, avoiding e.g. multiple re-gridding. With a total of less than 16 fields such an encoding would not require more than four bits per stage which over four stages could be stored in a single 16 bit parameter.

Geometry \ Measurand	M/0 - raw	M/1 - sensor calibrated	M/2 - target calibrated	M/3 - homogenised	M/4 - derived
G/A - raw	L0/L1A				
G/B - georeferenced		L1B			
G/C - georectified		L1C	L2(A)		
G/D - regridded1				L3	
G/E - regridded2					L4

Geometry \ Measurand	M/0 - raw	M/1 - sensor calibrated	M/2 - target calibrated	M/3 - homogenised	M/4 - derived
G/A - raw	L0/L1A				
G/B - georeferenced		L1B	L2B	L3B	?
G/C - georectified		L1C	L2C	L3C	L4C
G/D - regridded1		L1D	L2D	L3D	L4D
G/E - regridded2				L3E	L4E

ideal tolerable critical

Fig. 1. Proposal of a matrix of processing levels based on separated measurand and geometric refinement steps above: mapping of current popular CEOS processing levels below: a new set of levels and recommended workflow paths (see color codes)

doi:10.2760/46796

4. FUTURE ARD INTEROPERABILITY LEVELS

An issue which has recently come up in various discussions in the EO and ARD community and which is still not well represented in literature is the expansion of the concept of 'Analysis Ready Data' beyond the classical 'Level 2' products for which CEOS-ARD has come up with specifications [6].
Meanwhile a Standards Working Group was convened under OGC auspices to develop ARD into an official standard [7]. The first part of the new standard will have to deal with the question of 'What means ready and for whom?' and most naturally will require a deeper analysis and conceptualisation of respective 'readiness classes'.
The proposed categorisation along the measurand refinement opens a path also for a clearer usage-oriented definition of 'analysis readiness' at the different levels of the value chain. Below list gives a first hint at what such labels could look like and how they could inform about the type of analysis for which data and products at a specific level are meant:

- L0 (raw data)
- L1A (calibration ready data)
- L1B (orthorectification ready data)
- L2B (conflation/combination ready data)
- L3B (fusion ready data)
- L3C/D (analysis/model ready data)
- L4C/D (inference ready information)

These readiness classes as well as the matrix scheme overall are of course indicative and will have to undergo the scrutiny of the standards development process and the endorsement by CEOS. However, as turned out from the industry led 'ARD23' workshop [8], also the so called NewSpace sector is keen to see a user- and service-oriented evolution of the rather traditional and space agency centric value chains.

5. CONCLUSIONS AND OUTLOOK

The classical linear processing level scheme might benefit from an overhaul. Processing levels must be helpful for structuring processing chains and archives into blocks sharing relevant attributes and allowing definition of interoperability goals across more than one sensor or class of sensors. Separating geometric from measurand refinements in the way presented here offers several advantages over the current practice:

1) Possibilities for a more flexible and comprehensive lay-out of processing pathes (e.g. to track and minimise uncertainty contribution)
2) Clear and unambiguous indexing of processing levels to allow easy referencing and traceability of processing history down to pixel (sample) level
3) Accommodation also of non-imaging and non-satellite observations into one scheme, and expansion of interoperability beyond classical satellite based EO as a declared goal of CEOS [9]
4) Dedicated standardisation of ARD at each level for increased interoperability and user guidance

While with these the proposed scheme addresses many of the overall goals, it is meant as no more than a basis for future discussions. In particular, the used terminology will have to be elaborated and agreed among all stakeholders. For example, the term 'homogenisation' as used for labelling Level M/3 is far from being uniquely established and a clear definition is of key importance. However, time is pressing. The dynamic with which EO data pre-processing is changing due to new actors, increased demand, and vast amounts of classical and novel EO data, keeps pushing the current scheme to its limits, if not beyond. A discussion is therefore timely and the upcoming ARD standardisation and the BiDS conference will provide ideal occasion for a for a broad stakeholder consultation, which ideally will be followed up by a proper white paper or publication on this topic.

REFERENCES

[1] R. Weaver, Processing Levels, in: Njoku, E.G. (eds) Encyclopedia of Remote Sensing, Encyclopedia of Earth Sciences Series. Springer, New York, NY. https://doi.org/10.1007/978-0-387-36699-9_36, 2014

[2] National Aeronautics and Space Administration (NASA), *Earth Observing System (EOS) Reference Handbook*, ed. G. Asrar and D. J. Dokken. National Aeronautics and Space Administration, Earth Science Support Office, Document Resource Facility, Washington D.C., 1993

[3] G. Gutman and A. Ignatov, "Towards a common language in satellite data management: a new processing level nomenclature," *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, Singapore, pp. 1252-1254 vol.3, doi: 10.1109/IGARSS.1997.606413, 1997

[4] FIDUCEO Fidelity and uncertainty in climate data records from EO glossary, https://research.reading.ac.uk/fiduceo/glossary/, last accessed 09.10.2023

[5] M. Purss, P. Peterson, P. Strobl, C. Dow, Z. Sabeur, R. Gibb, J. Ben, "Datacubes: a discrete global grid systems perspective", Cartographica 54(1):63–71, doi:10.3138/cart.54.1.2018-0017, 2019

[7] OGC ARD Standards Working Group announcement, https://ceos.org/news/ogc-ard-swg/, posted 2.2.2023, last accessed 9.10.2023

[6] CEOS-ARD Specifications and summary of products: https://ceos.org/ard/, last accessed 09.10.2023

[8] Repository of agenda and recordings of the ARD23 workshop, https://www.ard.zone/ard23, last accessed 09.10.2023

[9] CEOS Interoperability Handbook https://ceos.org/document_management/Meetings/Plenary/34/Documents/CEOS_Interoperability_Terminology_Report.pdf, 2008, last accessed 14.07.2023