A REVISED PROCESSING LEVEL SCHEME FOR EARTH OBSERVATION DATA

Peter Strobl, EC-JRC

Big Data from Space 2023

Austria Center, Vienna, Austria

6-9 November 2023

Pertinent questions:

"My favourite slide of the entire week"
Scott Simmons, Chief Standards Officer OGC
https://portal.ogc.org/files/?artifact_id=104110

ANALYSIS

- What kind?
  visual interpretation, simple algebra, machine learning,
  physical modelling, AI, with same sensor, sensor fusion, data assimilation
  …

READY

- For whom?
  EO experts, EO users, analysts, decision makers, general public …

DATA

- Which types?
  "Remote Sensing", "in-situ", "Earth Observation", "gridded", "localised",
  "orthorectified", "spatiotemporal" …

European Commission

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 (‘**Level**’) and inform the user about what to expect
- Different **applications** and **user** types will build on **different Levels**
- At each Level, data products should have certain **communalities** to help in establishing **harmonised metadata, formats,** and other **standards**

Real world implementation

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

Why reopening the discussion?

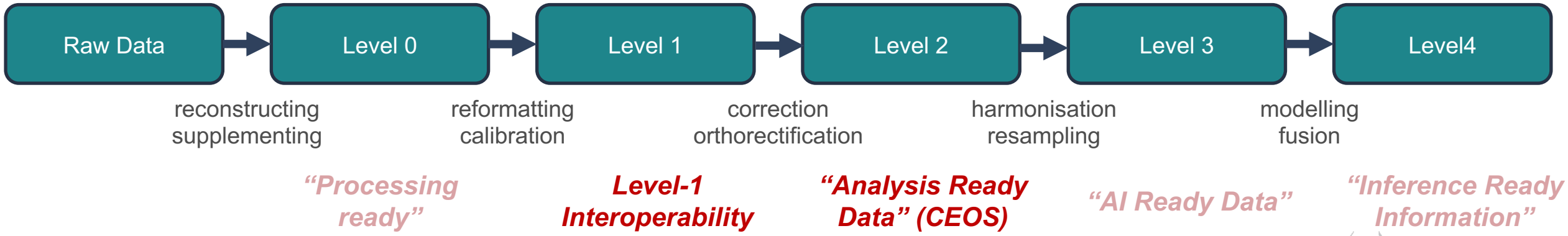
- The currently most familiar ‘Processing Level’ concepts (in particular that of CEOS) go back to the 1990’s and were mostly developed having radiometric (optical) instruments in mind
  - Interoperability was not the major driver
  - The processing chain was mostly linear and mono-track
  - Pre-processing was considered task of the instrument operator’s ground segment, while value adding was the domain of the user

➢ Today the multi-disciplinarity of Earth Sciences demands much broader data integration than in the past, at all Levels!

CEOS processing levels

In a nutshell:

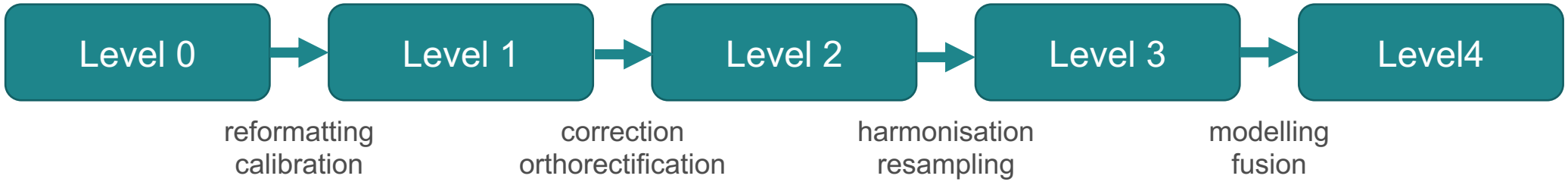
- linear sequence in which one level builds on the previous one
- rooted in ground segment technology of the 1990s
- somewhat biased towards spectroradiometric sensors
- never fully harmonised across sensor types and agencies



European Commission

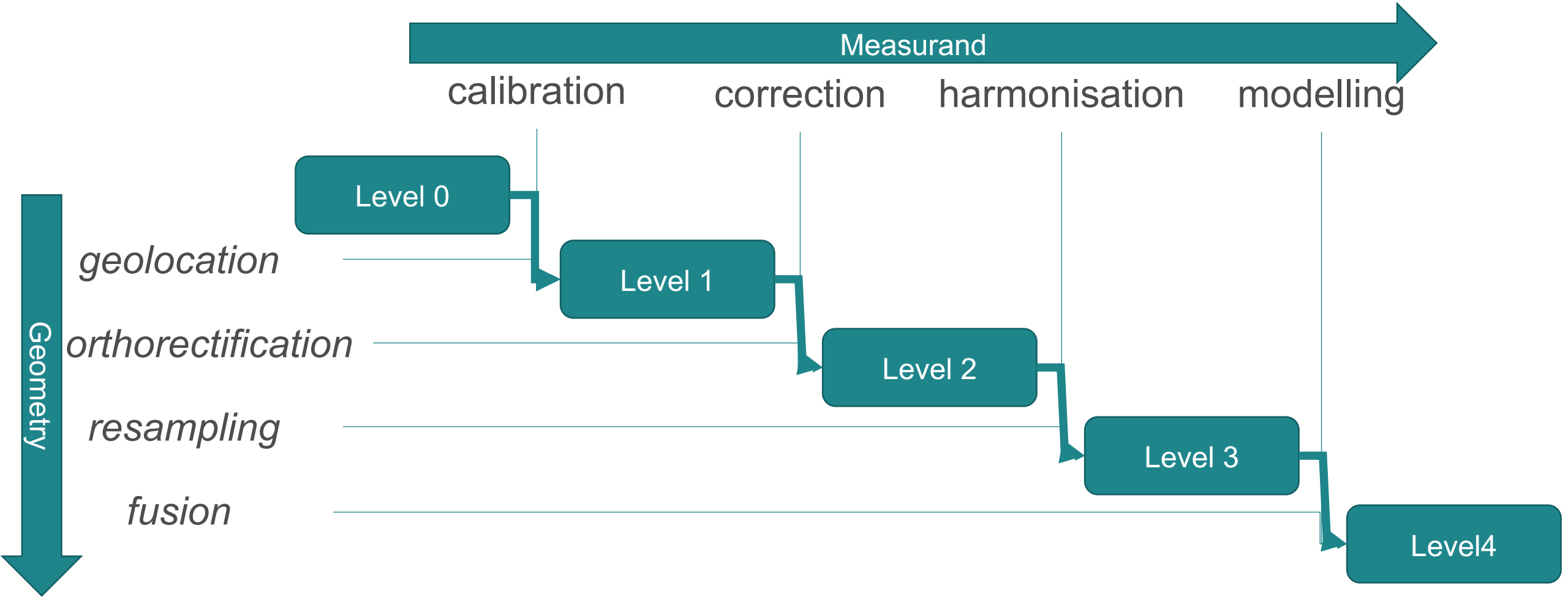
Disentangling the processing steps

Separate between interventions affecting the measurand and those affecting geometry



Disentangling the processing steps

Separate between interventions affecting the measurand and those affecting geometry



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 ancilliary data …
- **Level 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.
- **Level M/2 (target calibrated):** Level M/1 data processed to represent **geophysical properties (values and uncertainties) for a specified target** (object, feature of interest, e.g. surface reflectance, apparent temperature) derived (exclusively?) from M1 sensor data, as much as possible maintaining the sensors nominal spatial and temporal sampling (observation preserving).
- **Level M/3 (homogenised*):** 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 **values are agnostic of the originally acquiring sensor** 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.
- **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).

*For a definition of ‘homogenisation’ and disambiguation with ‘harmonisation’ see: https://research.reading.ac.uk/fiduceo/glossary/

The geometry dimension G

- **Stage G/A (raw):** individual observations (samples) are not geolocated
- **Stage G/B (geolocated):** Each observation is geolocated with documented uncertainty in a (traceable) 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’).
- **Stage G/C (georectified/gridded):** Observations have been spatially re-sampled to fall within a specified, usually regular, geodetic grid.
- **Stage G/D (regridded1):** Observations have been re-sampled from the original geodetic grid into another specified (geodetic) grid.
- **Stage 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

Old Levels in a new perspective

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

Measurand / Geometry	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

A new Processing Level matrix

Preliminary labelling scheme:

- Numbers are used to identify Measurand steps
- Letters indicate spatio(temporal) Geometry steps

Labelling scheme could also be revised to clearly distinguish from classical CEOS Levels. Colour coding is only indicative to illustrate possible suitability e.g. for spectroscopic data.

Measurand / Geometry	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

Future interoperability levels (tentative)

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

Summary & Conclusions

The classical linear Level scheme might benefit from an overhaul. Separating geometric (spatiotemporal discretisation?) from measurand interventions offers:

1) Possibilities for more flexible and comprehensive lay-out of processing paths (e.g. to track and minimise uncertainty contribution)

2) Clear and unambiguous indexing of Levels which allows easy referencing and traceability (4x4 states = 4bits permits storage even per measurement)

3) Accommodation of non-imaging sensors and ‘in-situ’ observations, thus not just classical satellite based EO

4) Dedicated standardising (‘ARD’) at each level for increased interoperability

➢ **tailored Analysis Ready Data**!

Thank’s for your attention!
Questions?

Peter.Strobl@ec.europa.eu

European Commission