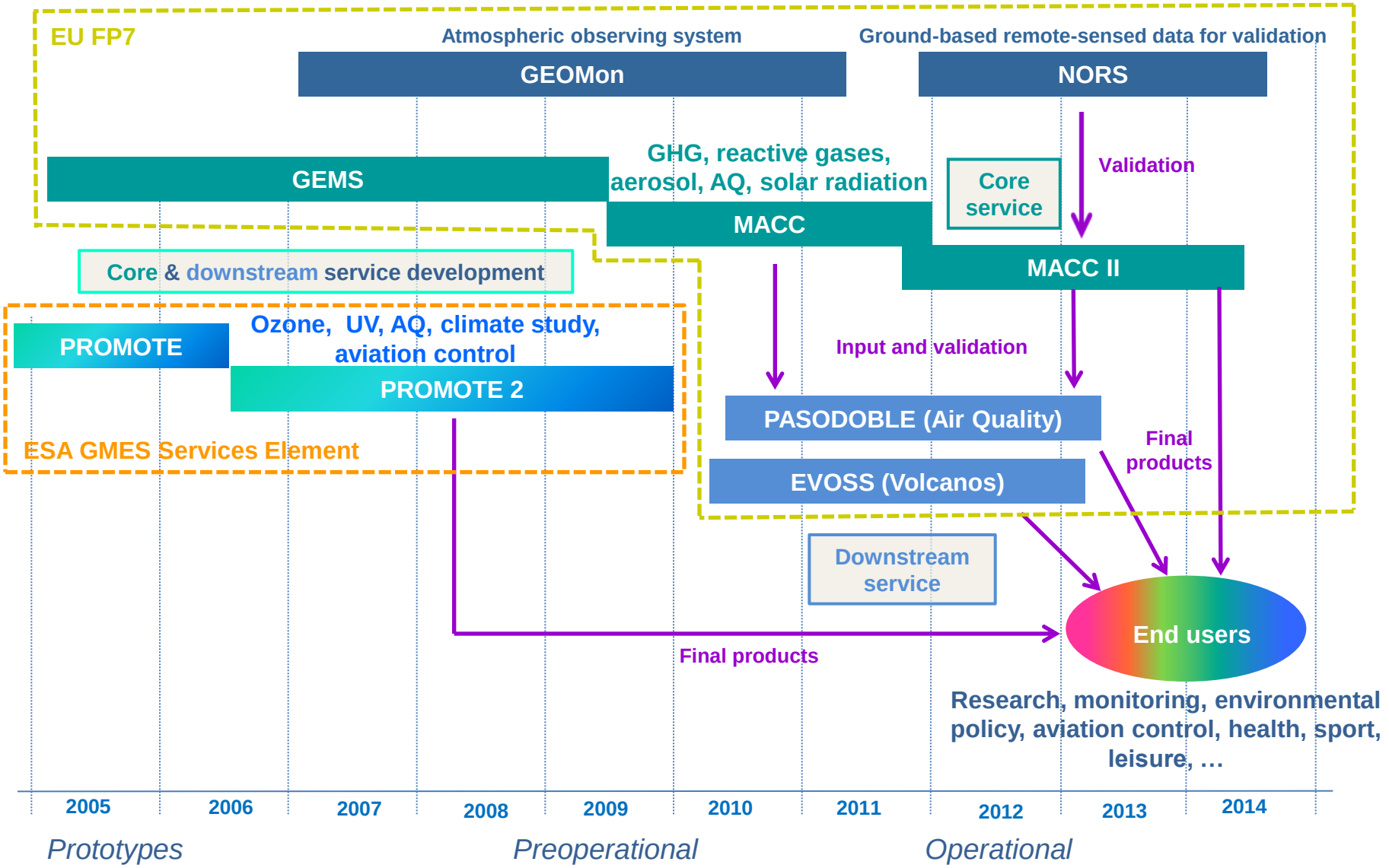
PASODOBLE – INSPIRE metadata mapping

Anne De Rudder

Belgian Institute for Space Aeronomy (IASB-BIRA), Brussels

Projects contributing to the GMES Atmospheric Service (GAS)



Preliminary questions, options, decisions

Granularity

What is the type of object we want to describe with one given metadata set?

In other terms, if a dataset is the entity we want to describe, what do we define as a dataset?

Note. *In INSPIRE, there are 3 resource types : dataset, dataset series (unclear what the difference is), service.*

- One data file (e.g. metadata in file header as in CF NetCDF)?
- One set of files containing the same type of data?
- A continuously updated dataset? → Do we want dynamical metadata to be automatically generated « on the fly » (retrieved from the data file)?
- May a dataset include several dependent variables? Ideally, it should. This has implications on the number of allowed values of other metadata fields, e.g. unit, instrument, domain of values, quality indicators, etc.

Preliminary questions, options, decisions (cont.)

Purpose

What is the metadata intended for?

- To be integrated into an existing framework (e.g. INSPIRE)?
- Underpin a tailored catalogue (possibly searchable by keyword) : discovery metadata?
- Inform the user on the verge of using the data (serve as a manual)? What does the user need to know in order to correctly read, understand and use the data?

Preliminary questions, options, decisions (cont.)

Volume

How detailed (= voluminous) do we want the metadata scheme to be?

Balance between

- Requirements related to the metadata purpose
- Technical implementation constraints
- Amount of work required to build up the conceptual scheme, then any tool based on that scheme
- Amount of work required from the metadata providers

Preliminary questions, options, decisions (cont.)

Compliance with standards

Which standards? What are the implications?

- Complying with standards and common practices ensures efficient communication with the outside world
- When the standard is well designed, complying with it provides some guarantee on the metadata quality (e.g. QA4EO)
- Standards help internal harmonisation as they provide a ready-made framework
- Complying with a standard may be a challenge when the standard is not perfectly adapted to its target
- Do we agree on the terminology? E.g. in INSPIRE, space, scale, accuracy have different – and sometimes opposite – meanings than in Earth science

Metadata quality

What do we need to care for in order to get good (pertinent, useful, complete) metadata?

- A good metadata scheme does not guarantee good metadata! − It does not even guarantee compliance with some adopted standard
- Provide clear definitions of metadata elements (+ possibly examples)
- Where possible, provide lists of preselected metadata values
- Make room for atypical cases by offering options such as
  - *other*
  - *not applicable*
  - *unknown (yet)*
  - [void] ↔ either appears as *not provided* in the output or does not appear at all

  Applies even to *mandatory* metadata fields, except if the field value determines which branch of the scheme must be followed (e.g. if the resource is a dataset, the applicable metadata fields will be different from those for a service)
- Avoid redundancies

INSPIRE discovery metadata

- Described in the INSPIRE Metadata Regulation (*)
- Includes the set of top-level metadata supposed to apply to any *spatial* resource − Completed by data specifications – one for each INSPIRE data theme – there is one for *Atmospheric Conditions* and *Meteorological Geographical Features*

(*) Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, Official Journal of the European Union, December 4, 2008.

PASODOBLE detailed metadata

A Product description
- A1 Identification and nature
- A2 Domain of definition
- A3 Domain of values
- A4 Quality
- A5 History
- A6 Presentation, coding, format
- A7 References

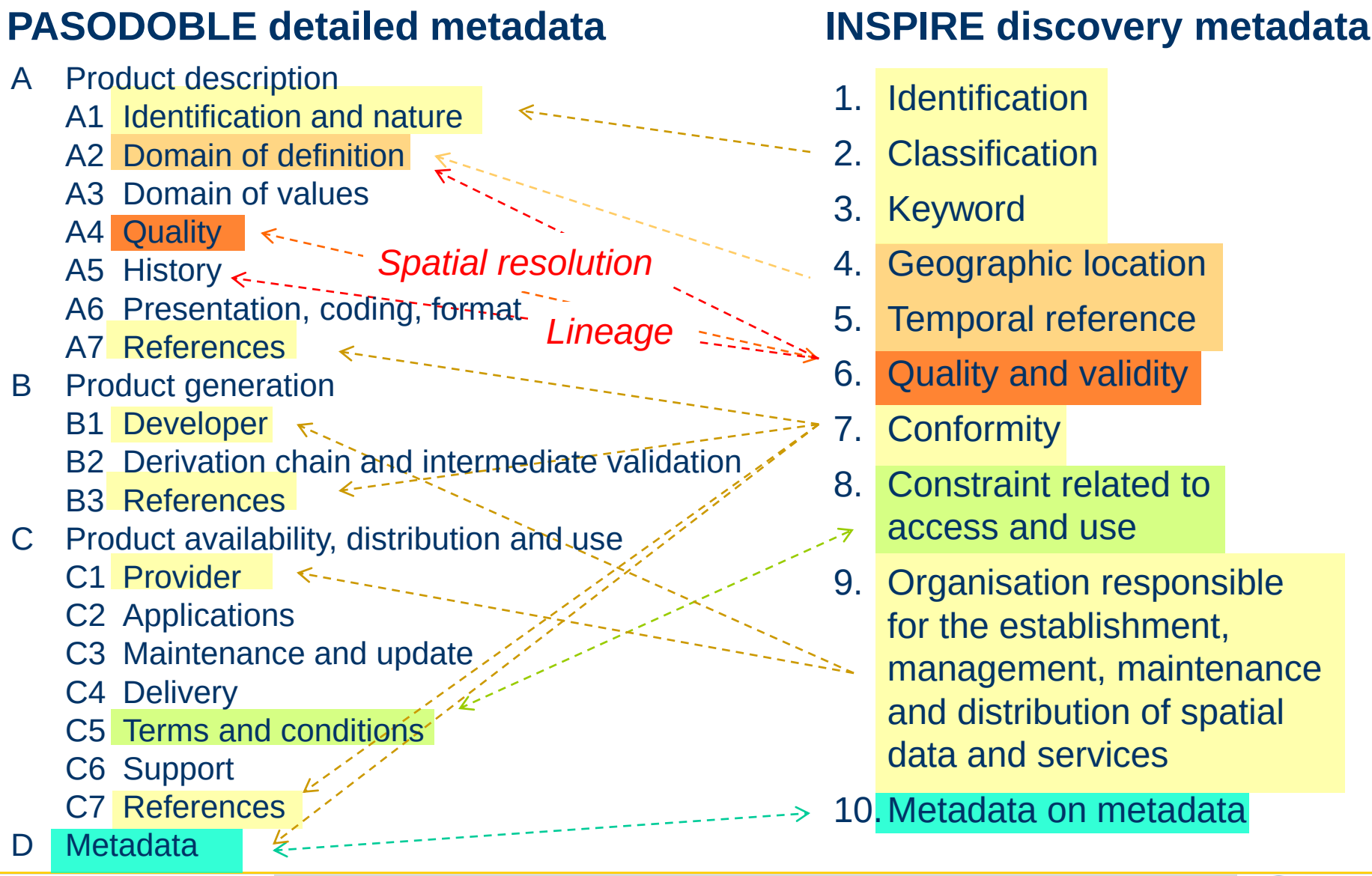
B Product generation
- B1 Developer
- B2 Derivation chain and intermediate validation
- B3 References

C Product availability, distribution and use
- C1 Provider
- C2 Applications
- C3 Maintenance and update
- C4 Delivery
- C5 Terms and conditions
- C6 Support
- C7 References

D Metadata

\+ variable names harmonisation (CF-like)

PASODOBLE versus INSPIRE



PASODOBLE detailed metadata

- A Product
  - A1 Identification and nature
  - A2 Domain of definition
  - A3 Domain of values
  - A4 Quality
  - A5 History
  - A6 Presentation, coding, format
  - A7 References
- B Product generation
  - B1 Developer
  - B2 Derivation chain and intermediate validation
  - B3 References
- C Product availability, distribution and usage
  - C1 Provider
  - C2 Applications
  - C3 Maintenance and update
  - C4 Delivery
  - C5 Terms and conditions
  - C6 Support
  - C7 References
- D Metadata

A4 Quality:

- Existence of quality assessment
- Validation method
- Validation datasets
- Validation results
- Content analysis
- Quality indicators

↕ a major user requirement

QA4EO A QUALITY ASSURANCE FRAMEWORK FOR EARTH OBSERVATION

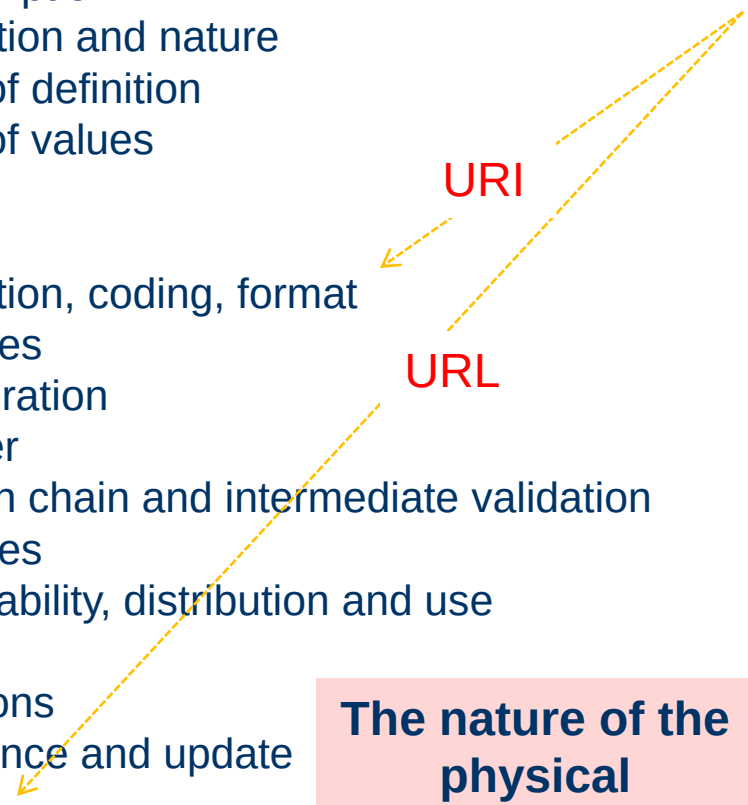
INSPIRE top-level metadata compliance issue

identification of horizontal resolution with data quality

PASODOBLE versus INSPIRE (cont.)

PASODOBLE detailed metadata

- A Product description
  - A1 Identification and nature
  - A2 Domain of definition
  - A3 Domain of values
  - A4 Quality
  - A5 History
  - A6 Presentation, coding, format
  - A7 References
- B Product generation
  - B1 Developer
  - B2 Derivation chain and intermediate validation
  - B3 References
- C Product availability, distribution and use
  - C1 Provider
  - C2 Applications
  - C3 Maintenance and update
  - C4 Delivery
  - C5 Terms and conditions
  - C6 Support
  - C7 References
- D Metadata



The nature of the physical quantities and their units are not documented in INSPIRE

INSPIRE discovery metadata

1. Identification
2. Classification
3. Keyword
4. Geographic location
5. Temporal reference
6. Quality and validity
7. Conformity
8. Constraint related to access and use
9. Organisation responsible for the establishment, management, maintenance and distribution of spatial data and services
10. Metadata on metadata

Atmospheric data

Atmospheric conditions are numerically described through the values of physical atmospheric variables (or parameters).

Physical atmospheric variables include

- concentrations of charged particles (ions, electrons) in air;
- absolute or relative amounts (expressed as molecular or mass concentrations, or mass or volume mixing ratios, or partial pressure) of airborne gaseous chemical species (molecules, radicals, atoms);
- these amounts integrated along some spatial direction (e.g. vertical columns);
- concentrations of airborne liquid and solid particles;
- aerosol and cloud optical properties;
- dynamical variables (air temperature, pressure, density; wind direction and speed, turbulence; absolute or relative humidity; precipitation; etc.).

Physical atmospheric variables are governed by **physical phenomena** such as

- electromagnetism;
- photochemical reactions between molecules or molecules and photons;
- heterogeneous chemical reactions;
- cloud microphysics;
- phase transformation;
- energy transfer;

and by **boundary inputs, outputs or conditions** such as

- the Earth's magnetic field;
- solar and cosmic radiation;
- terrestrial radiation;
- molecular and energy exchanges (fluxes) with the ocean, land, cryosphere, biosphere and man-made apparatus;
- surface albedo.

Conclusions

- The main obstacles encountered in mapping scientific metadata to INSPIRE metadata are related to
  - the lack of an appropriate conceptual framework
    - datasets = functions of independent variables
    - we know that what we see is not the truth (accuracy is not guaranteed by a fine resolution)
  - the fact that the use of already existing pieces of software tends to be preferred to building up new tools
    - the prevalence of formal aspects (allowing easy automation) on the information content
- A new smaller INSPIRE / ISO compliant scheme has been derived for PASODOBLE from the detailed scheme, as a rescue solution
- The INSPIRE Metadata Regulation might be reviewed at some point…


Thank you