A revised Processing Level scheme for increased flexibility and interoperability

Peter Strobl, EC-JRC

2ND WORKSHOP ON INTERNATIONAL COOPERATION IN SPACEBORNE IMAGING SPECTROSCOPY

"La Collinetta Events", Rome – Italy

19 October 2022

Why processing levels?

- Going from **raw data** acquired by a sensor to **information** meaningful to an end user requires a series of interventions on the data, often referred to as **pre-processing and value-adding**
- Structuring this process is meant to assign specific tasks and resulting qualities to each step and inform the user about what to expect
- Different applications and user types will build on different Levels
- At each Level, data products will have certain communalities which should help in establishing harmonised formats and standards

Processing Level definitions

Over the years a number of Level definitions were developed, they are similar but not identical…

	CEOS 1996	LTDP Guidelines	NASA EOSDIS	ESA PDGS Glossary
Raw	Data in their original packets, as received from a satellite.	The physical telemetry payload data as received from the satellite, i.e. a serial data stream without de-multiplexing. These data are not computer compatible.	Missing	Missing
Level0	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.	Reconstructed unprocessed data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g. ephemeris, health and safety) appended.	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. (In most cases, the EOS Data and Operations System (EDOS) provides these data to the data centers as production data sets for processing by the Science Data Processing Segment (SDPS) or by a SIPS to produce higher-level products.)	Reconstructed unprocessed data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g. ephemeris, health and safety) appended.
Level1	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.	Reconstructed unprocessed data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and geo-referencing parameters (e.g. ephemeris) computed and appended but not applied to the Level 0 data.	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to Level 0 data.	Reconstructed unprocessed data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and geo-referencing parameters (e.g. ephemeris) computed and appended but not applied to the Level 0 data.
Level1A	Missing	Radiometrically corrected and calibrated data in physical units at full instrument resolution as acquired.	Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).	Radiometrically corrected and calibrated data in physical units at full instrument resolution as acquired.
Level1B	Missing	L1B data orthorectified, re-sampled to a specified grid.	Missing	L1B data orthorectified, re-sampled to a specified grid
Level2	Retrieved environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the level 1 source data.	Derived geophysical parameters (e.g. sea surface temperature, leaf area index) at the same resolution and location as Level 1 source data.	Derived geophysical variables at the same resolution and location as Level 1 source data.	Derived geophysical parameters (e.g. sea surface temperature, leaf area index) at the same resolution and location as Level 1 source data.
Level3	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.	Data or retrieved geophysical parameters which have been spatially and/or temporally re-sampled (i.e. derived from Level 1 or 2 products), usually with some completeness and consistency. Such re-sampling may include averaging and compositing.	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.	Data or retrieved geophysical parameters which have been spatially and/or temporally re-sampled (i.e. derived from Level 1 or 2 products), usually with some completeness and consistency. Such re-sampling may include averaging and compositing.
Level4	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).	Outputs or results from models using lower level data as inputs and, thus, not directly derived from the instruments.	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).	Missing

Levels in a nutshell

- Linear sequence in which one Level builds on previous ones
- Not fully harmonised across sensor types
- Recent linkage with ‘Analysis Ready Data’ and the ‘ready for what’ discussion

Raw Data → Level 0 → Level 1 → Level 2 → Level 3 → Level4

reconstructing
supplementing

reformatting
calibration

correction
orthorectification

harmonisation
resampling

modelling
fusion

?

Level-1 Interoperability

Analysis Ready Data

“AI Ready Data”

“Inference Ready Information”

European Commission

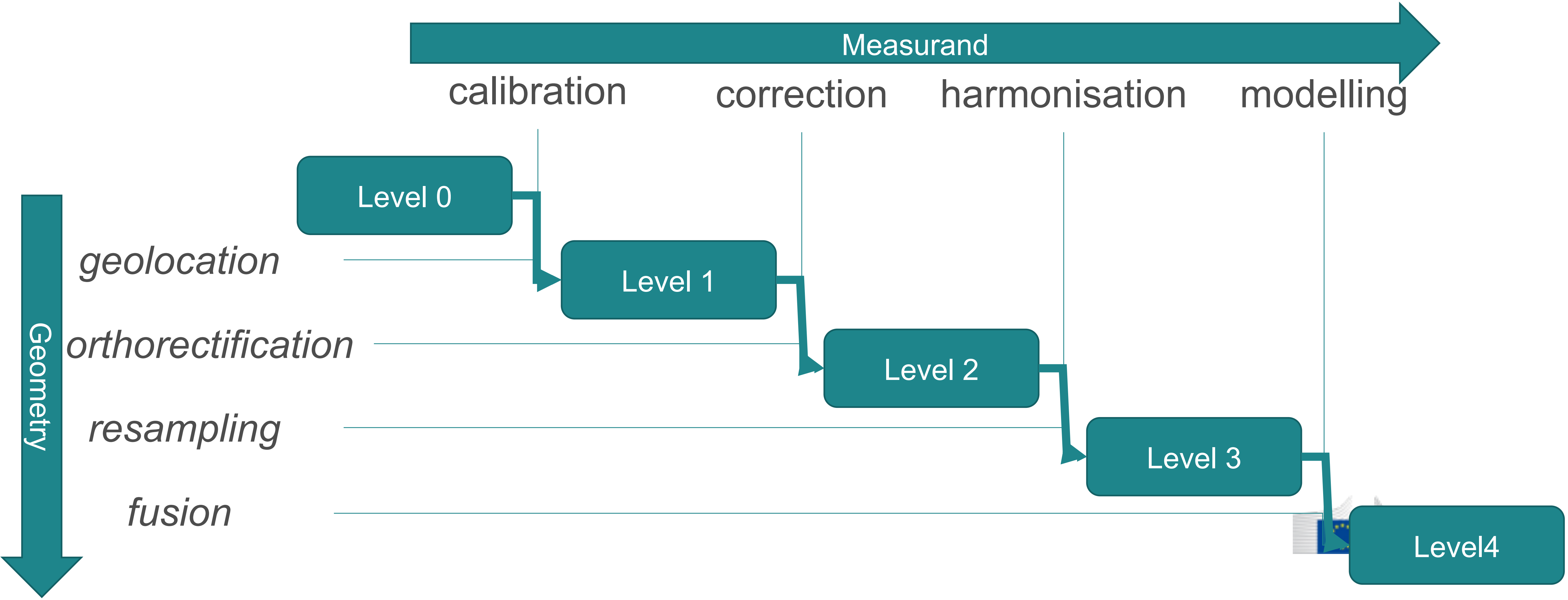
Why reopening the discussion?

- The currently most familiar ‘Level’ concepts (in particular that of CEOS) go back to the 1990’s and were mostly developed having radiometric (optical) instruments in mind
- Interoperability was not the major driver
- The processing chain was mostly linear
- Pre-processing was considered task of the instrument operator’s ground segment, while value adding was the domain of the user
- Multi-disciplinarity of Earth Sciences demand much broader data integration than in the past, at all Levels!

European Commission

Disentangling the processing steps

Separate between interventions affecting the measurand and those affecting geometry



re-thinking CEOS Processing Levels

- Considering these two types of refinement strains separate, a matrix could be built in which classical Processing Levels would (roughly) appear as below:

Geometry \ Measurand	M/0 - raw	M/1 - sensor calibrated	M/2 - target calibrated	M/3 - harmonised	M/4 - derived
G/A - raw	L0/L1A				
G/B - geolocated		L1B			
G/C - orthorectified		L1C	L2(A)		
G/D - resampled1				L3	
G/E - resampled2					L4

The measurand dimension M

- **Level - (raw):** The complete and unaltered/unprocessed set of *data* ..
- **Level M/0 (uncalibrated):** Unaltered/unprocessed Level 0 *(main) sensor data* annotated with processed …
- **Level M/1 (sensor-calibrated):** Level M/0 *sensor data* which have been calibrated and spatially aligned (co-located, eventually co-gridded) to represent *at-sensor observations* (value and uncertainty) in physical units in sensor nominal spatiotemporal sampling and viewing geometry, geolocated and supplemented by appropriate *ancillary* and *auxiliary data* for further processing
- **Level 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, sea surface temperature, leaf area index, soil moisture) derived from M1 sensor data maintaining the sensors nominal spatial and temporal sampling (observation preserving).
- **Level M/3 (harmonised):** Level M/1 or M/2 data which have been harmonised and combined 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 **values are independent of the originally acquiring sensor**. Harmonisation and fusion may include data resampling to external references making use of modelling, aggregation and interpolation.
- **Level 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 (incl. AI).

The geometry dimension G

- **Stage G/A (raw):** individual observations are not geolocated
- **Stage G/B (geolocated):** Each observation (sample) is geolocated with documented uncertainty. At this stage the individual observations can be considered forming a point cloud which might also be regularised to enhance storage efficiency (‘sensor grid’).
- **Stage G/C (orthorectified):** Observations have been re-sampled to fall within a specified, usually regular, geodetic grid.
- **Stage G/D (resampled):** Observations have been re-sampled from the original geodetic grid into another specified (geodetic) grid.
- **Stage G/E (twice resampled):** 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 original geodetic grid

A new Processing Level matrix

For the discussion of 'Analysis Readiness' of data, a clearer separation of these two 'dimensions' of processing yields a chance to obtain a transparent scheme in which also recommendations about best possible paths (processing sequences) are feasible. This would be advantageous for defining 'Analysis Ready Data' standards at different processing Levels and for their respective interoperability.

Measurand / Geometry	M/0 - raw	M/1 - sensor calibrated	M/2 - target calibrated	M/3 - harmonised	M/4 - derived
G/A - raw	L0/L1A				
G/B - geolocated		L1B	L2B	?	??
G/C - orthorectified		L1C	L2C	L3C	L4C
G/D - resampled1		L1D	L2D	L3D	L4D
G/E - resampled2				L3F	L4F

ideal	tolerable	critical

Future interoperability levels?

- L0 (raw data)
- L1A (calibration ready data)
- L1B (orthorectification ready data)
- L2 (conflation/combination ready data)
- L2 (fusion ready data)
- L3C/D (analysis/model ready data)
- L4C/D (inference ready information)

European Commission

Summary & Conclusions

- The classical linear Level scheme might benefit from an overhaul
- Separating geometric (spatiotemporal?) from measurand interventions offers possibilities for more flexible and comprehensive characterisation of processing paths
- Clear and unambiguous indexing of Levels for easy referencing and traceability (4x4 states = 4bits allows storage even per measurement)
- Generic enough to accommodate non-imaging sensors and in-situ observations
- Tailored standardising for better interoperability

European Commission

Thank you!

Peter.Strobl@ec.europa.eu

European Commission