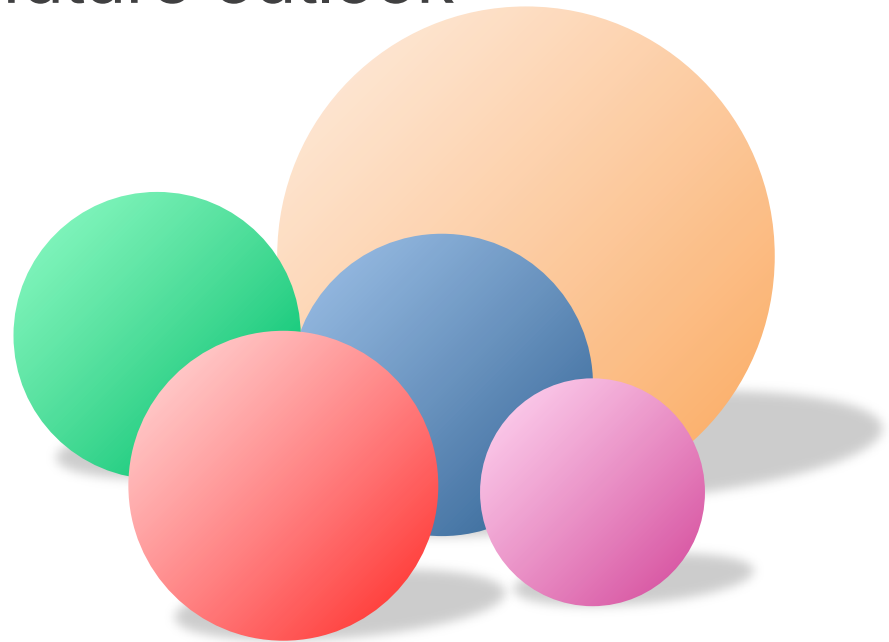# ICOS near real-time data processing

Current status and future outlook

Maggie Hellström

ICOS Carbon Portal
& Lund University

In collaboration with...

Jérôme Tarniewicz, Lynn Hazan and Leonard Rivier

ICOS Atmosphere Thematic Centre
Laboratoire des Sciences du Climat et de l'Environnement
Gif-sur-Yvette, France

Simone Sabbatini, Domenico Vitale and Dario Papale

ICOS Ecosystem Thematic Centre
University of Tuscia
Viterbo, Italy

Steve D. Jones, Benjamin Pfeil and Truls Johannessen

ICOS Oceanic Thematic Centre
University of Bergen
Bergen, Norway

Oleg Mirzov and Alex Vermeulen

ICOS Carbon Portal
Lund University
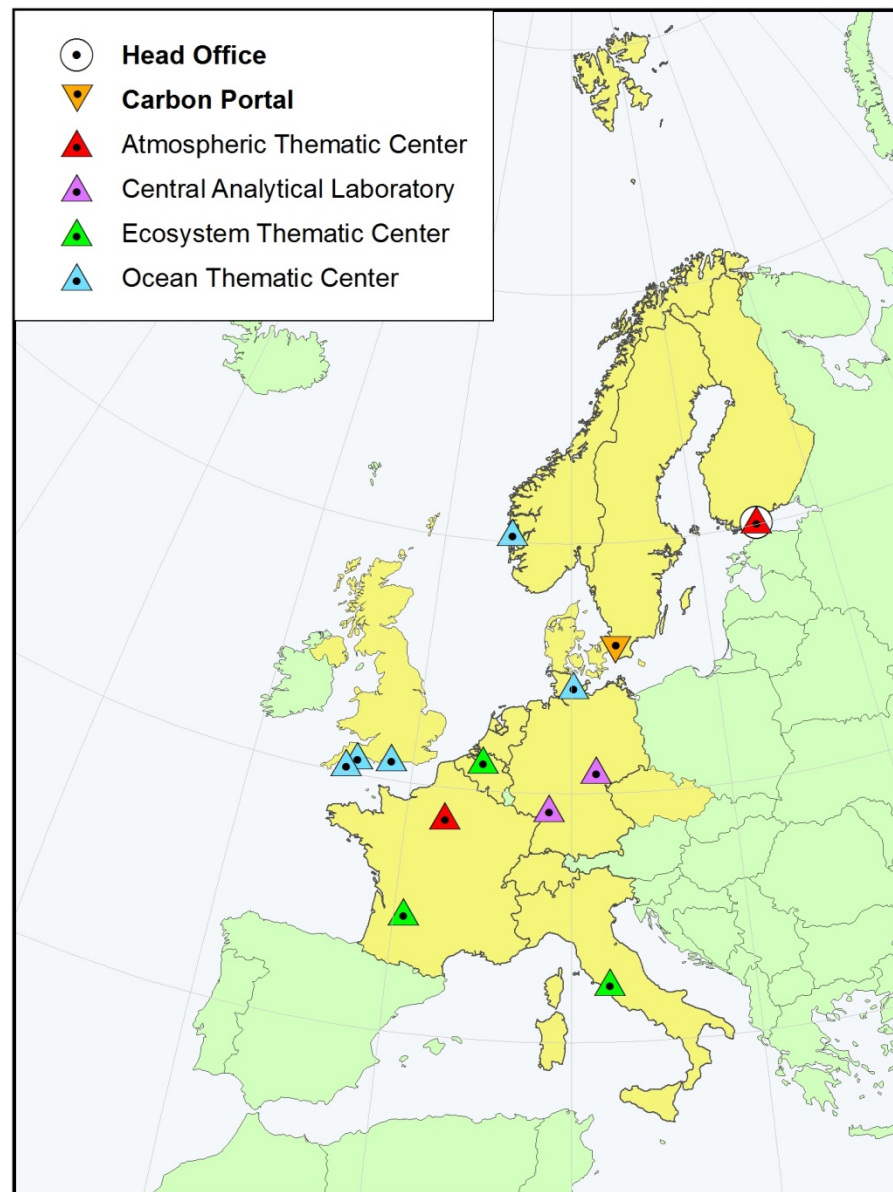Lund, Sweden

ICOS | Carbon Portal

What is ICOS?

- ICOS stands for Integrated Carbon Observation System

- It is a pan-European research infrastructure for observing and understanding the greenhouse gas (GHG) balance of Europe and its adjacent regions.

- ICOS brings together researchers from three domains – atmosphere, ecosystem and ocean.

- The major mission of ICOS is to collect and make available high-quality observational data from its state-of-the-art measurement stations operated with a long-term perspective.

- To achieve this, ICOS is developing a series of sustainable data management and computing services based on Open Source technology.

ICOS | Carbon Portal

ICOS organization

ERIC since November 2015

- Head Office in Finland

- Carbon Portal data center in Sweden

- Atmospheric Thematic Centre in France & Finland

- Ecosystem Thematic Centre in Italy, Belgium & France

- Ocean Thematic Centre in Norway, UK & Germany

- Central Laboratory in Germany

http://www.icos-ri.eu

- Head Office
- Carbon Portal
- Atmospheric Thematic Center
- Central Analytical Laboratory
- Ecosystem Thematic Center
- Ocean Thematic Center
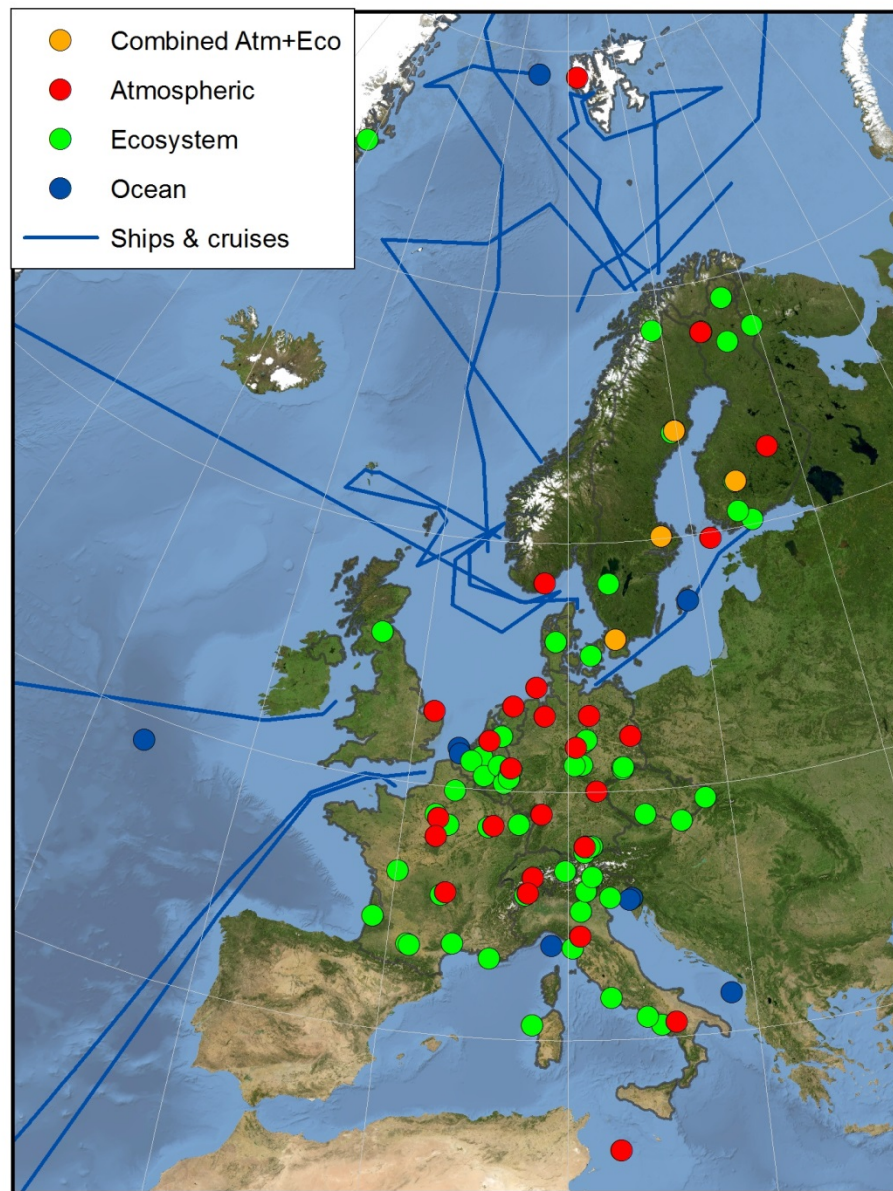
ICOS station networks

Stations

- 33 atmosphere
- 70 ecosystem
- 17 ocean

Member countries

- Belgium
- Czech republic
- Denmark
- Finland
- France
- Germany
- Italy
- Netherlands
- Norway
- Sweden
- Switzerland
- UK

Station networks are operated on a national level by each member country


ICOS | Carbon Portal

ICOS station networks (II)

ICOS observations

Continuous measurements
- Atm. concentrations of greenhouse gases: CO_2, CH_4, N_2O
- Exchange fluxes of CO_2 (+ CH_4 & N_2O) + latent & sensible heat
- Meteo: air temperature & pressure, precipitation, wind
- Soil & vegetation parameters (temperature, humidity, growth)
- Sea surface water parameters (p_{CO2}, T_{water}, pH)

Periodic sampling
- Rarer trace gases (CO, ...) & isotopes (^{14}C, radon)
- Biomass & vegetation samples
- Soil composition (organics/mineral, nutrients, ...)
- Water (oceans & land)

ICOS | Carbon Portal

ICOS data flow
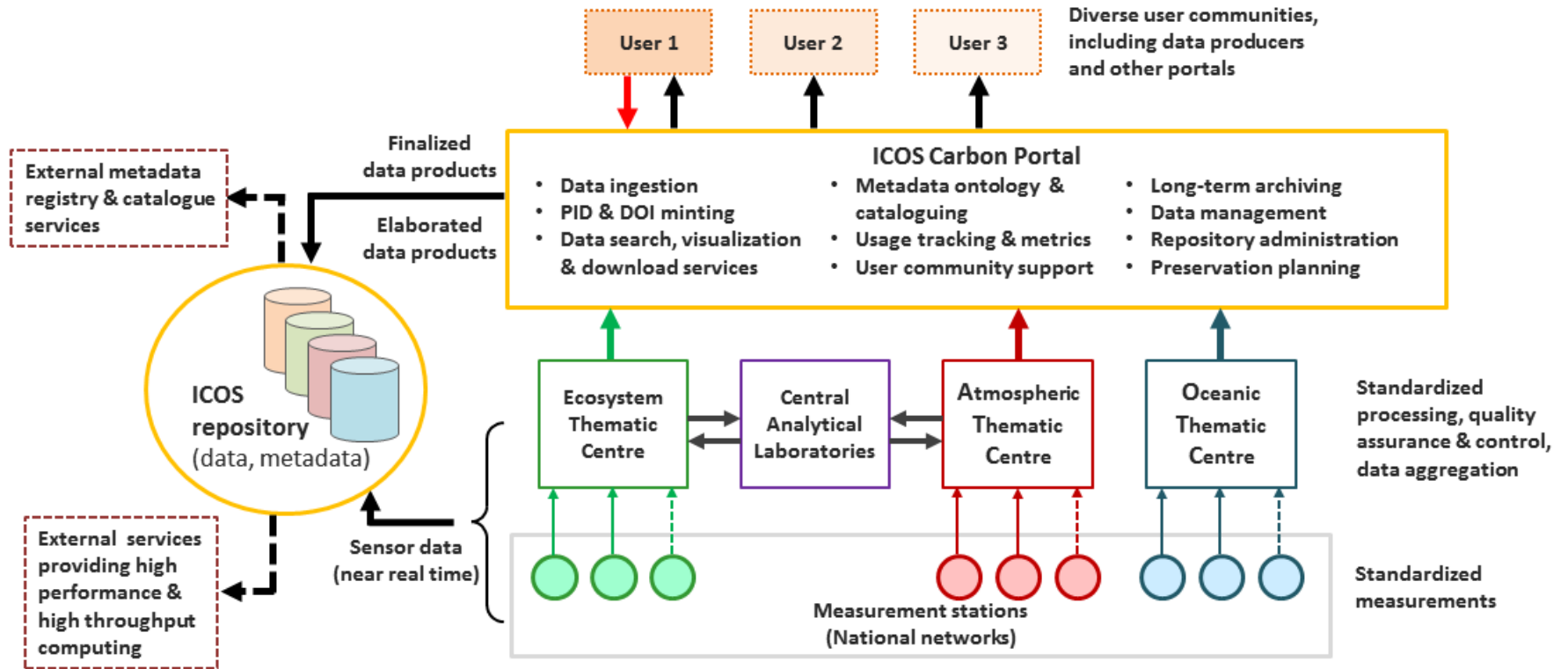
ICOS data types
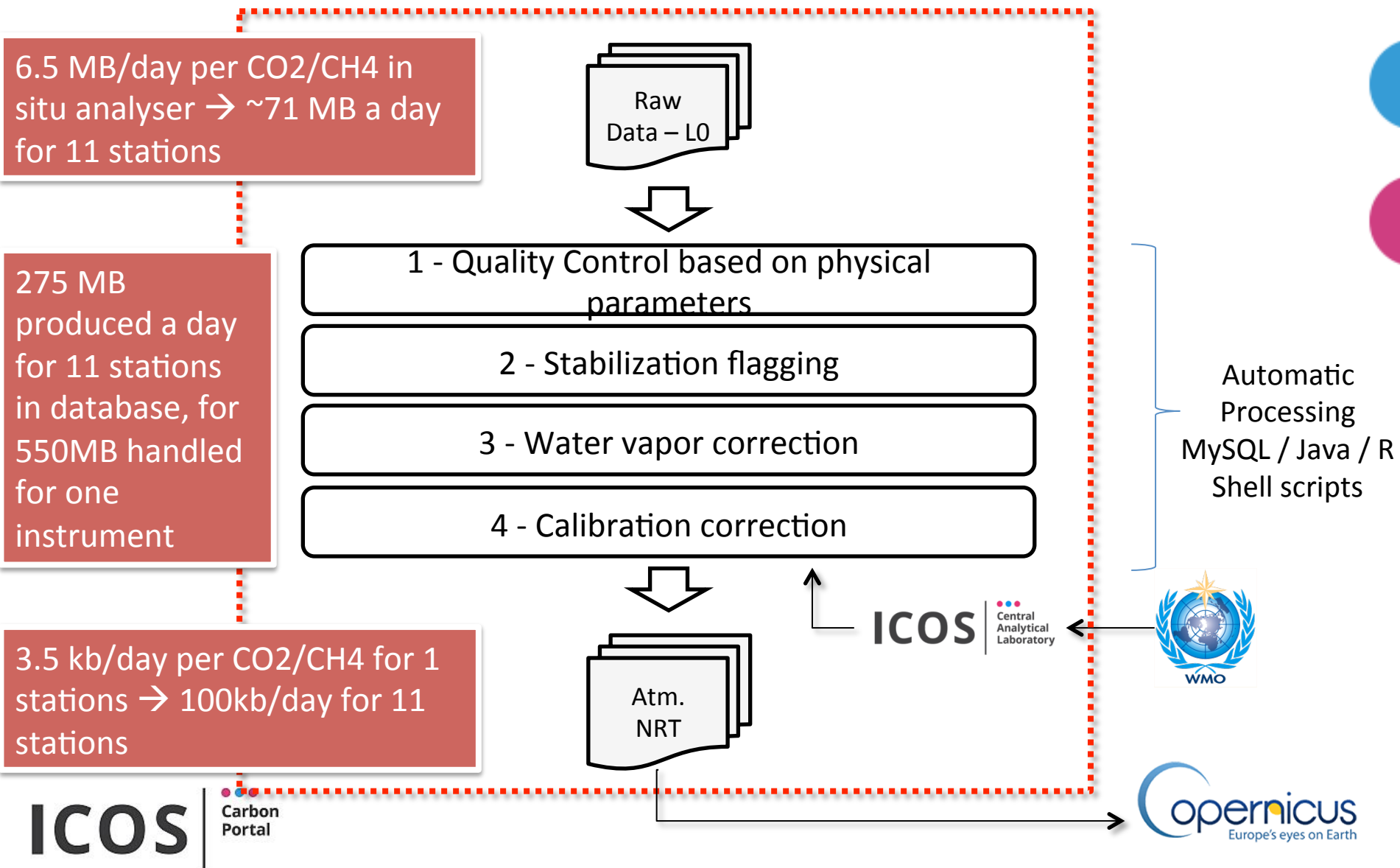
i) <u>raw sensor data</u> collected at the measurement stations associated with ICOS RI (known as Level 0 data);

ii) <u>near real-time data</u> sets (of e.g. greenhouse gas concentrations and/or fluxes) provided to users after applying basic processing & quality controls (NRT);

iii) <u>aggregated</u> & fully processed, quality-controlled <u>observational data</u> that are produced by ICOS expert centres based on the sensor data (Level 1 & 2);

iv) <u>"elaborated" products</u> produced by researchers external to ICOS, but based (in part) on ICOS observational data (Level 3). These are typically results from *model calculations*.

ICOS | Carbon Portal

ICOS NRT data processing: why?

Near real-time data processing occurs at several levels of ICOS
- *Station level:* quality checks & error detection
- *ICOS data repository:* sensor data should be stored as soon as possible (trust)
- *Thematic Centre level*: start data processing, including basic QA/QC, as soon as possible
- *Global monitoring system level:* atmosphere forecasting needs input data within ca 24 hours
- *Extreme event detection:* stakeholders & authorities need info within days or weeks
- *Outreach activities:* "live" data plots are interesting & useful

ICOS | Carbon Portal

ICOS atmosphere NRT processing

6.5 MB/day per CO2/CH4 in situ analyser → ~71 MB a day for 11 stations

275 MB produced a day for 11 stations in database, for 550MB handled for one instrument

Raw Data – L0

1 - Quality Control based on physical parameters

2 - Stabilization flagging

3 - Water vapor correction

4 - Calibration correction

Automatic Processing MySQL / Java / R Shell scripts

ICOS | Central Analytical Laboratory

WMO

Atm. NRT

3.5 kb/day per CO2/CH4 for 1 stations → 100kb/day for 11 stations

ICOS | Carbon Portal

Copernicus Europe's eyes on Earth

ICOS ecosystem NRT processing

Biomet Raw Data
sampled at 1s / 30min
collected in daily files
~few Kb

ICOS CP

EC Raw Data
sampled at 10 Hz
collected in
half-hourly files ~2
Mb

PROCESSING
(gCube VRE)

NRT product
Quality checked
half-hourly time series
For the Last 10 days
ASCII ~ few Mb

L2 product
Quality checked
half-hourly time series
For the whole year
ASCII ~ 30 Mb

Metadata

ICOS ETC

EC Setup
Processing Options

ICOS-CP
Release

ICOS | Carbon Portal

ICOS ocean NRT processing
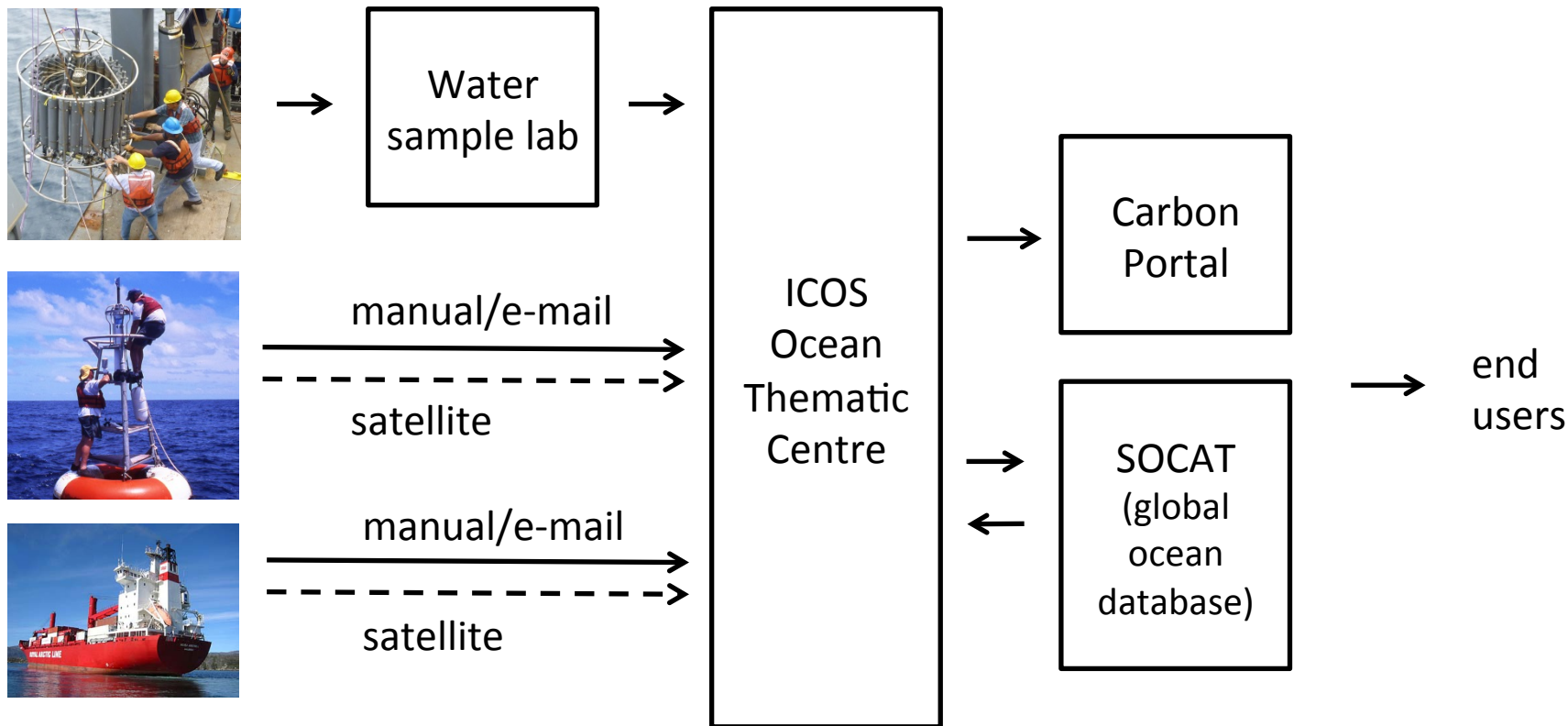
- Much can be automated, but still needs expert human analysis
- "Near"-real-time can mean weeks or months

ICOS | Carbon Portal

Challenges (I)

- Several parallel data streams per station; need to synchronize sensor clocks & collate data appropriately

- Sensors deliver mix of basic (volts, amperes) & physical (temperature, concentration) variables

- Physical sampling and "campaigns" complement continuous measurements

- Daily data volumes range from few MB to 1-2 GB per station

- Remote stations and platforms (especially for ocean domain) may have poor or non-existent internet connection

- Level of ICT competence varies much between stations, requiring common, easy to use solutions developed by Thematic Centres

ICOS | Carbon Portal

Challenges (II)

- ICOS must minimize the delay between data collection and storage at trusted repository (via Carbon Portal)

- Stored raw sensor data should be pushed/pulled of to Thematic Centres for processing; this may require staging operations (for large ecosystem files) to HTC resources

- Calibrations (coefficients, timing, standards used) must be documented & uploaded to Thematic Centres in near real-time

- Overall, efficient metadata collection, handling and curation is crucial

- Identification, Authentication & Authorization are very important at all stages of the data handling, but is also complicated

ICOS | Carbon Portal

Challenges (III)

- Quality assessment is complex, and domain-specific
- Automated QA often based on thresholds and outlier detection, but these can miss "subtle" problems
- Input from experts and station personnel required before finalizing data products (not always possible for NRT data)
- GUI-based tools must be developed to support station personnel in their data assessment
- "Flagging" schemes used for QA can be complicated and difficult to work with
- Data identified by QA/QC as bad or questionable may be "gap-filled" – e.g. via lookup tables or neural networks based on knowledge of biogeochemical processes

ICOS | Carbon Portal

Summary & outlook

- ICOS delivers NRT data products – but the time delay, the variable range and end users vary across domains

- NRT data users include global and European projects for atmospheric quality & greenhouse gas monitoring

- Technical challenges include

    - improving data transfer rates
    - automating QA/QC procedures
    - optimizing workflows involving HTC and VRE e-Service providers

- Applying persistent identifiers and provisioning relevant metadata supports machine-based processing of ICOS data – both in near real time and at longer timescales
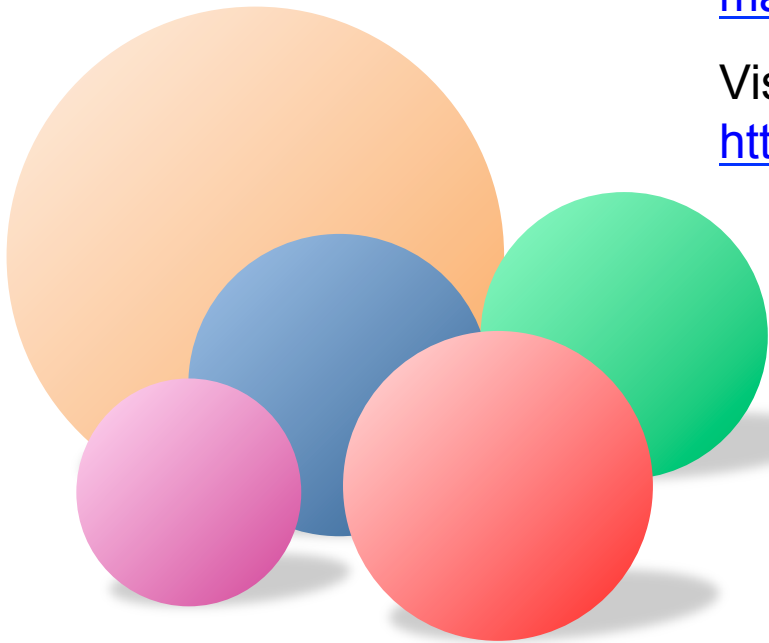
ICOS | Carbon Portal

Thank you!

Questions or comments?

Contact Maggie at
margareta.hellstrom@nateko.lu.se

Visit the ICOS Carbon Portal at
https://www.icos-cp.eu/