Evaluation and Quality Control for the Copernicus Seasonal Forecast Systems

Nicolau Manubens

Computer Scientist

Barcelona Supercomputing Center
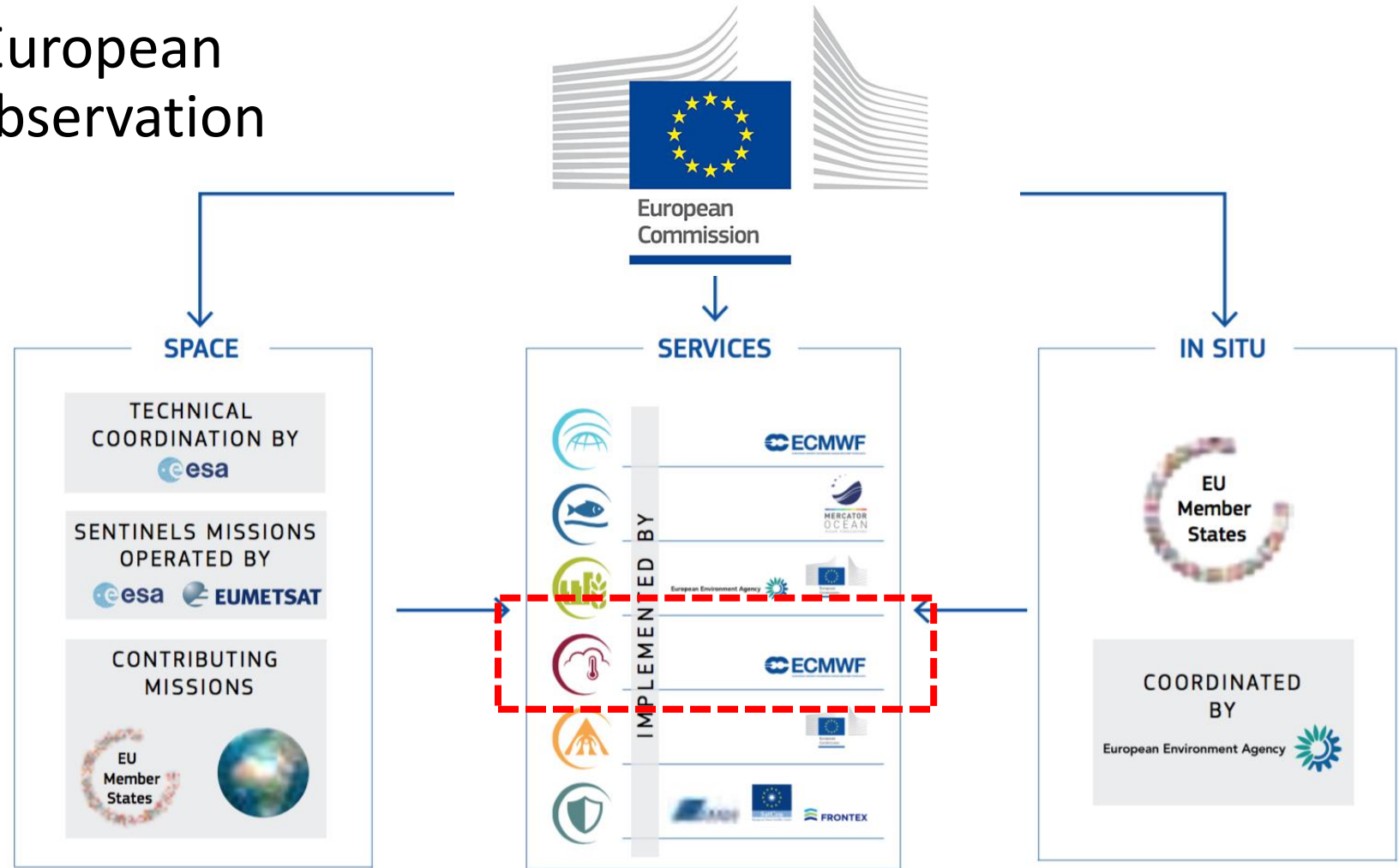
nicolau.manubens@bsc.es

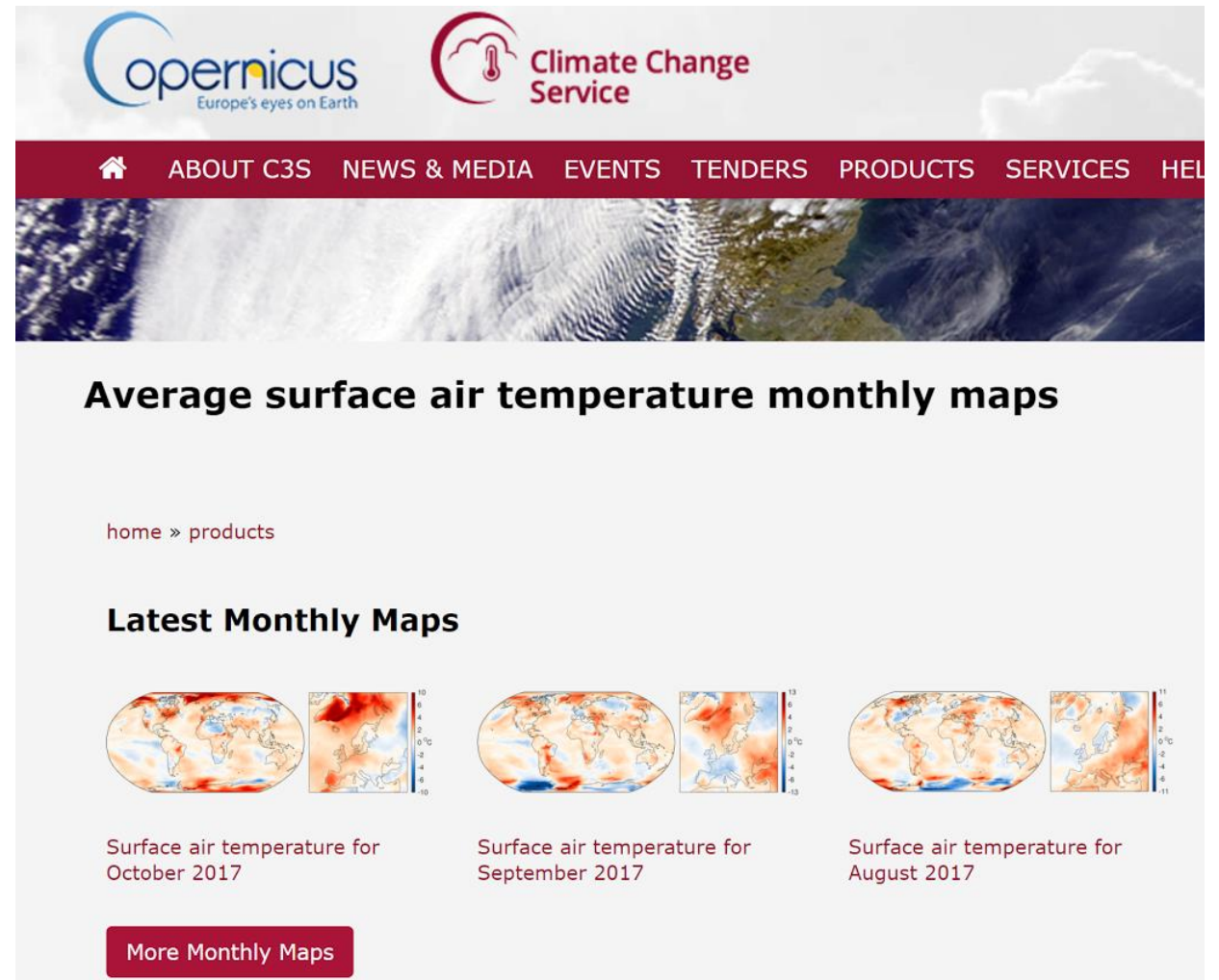ESIP IQC monthly telco

The Copernicus Programme

- Establishment of a European capacity for Earth Observation

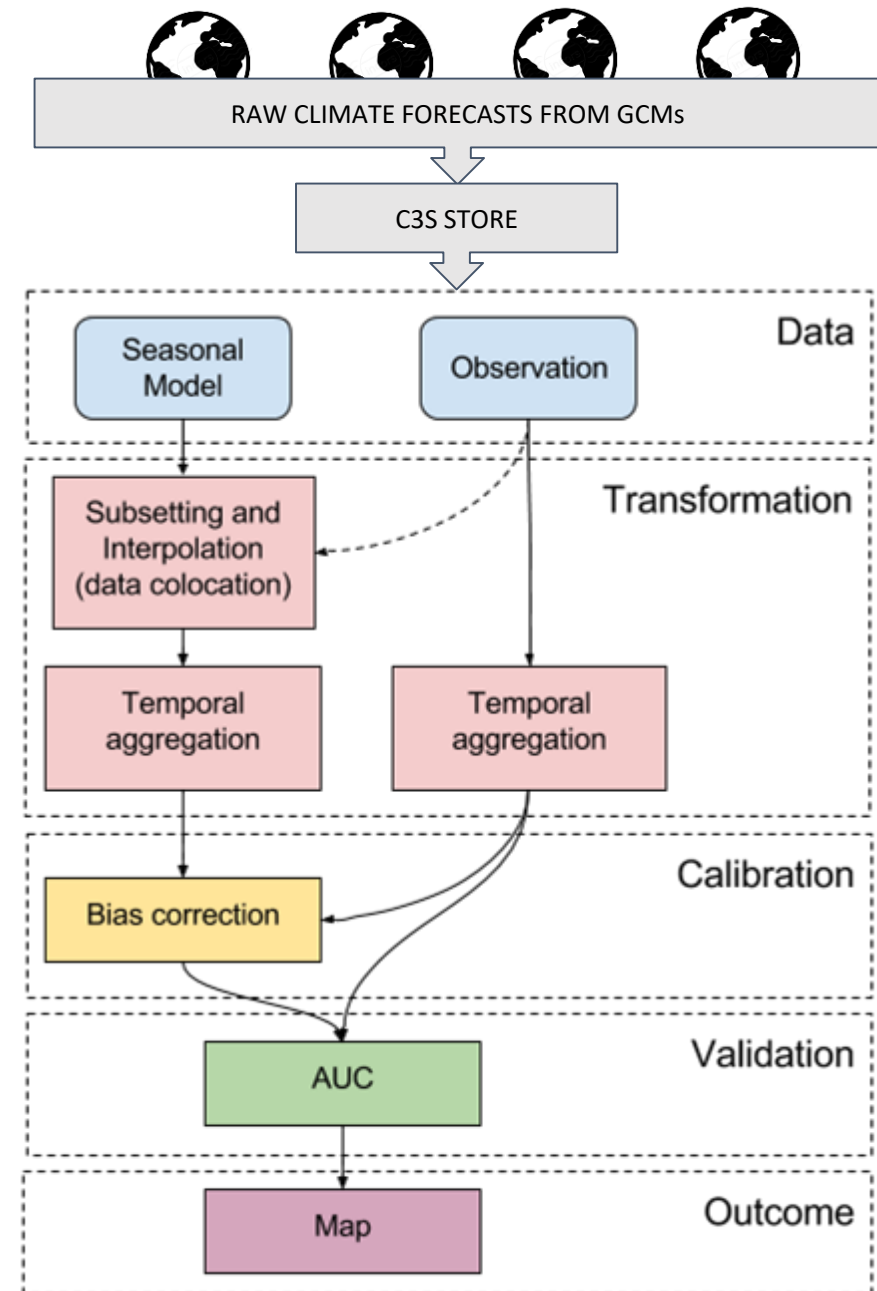# The Copernicus Climate Change Service (C3S)

- The C3S will openly provide authoritative information about past, present and future climate

- Targetting the private and public sectors, e.g. for decision making in the energy sector
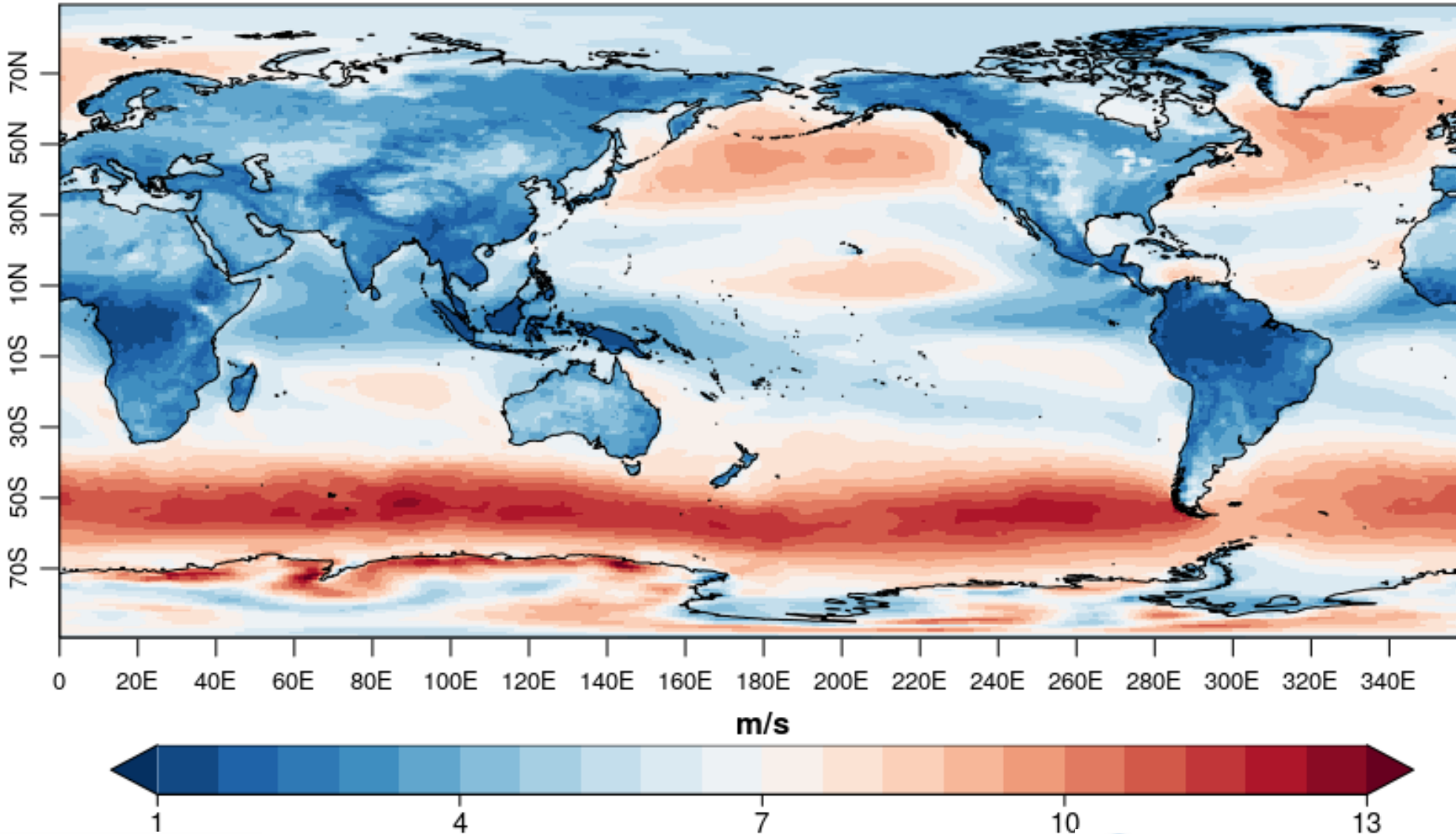
https://climate.copernicus.eu

Quality assurance strategy for seasonal forecasts

RAW CLIMATE FORECASTS FROM GCMs

C3S STORE

Data
- Seasonal Model
- Observation

Transformation
- Subsetting and Interpolation (data colocation)
- Temporal aggregation
- Temporal aggregation

Calibration
- Bias correction

Validation
- AUC

Outcome
- Map

EXAMPLE C3S PRODUCT GENERATION WORKFLOW

European Commission

Copernicus

ECMWF

Example forecast product (no QC)

ECMWF System 4 ensemble mean forecast of DJF average wind speed (issued in November 2017)

Quality assurance strategy for seasonal forecasts

RAW CLIMATE FORECASTS FROM GCMs

C3S STORE

Data
- Seasonal Model
- Observation

Transformation
- Subsetting and Interpolation (data colocation)
- Temporal aggregation
- Temporal aggregation

Calibration
- Bias correction

Validation
- AUC

Outcome
- Map
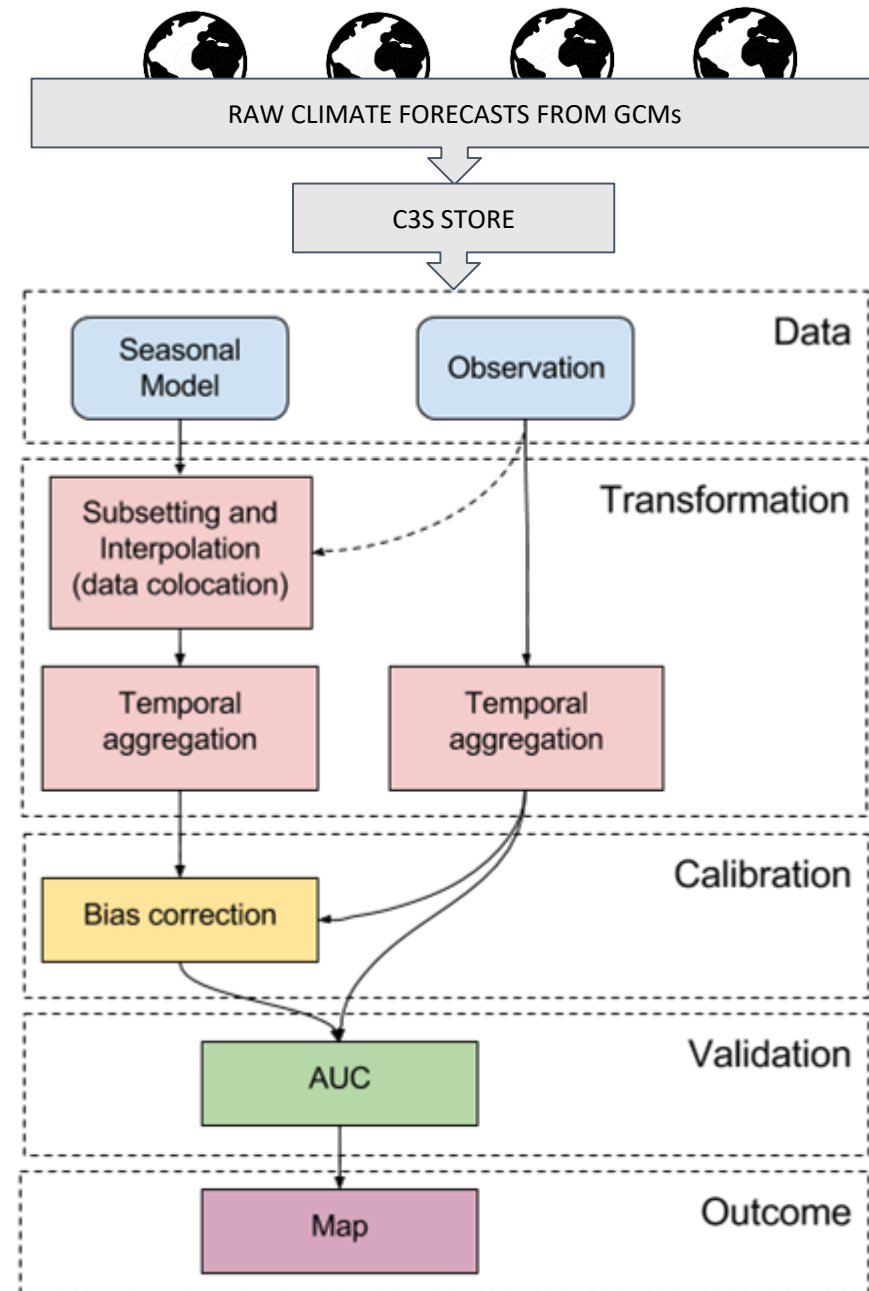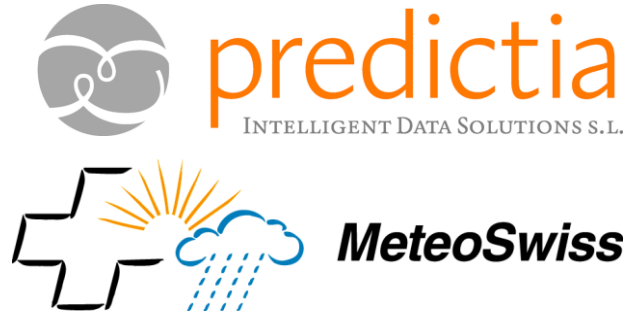
EXAMPLE C3S PRODUCT GENERATION WORKFLOW

Principles of a climate service

- According to the Climate Services Partnership's Ethical Framework for Climate Services[1]:

    "Climate service providers should consider the consequences of their actions for those who may use or be affected by the use of climate service products."

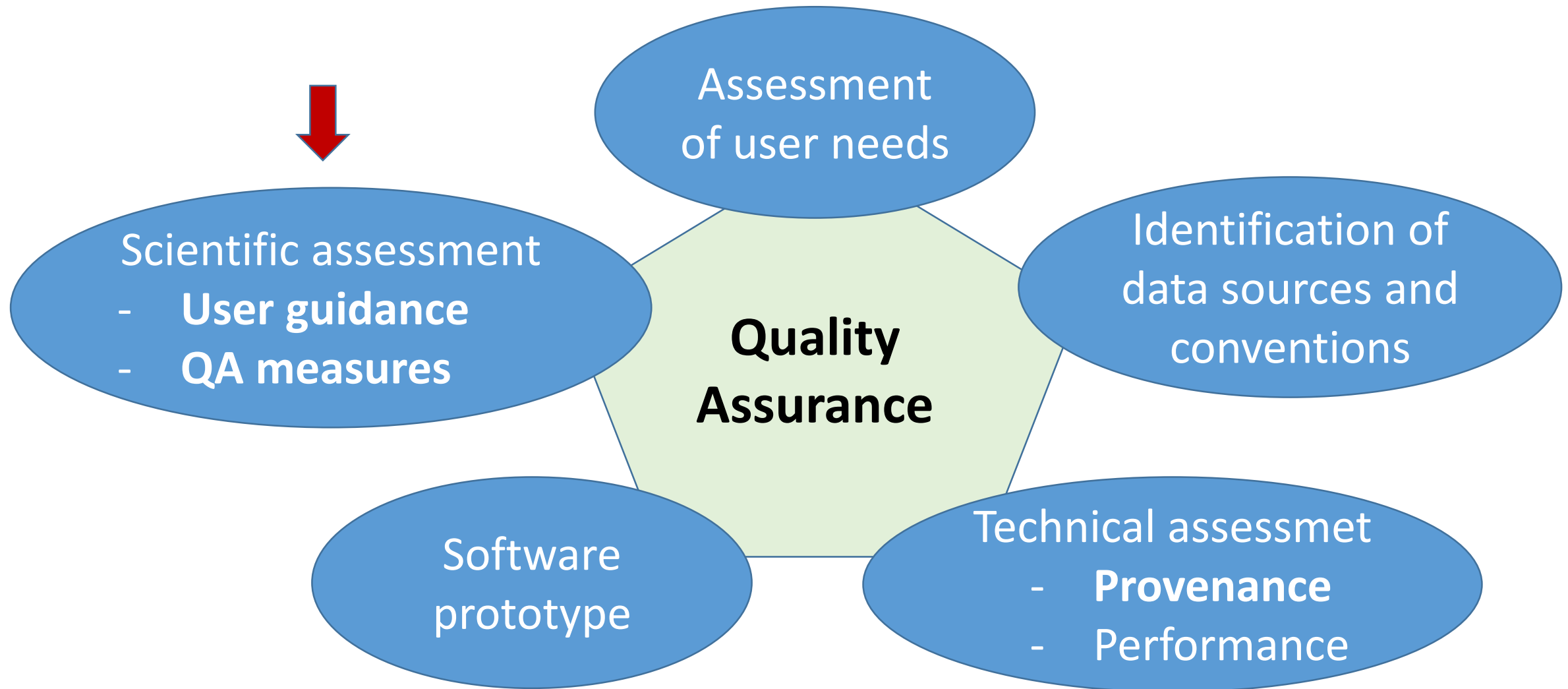   "Climate service products should be open to scrutiny and comparison."

[1] www.climate-services.org/ethics

Quality assurance strategy for seasonal forecasts

Scientific assessment

- Scientific assessment activities include:

  - Definition of a set of **forecast quality measures** needed
    - Different forecast types
    - Different forecast quality aspects

  - Elaboration of **rhetoric questions for end users to understand** the meaning/value of the quality measures

  - Identifying sources of uncertainty and methods to quantify

  - **Recommendations on EQC graphical products**

Example forecast product (with QC)
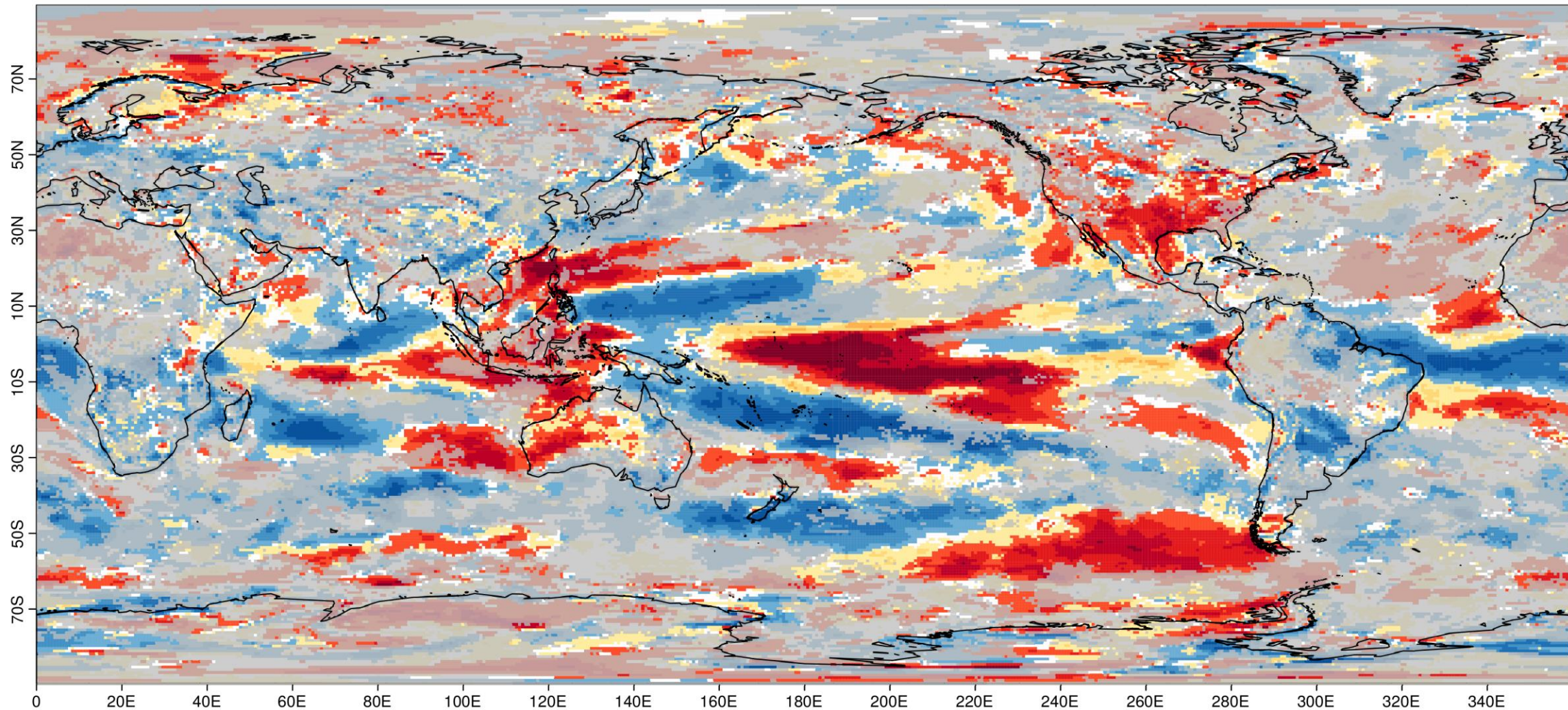
Quality assurance strategy for seasonal forecasts

Assessment of user needs

Identification of data sources and conventions

Scientific assessment
- **User guidance**
- **QA measures**

Quality Assurance

Software prototype

Technical assessmet
- **Provenance**
- Performance
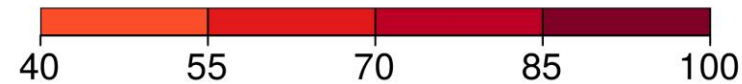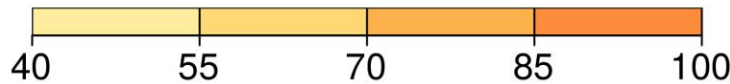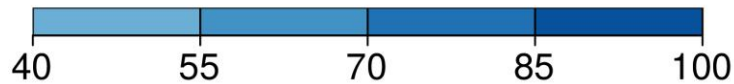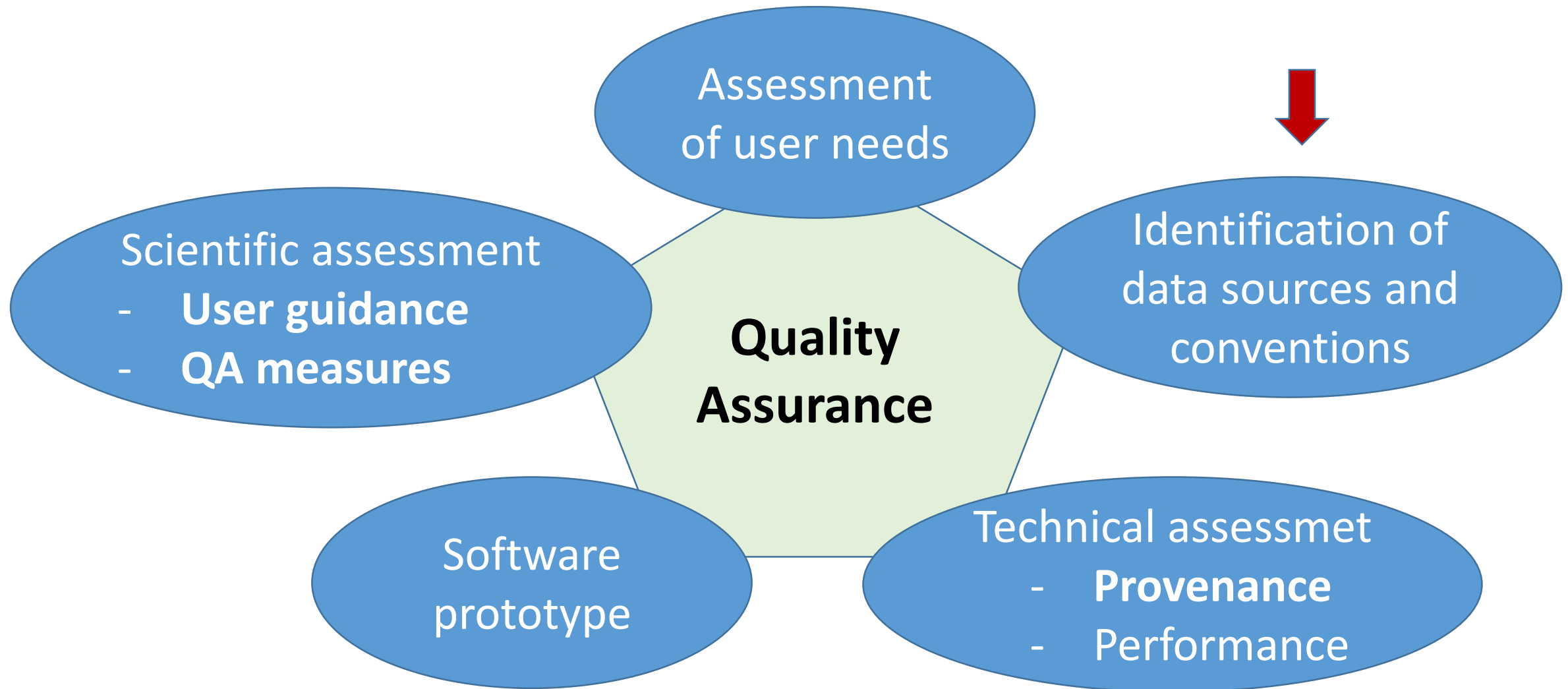
Data sources and conventions

- Essential Climate Variable inventory.

- Quality Brief – collection of quality aspects of a dataset which the user should be informed about.

- C3S data conventions

Data sources and conventions

- The C3S data conventions cover:
  - File content
  - File format
  - File structure
  - Metadata

- Based on **CF conventions**, **SPECS**, **CMIP**, and **ACDD**
  - NetCDF
  - File naming and distribution
  - Attributes (e.g. **source**, **commit**)
  - Spatial and time coordinates and axes: names, valid_min/max, units, bounds
  - Variables: names, units, dimension names, cell methods

Quality assurance strategy for seasonal forecasts

Assessment of user needs

Scientific assessment
- **User guidance**
- **QA measures**

Identification of data sources and conventions

Quality Assurance

Software prototype

Technical assessmet
- **Provenance**
- Performance
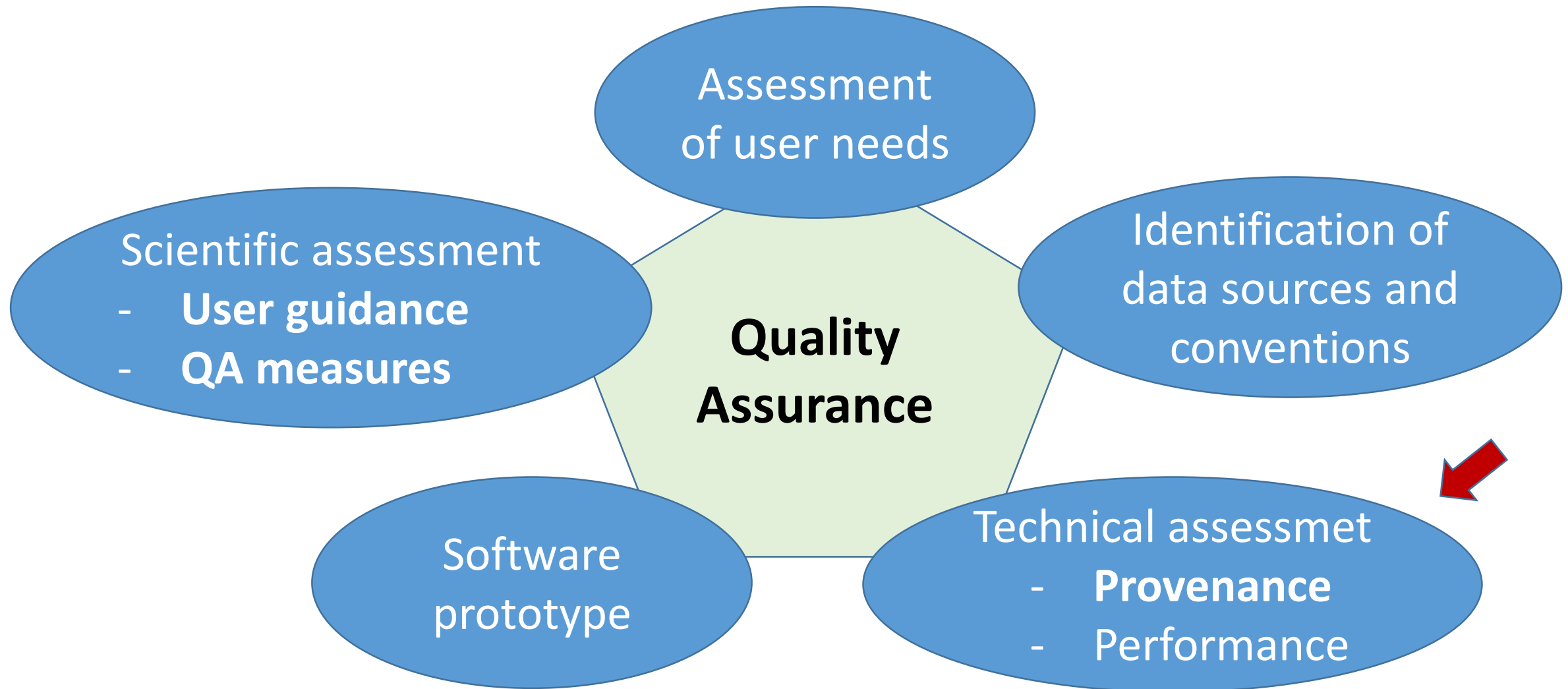
Technical assessment - Provenance

- **Metaclip** is a metadata and provenance model. It allows encoding meta-data and provenance as a graph

  http://metaclip.predictia.es

- The graph can be serialized and attached to products

- Metaclip info. of a product can be visually inspected
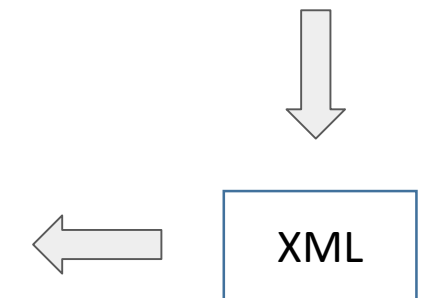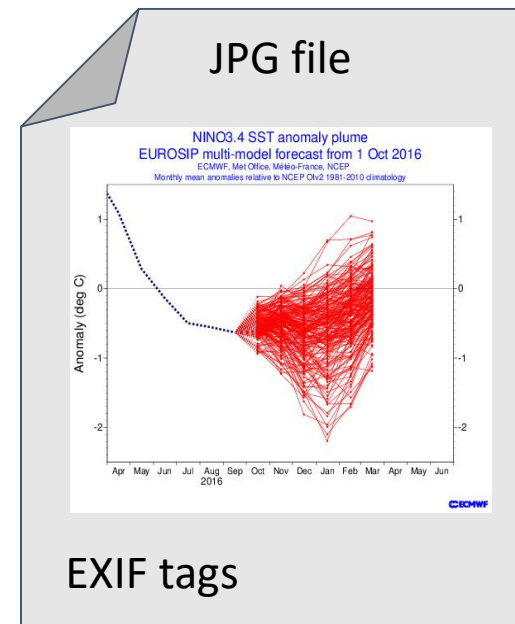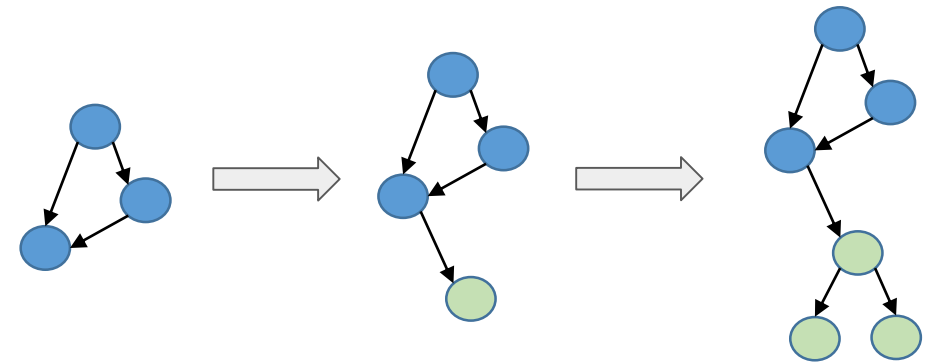
The metadata and provenance information is delivered within the image

Technical assessment - Provenance

METACLIP DEMO VIDEO REMOVED FOR SHARING PURPOSES. YOU CAN DOWNLOAD A VERSION OF THE OVERHEADS WITH THE ATTACHED VIDEO AT

https://drive.google.com/file/d/15VBG5jRS6f2apkkRjHYwFHC4uDv96nBJ/view?usp=sharing

Technical assessment - Provenance

- Based on the **W3C RDF** (Resource Description Framework)

  Reusing existing standard ontologies developed by the community:
  PROV, **SEQ**, **GEO**

- Modular and extensible

- Abstract

- Provides meaningful provenance traces

Technical assessment - Performance

- EQC information should be timely. Efficiency and computing performance of verification is being evaluated in controlled computing environments.
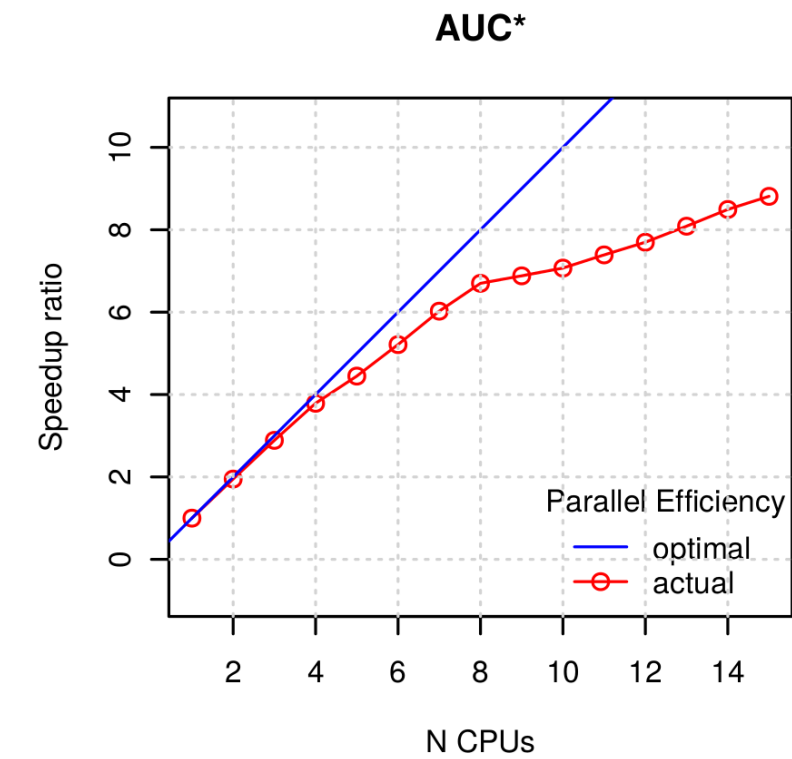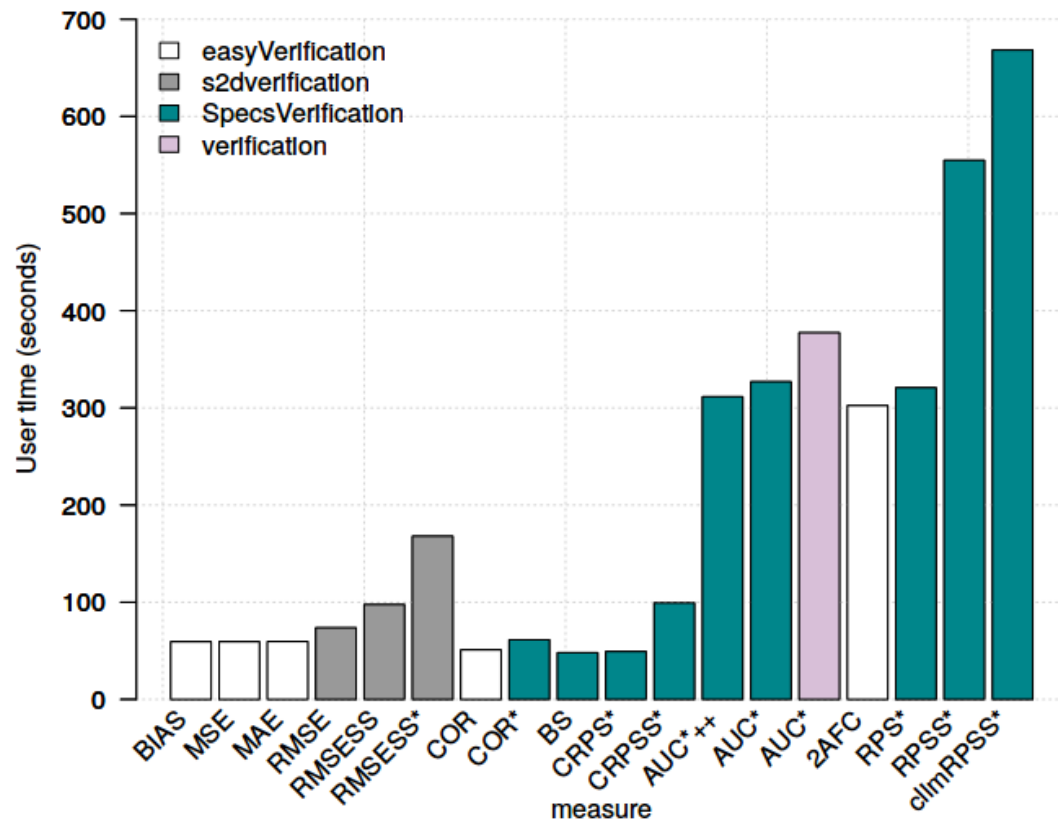
J. Bedia, D. Sanmartín (PREDICTIA)

Quality assurance strategy for seasonal forecasts

Quality Assurance

Assessment of user needs

Identification of data sources and conventions

Scientific assessment
- **User guidance**
- **QA measures**

Software prototype

Technical assessmet
- **Provenance**
- Performance
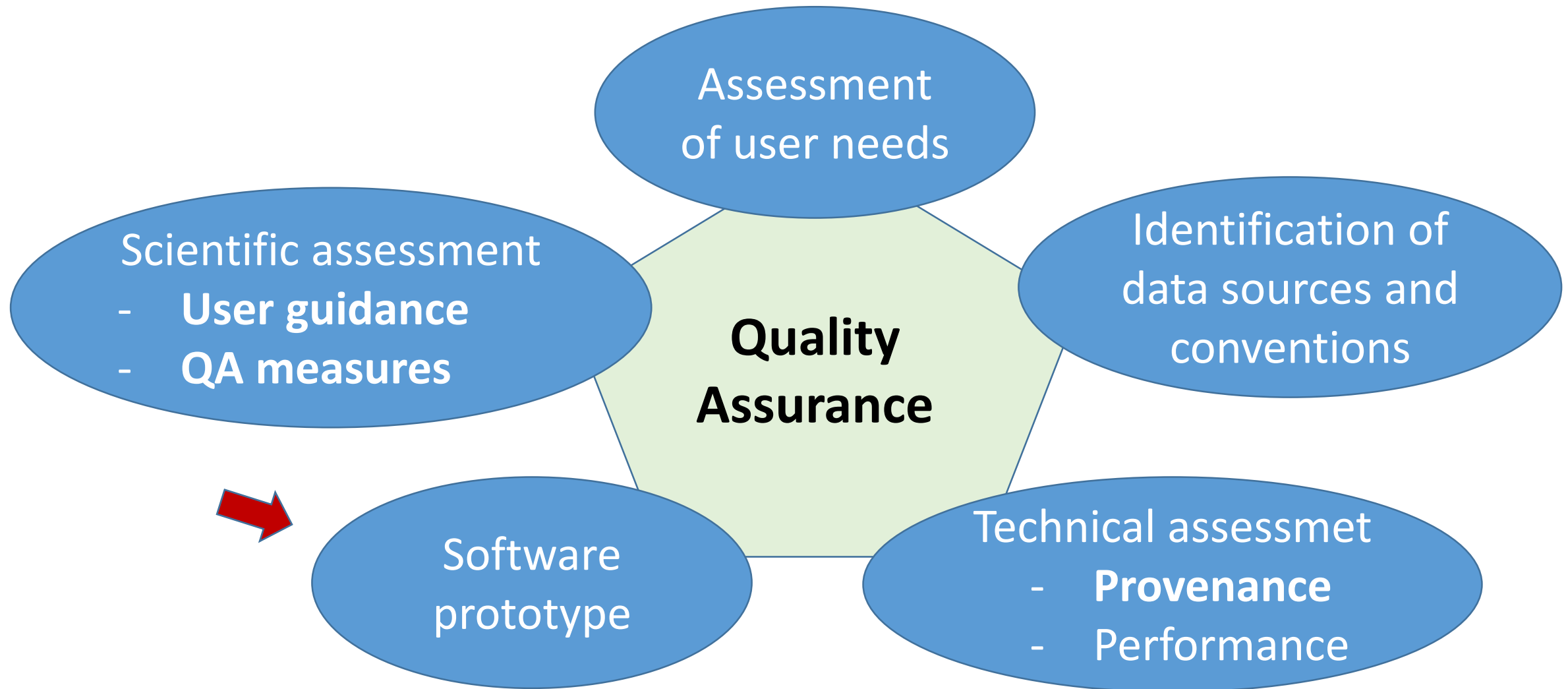
Prototype

- Proof-of-concept software

- Existing open-source R packages are being evolved

- Performance ('timeliness') is also being addressed

Quality assurance strategy for seasonal forecasts

Quality Assurance

Assessment of user needs

Identification of data sources and conventions

Scientific assessment
- **User guidance**
- **QA measures**

Software prototype

Technical assessmet
- **Provenance**
- Performance
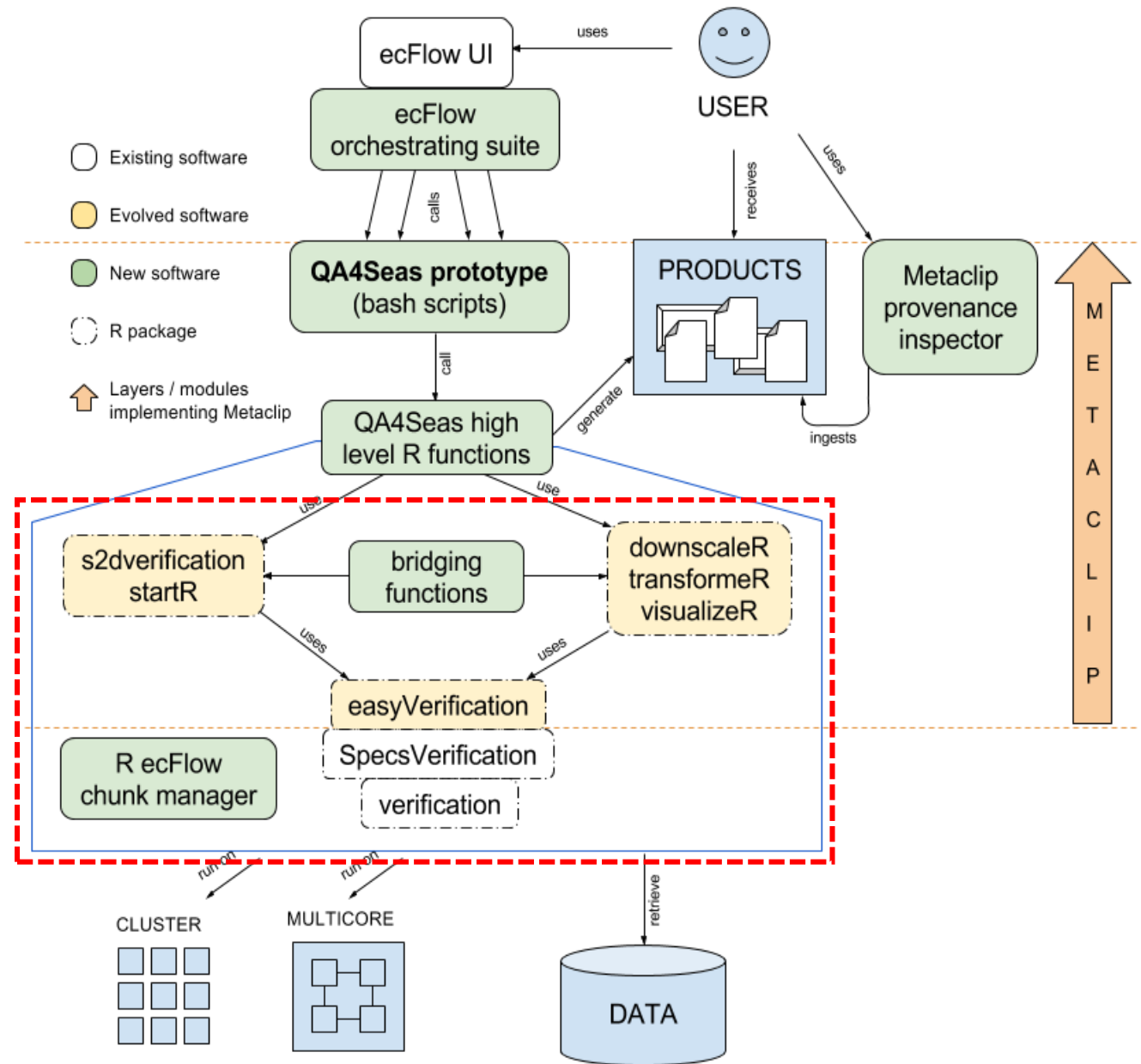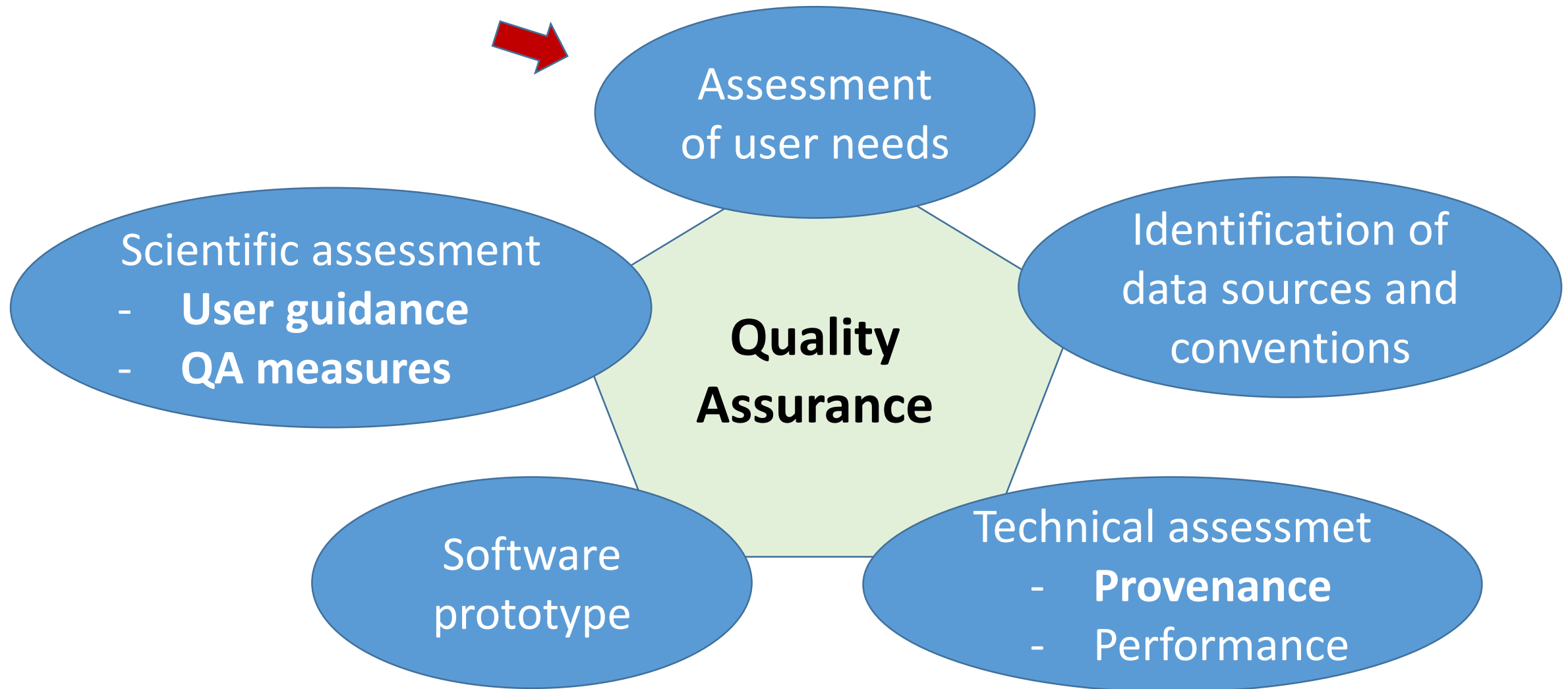
Assessment of user needs

- First consultation stage – carried out
  - Survey
    - A survey has been conducted, on themes such as:
      - Use of ECVs and CIIs
      - Using and accessing climate forecasts
      - Post-processing
      - Uncertainty
      - Metadata and software tools
    - ~60 respondents

  - Interviews
    - An interview protocol has been established. Same themes as survey
    - ~10 respondents

Assessment of user needs

Relative importance of forecast quality attributes

Legend: 5 - Very important | 4 | 3 | 2 | 1 - Not important at all | Don't know

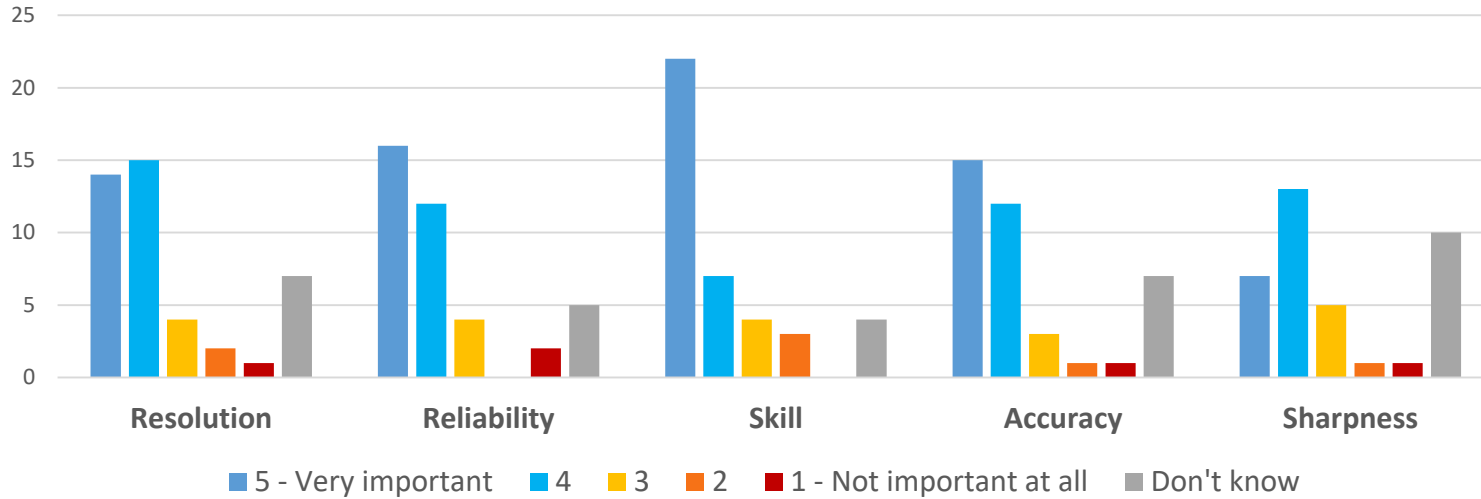Categories: Resolution, Reliability, Skill, Accuracy, Sharpness

Interest in sources of uncertainty in seasonal forecasts

Legend: 5 - Very interested | 4 | 3 | 2 | 1 - Not interested at all | Not sure

Categories: Initial state of the climate system, Numerical models' assumptions/formulations, Observations (including reanalysis), Post-processing, The full chain of uncertainty

M. Soares, A. Taylor (Univ. Leeds)
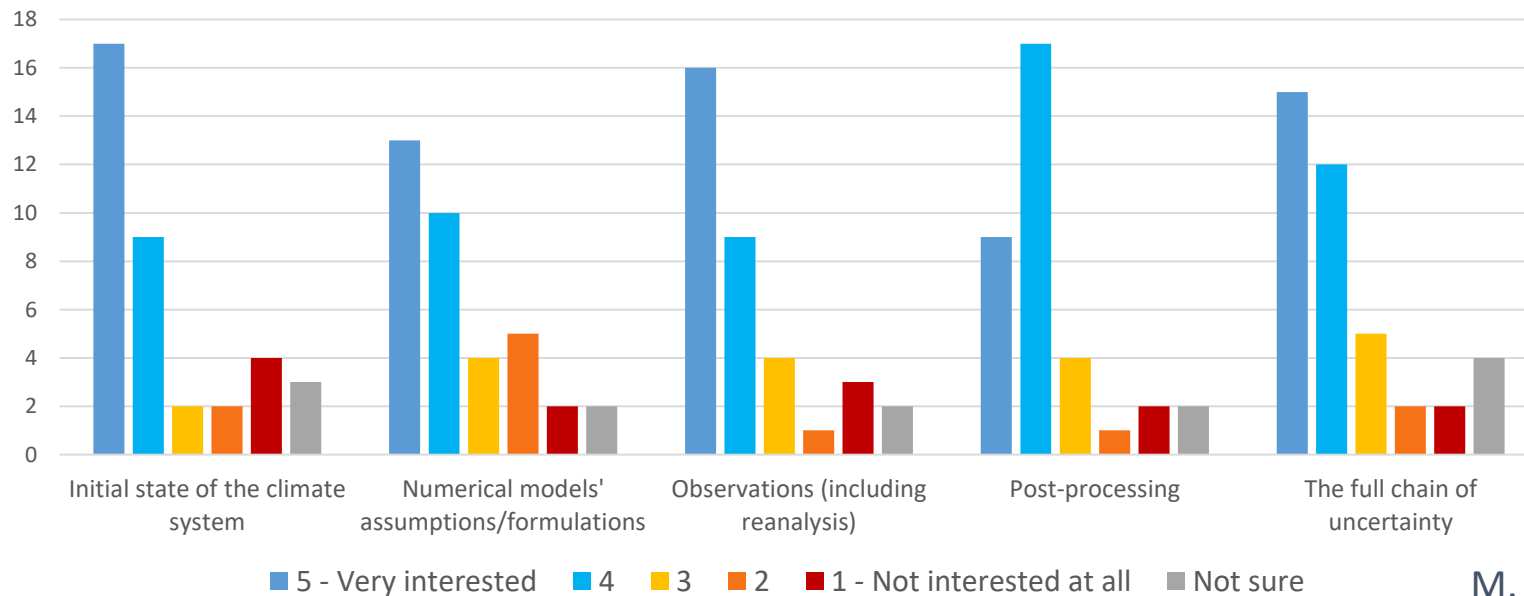
Assessment of user needs

- Second consultation stage – starting soon

  - Survey

  Please send an email to the following address if you would like to participate. The survey should take approximately 20 minutes.

  nicolau.manubens@bsc.es

Points in common with the ESIP IQC activities

- According to AGU Fall Meeting 2017 ESIP IQC abstract:

"What was discovered to be most lacking is the transparency of data lineage (i.e., provenance and maturity), uniform methods for uncertainty characterization, and uniform quality assurance data and metadata. While solutions to these types of issues exist, most data producers have little time to investigate and collaborate to arrive at and conform to a consensus approach."

We have worked on all of these aspects and we are looking for collaboration and interaction to tend as much as possible to a consensus approach.

Points in common with the ESIP IQC activities

- Technical topics on ESIP IQC wiki:
  - Quality aspect terminlogy/taxonomy
  - Data intercomparison and quality harmonization
  - Best practices in different communities
  - Level 2 Quality flags (and what do they tell about the product quality)
  - Level 3 "Quality" (what's that?)
  - Propagation of L2 uncertainty to L3
  - Effects of missing data on aggregates
  - Data quality ontology (The Semantic Web Cluster is already working on logistics of the Data Quality Ontology breakout session)
  - Presenting data quality to users

Points in common with the ESIP IQC activities

- ESIP IQC maturity framework (Ramapriyan, Peng, Moroni, Shie, 2017)
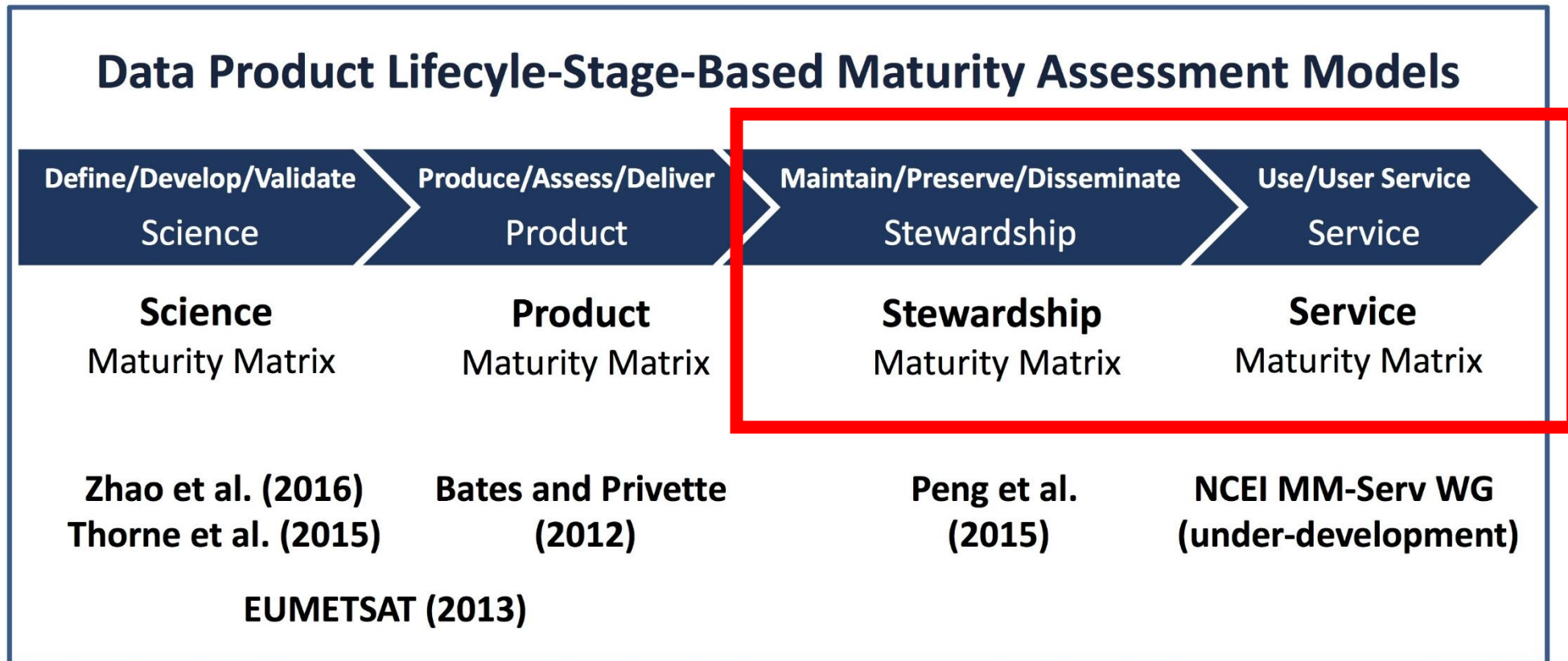
G. Peng, 2018

Conclusions

- **A quality control and provenance strategy for climate seasonal forecasts has been shown from a producer's perspective**

- The solution consists in defining **quality measures** for the products and in establishing a **provenance mechanism**

- **User guidance and data conventions** are fundamental

- An extensible **provenance framework called Metaclip** has been shown

- Open-source software for the QC of seasonal forecasts is being developed

Conclusions

- The work done is tightly related to ESIP IQC activities

- **IQC recommendations** have not been identified soon enough, but **should become a reference point for** upcoming iterations of **the C3S/Copernicus EQC strategy** (presented contract ending soon)

- ESIP IQC seems an appropriate place for the discussions we have been looking for on provenance and other QC aspects. We would like to stay close to the IQC

Thank you

This project has received funding from the EU Copernicus Programme under Framework Agreement No. 2016/C3S_51_Lot3_BSC

nicolau.manubens@bsc.es