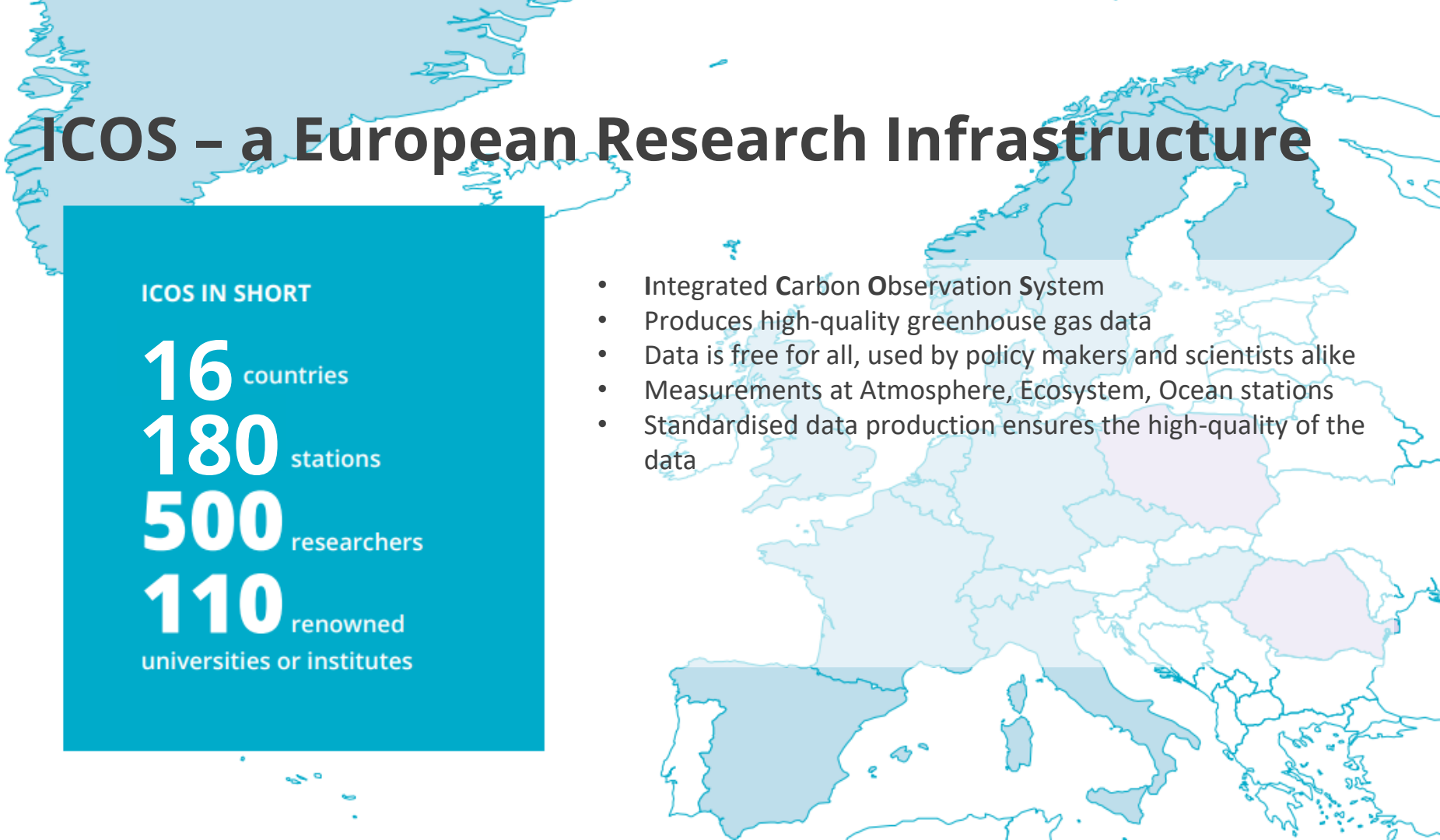
Challenges and opportunities from an in-house cross collaboration between three research infrastructure data repositories

Claudio D'Onofrio, EGU2024, ESSI 3.5

Acknowledgment for the ICOS Carbon Portal

Alex Vermeulen
Harry Lankreijer
Anders Dahlner
Maggie Hellström
Wouter Peters
Ute Karstens
Oleg Mirzov
André Bjärby
Jonathan Thiry
Angeliki Adamaki
Klara Broman
Ida Storm
Zhendong Wu
Remco de Kok
Jonathan Schenk
Zois Zogopoulos

ICOS – a European Research Infrastructure



ICOS IN SHORT

16 countries

180 stations

500 researchers

110 renowned universities or institutes

- **I**ntegrated **C**arbon **O**bservation **S**ystem
- Produces high-quality greenhouse gas data
- Data is free for all, used by policy makers and scientists alike
- Measurements at Atmosphere, Ecosystem, Ocean stations
- Standardised data production ensures the high-quality of the data

What is this all about?

The ICOS Carbon Portal is hosting three data repositories:

Integrated Carbon Observation System
ICOS RI https://data.icos-cp.eu/portal/

The Swedish Infrastructure for Ecosystem Science
SITES https://data.fieldsites.se/portal/

Pilot Applications in Urban Landscapes – Towards Integrated City Observatories for Greenhouse Gases
ICOS CITIES https://citydata.icos-cp.eu/portal/

What is this all about?

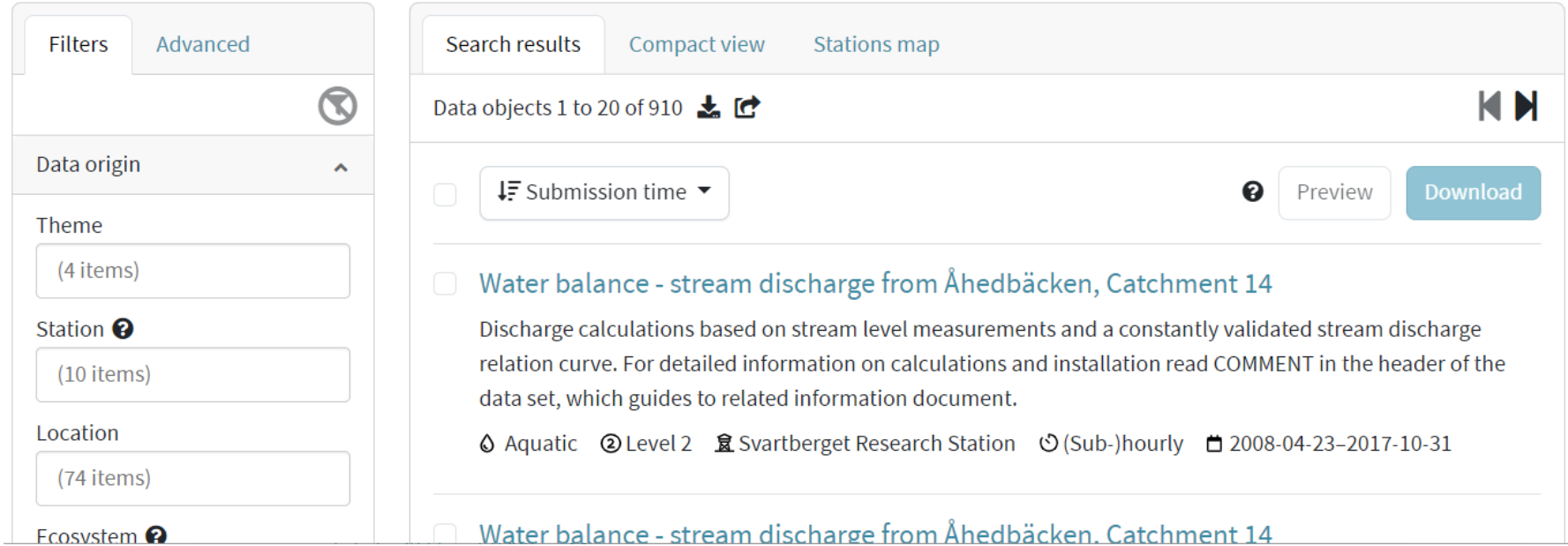
Each of the data portals has an instance based on the same software stack
(backend, frontend, triple store, programmatical access, authentication, …)

Hence, we have tried to answer the following questions:

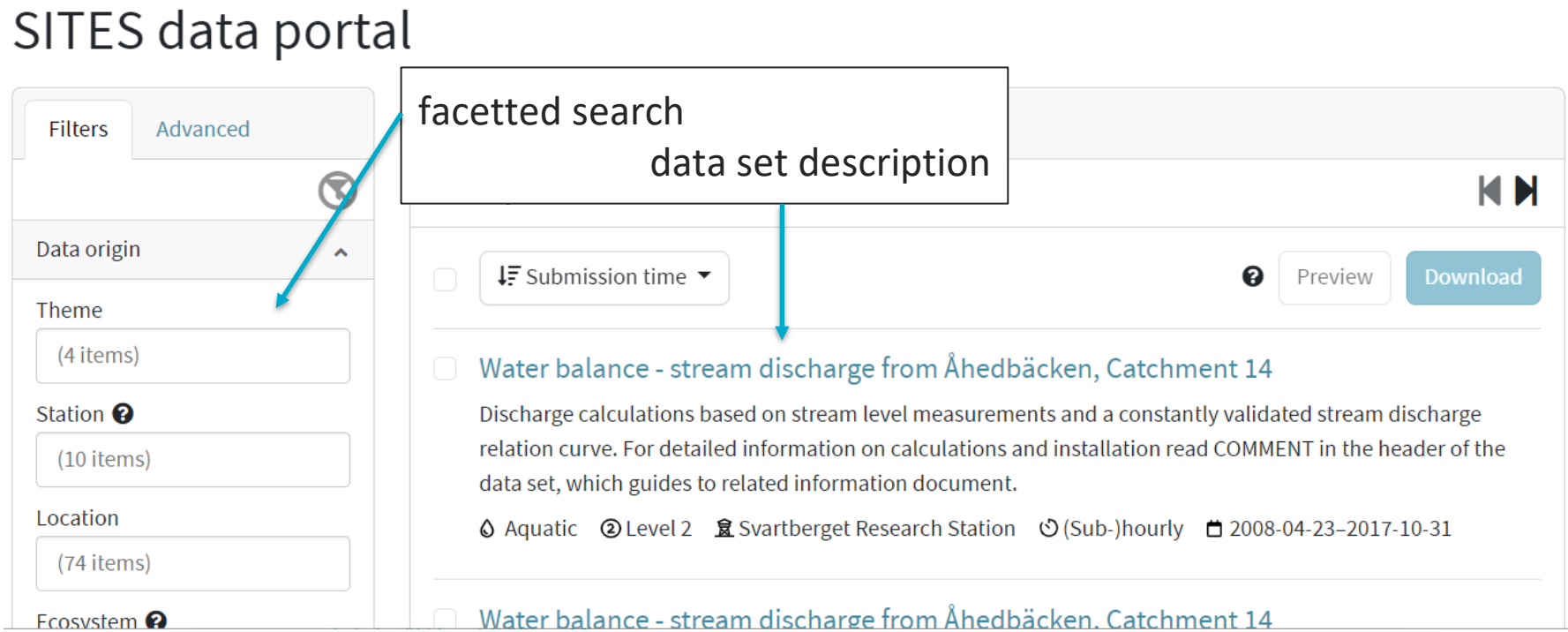
- FAIRness of data portal?
- Is it easy to load/access data from the Research Infrastructure?
- Can we load/analyze/visualize data ACROSS Research Infrastructure?
- What are the challenges?
- What are the opportunities?

Data Portal and the very basics

SITES data portal



Data Portal and the very basics

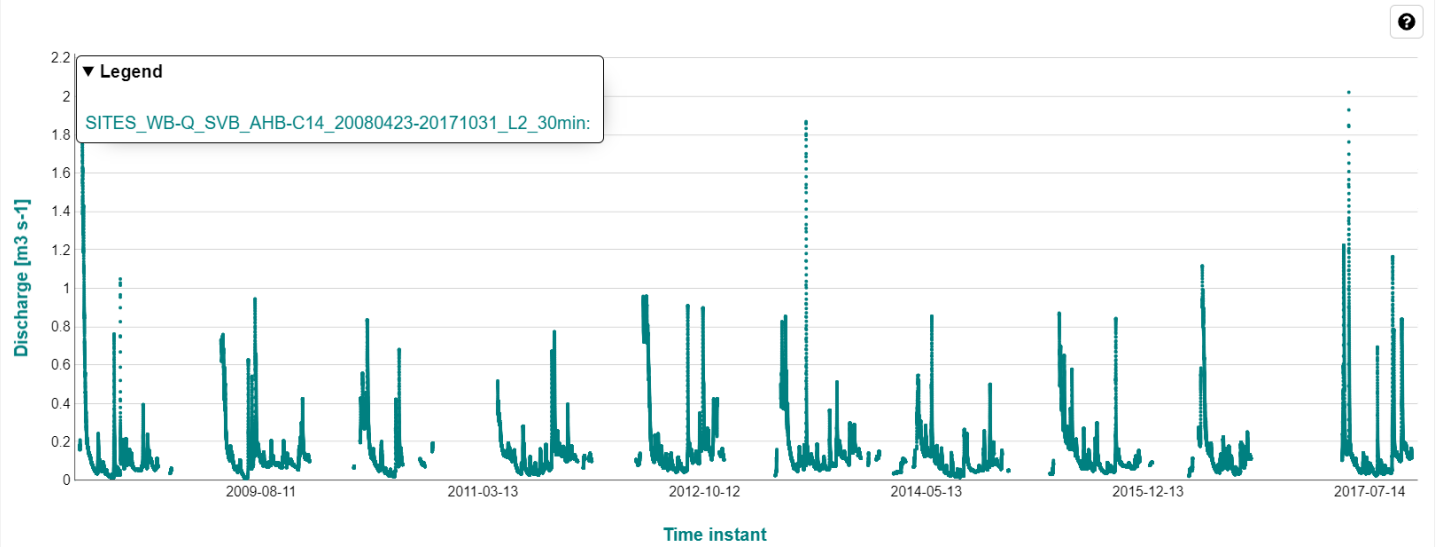


Data Portal and the very basics

2008-04-23–2017-10-31

Metadata | Preview

Q—Discharge | Select a second parameter | Scatter | Line | Use preview data



Data Portal and the very basics

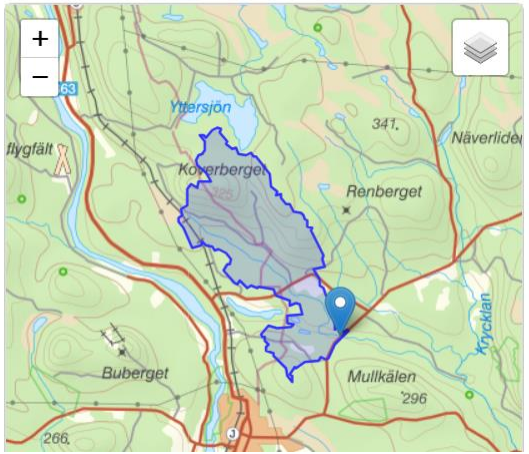
Water balance - stream discharge from Åhedbäcken, Catchment 14

Add to cart | Download

2008-04-23–2017-10-31

Metadata | Preview

PID	11676.1/kyCDtIXhWr-xJcyah0_hOyBt (link)
Description	Discharge calculations based on stream level measurements and a constantly validated stream discharge relation curve. For detailed information on calculations and installation read COMMENT in the header of the data set, which guides to related information document.
Data affiliation	SITES Water
Citation	Svartberget Research Station (2024). Water balance - stream discharge from Åhedbäcken, Catchment 14, 2008-04-23–2017-10-31 [Data set]. Swedish Infrastructure for Ecosystem Science (SITES). https://hdl.handle.net/11676.1/kyCDtIXhWr-xJcyah0_hOyBt BibTex RIS
File name	SITES_WB-Q_SVB_AHB-C14_20080423-20171031_L2_30min.csv
File size	4 MB (4429994 bytes)



Are we FAIR?

Assessment by F-UJI https://www.f-uji.net/index.php?action=test

F-UJI is a web service to programatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics developed by Anusuriya Devaraju & Robert Huber (PANGAEA) under the umbrella of the FAIRsFAIR project.

F-UJI is a result of the FAIRsFAIR “Fostering FAIR Data Practices In Europe” project which received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 (grant agreement 831558).

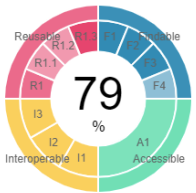
Are we FAIR?

ETC L2 ARCHIVE, Tharandt, 2019-12-31–2023-10-31

Save | {JSON} | New

FAIR level:	advanced
Resource PID/URL:	https://meta.icos-cp.eu/objects/e-1t32ycaNdDBiWU-6jWiW3i
DataCite support:	enabled
Metric Version:	metrics_v0.5
Metric Specification:	https://doi.org/10.5281/zenodo.6461229
Software version:	3.1.0
Download assessment results:	{JSON}
Save and share assessment results:	

Summary:



	Score earned:	Fair level:
Findable:	6 of 7	advanced
Accessible:	2 of 3	moderate
Interoperable:	4 of 4	advanced
Reusable:	7 of 10	moderate

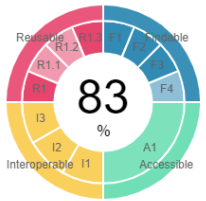
Are we FAIR?

Chemical variables - stream from Röbäcksdalen Catchment, Sampling point 5, 2016-11-17–2023-11-15

Save | {JSON} | New

FAIR level:	advanced
Resource PID/URL:	https://meta.fieldsites.se/objects/NbNbJ33f_K47Ndmcc_0Lw9wp
DataCite support:	enabled
Metric Version:	metrics_v0.5
Metric Specification:	https://doi.org/10.5281/zenodo.6461229
Software version:	3.1.0
Download assessment results:	{JSON}
Save and share assessment results:	

Summary:



	Score earned:	Fair level:
Findable:	6 of 7	advanced
Accessible:	2 of 3	moderate
Interoperable:	4 of 4	advanced
Reusable:	8 of 10	moderate

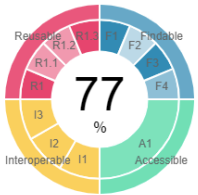
Are we FAIR?

Save | {JSON} | New

FAIR level:	advanced
Resource PID/URL:	https://citymeta.icos-cp.eu/objects/bO_M0cDx8SeVfxzGo1CbxDWZ
DataCite support:	enabled
Metric Version:	metrics_v0.5
Metric Specification:	https://doi.org/10.5281/zenodo.6461229
Software version:	3.1.0
Download assessment results:	{JSON}
Save and share assessment results:	

Title missing

Summary:



	Score earned:	Fair level:
Findable:	4.5 of 7	advanced
Accessible:	2 of 3	moderate
Interoperable:	4 of 4	advanced
Reusable:	8 of 10	moderate

Put it to the test

Two use cases:

- Load data from Svartberget (co-located) research station in Sweden from ICOS and SITES and compare air temperature

- Compare CO_2 concentrations from ICOS Cities, Pilot City Zurich, and ICOS Atmospheric Station of Beromunster
  - Load a single dataset from a low-cost sensor in Zurich and access the concentration measurements and create a fingerprint plot, a day profile
  - Load more data from ICOS Switzerland and compare

Jupyter Service @ ICOS

https://www.icos-cp.eu/data-services/tools/jupyter-notebook

Load Library

[1]:
```
from icoscp_core.sites import meta as sitesmeta, data as sitesdata
```

Access the metadata

[2]:
```
metaSITES = sitesmeta.get_dobj_meta('https://meta.fieldsites.se/objects/v0bn_ufBJ4vgq8Nen9d-Vqe5')
```

[ ]:
```

```

Svartberget FIELDSITES & ICOS

Top functions to access meta data

'accessUrl’
'coverageGeo’
'doi’
'fileName’
'hash’
'latestVersion’
'nextVersion’
'parentCollections’
'pid’
'previousVersion’
'references’
'size’
'specificInfo’
'specification’
'submission'

Svartberget FIELDSITES & ICOS

Use the locations provided to calculate distance in KM, where the measurments are coming from

```
coordSITES = (metaSITES.specificInfo.acquisition.station.location.lat, metaSITES.specificInfo.acquisition.station.location.lon)
```

```
coordICOS = (metaICOS.specificInfo.acquisition.station.location.lat, metaICOS.specificInfo.acquisition.station.location.lon)
```

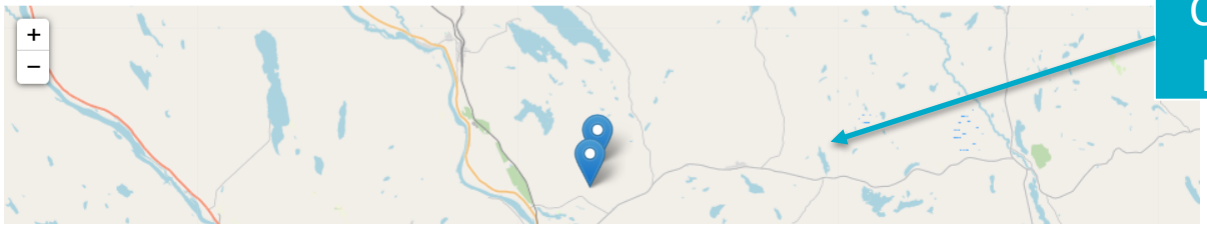
```
distance = geodistance.distance(coordSITES, coordICOS).km
print(f'Distance in KM between the two research stations: {distance:.2f}')
```

```
Distance in KM between the two research stations: 1.37
```

Distance between research stations

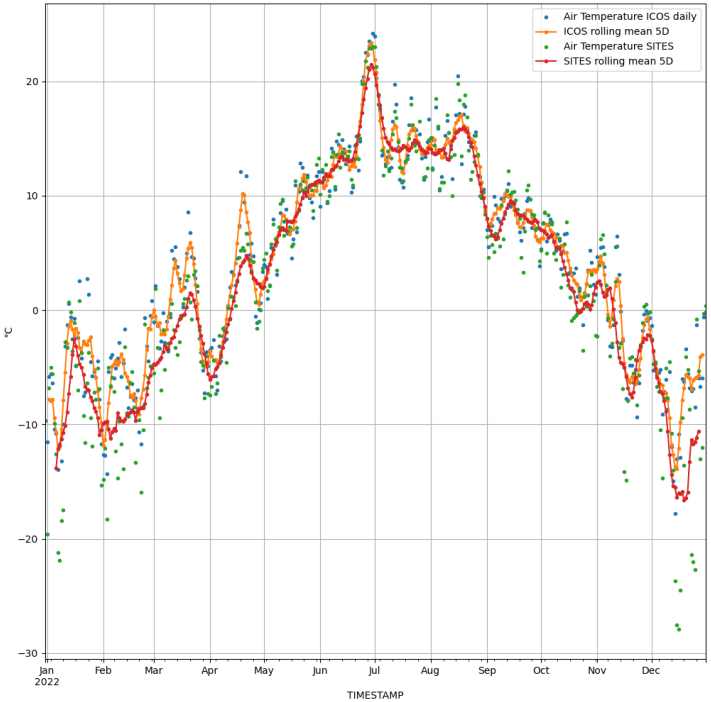
```
# create a map with all stations
# define the initinal paramters of centre point and zoom
Center = [64, 19]
Zoom = 8
#create the map
myMap = folium.Map(location=Center,zoom_start=Zoom,no_wrap=True, preferCanvas=True)
```

```
#create a marker to display the station on the map
myMap.add_child(folium.Marker(location=[coordSITES[0], coordSITES[1]],popup='SITES'))
myMap.add_child(folium.Marker(location=[coordICOS[0], coordICOS[1]], popup='ICOS'))
```

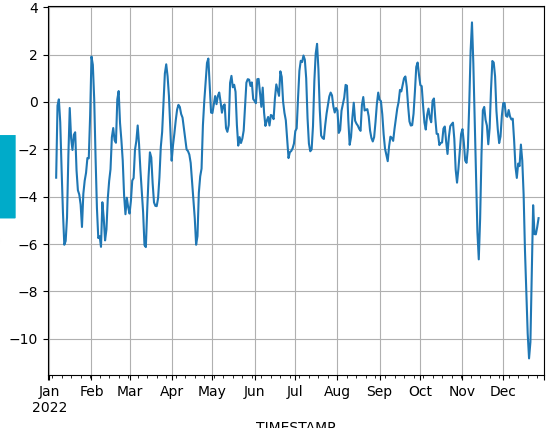


Create a map and plot the location

Svartberget FIELDSITES & ICOS

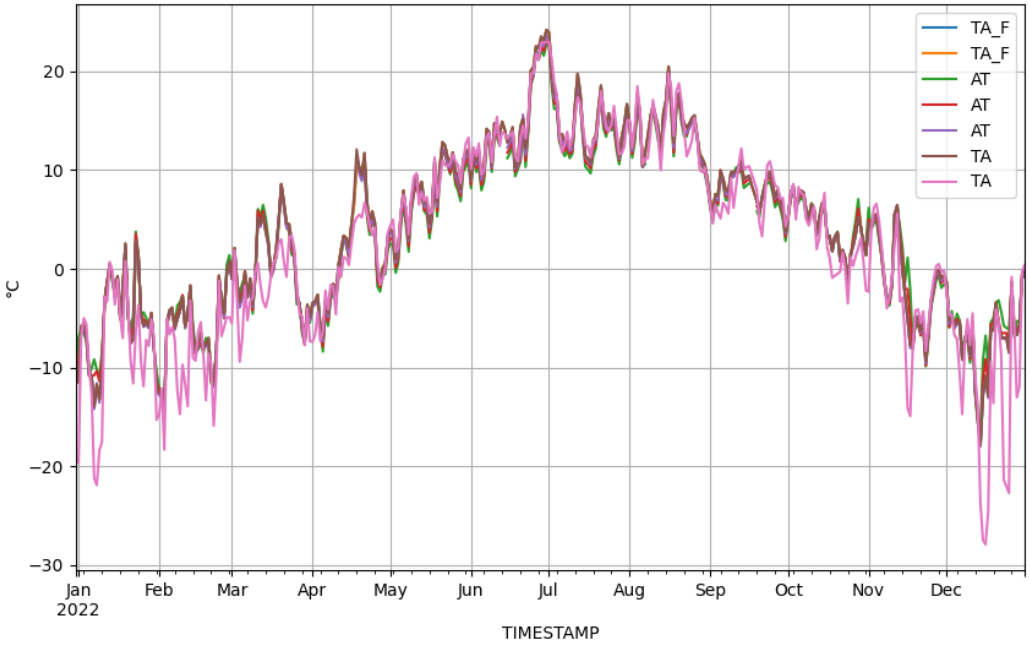


difference



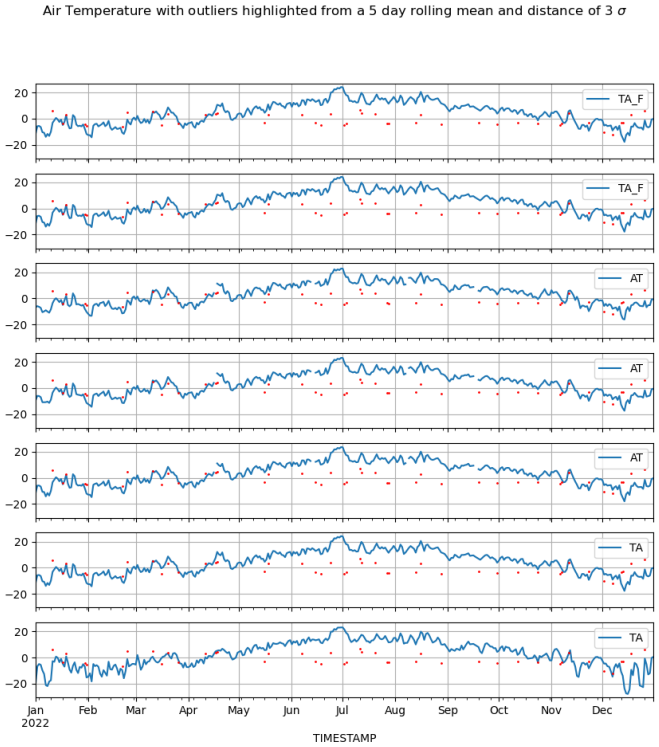
Air Temperature from Svartberget, ICOS and FIELDSITES plus rolling mean and the difference.

Svartberget FIELDSITES & ICOS



Compare more datasets.
The feature of ‘Air Temperature’ helped to find more datasets, although the variables have different naming schemes.

Svartberget FIELDSITES & ICOS



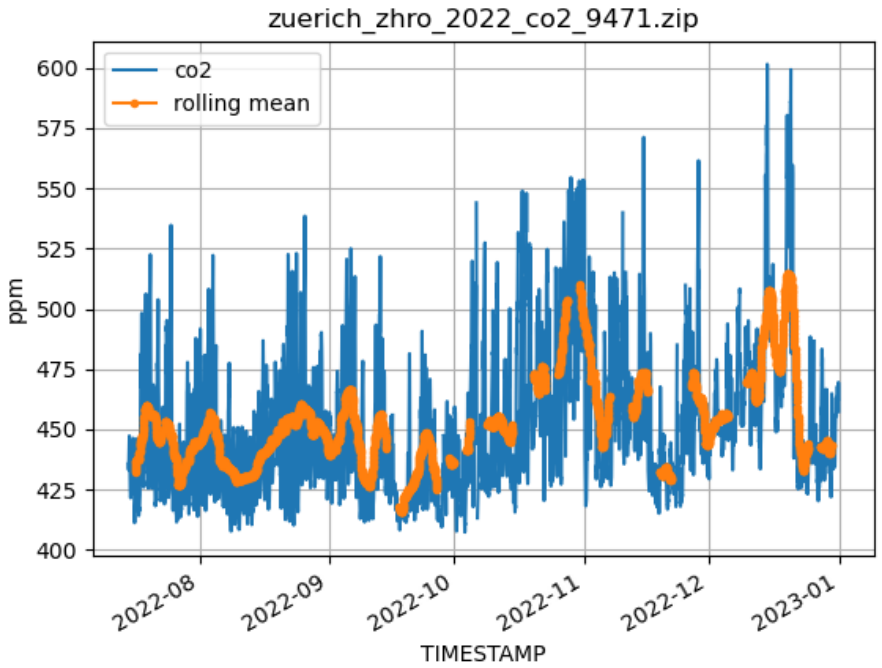
Easy to start analyzing and comparing.

Is it useful, do you gain insight?
Just because you can, not always helpful to make pretty pictures.

Svartberget FIELDSITES & ICOS

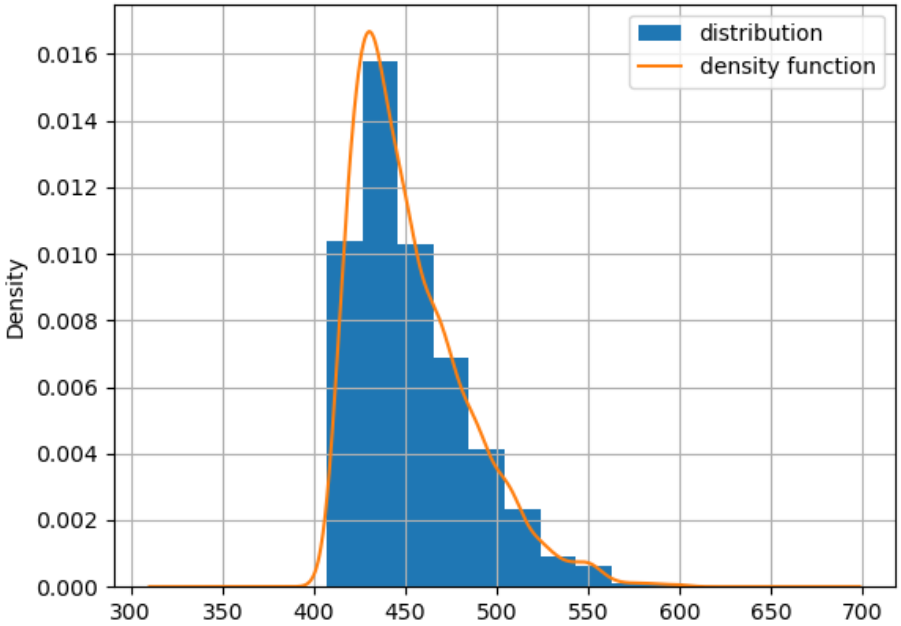
- Peichl, M., Nilsson, M., Smith, P., Marklund, P., De Simon, G., Löfvenius, P., Dignam, R., Holst, J., Mölder, M., Andersson, T., Jensen, R., Kozii, N., Larmanou, E., Linderson, M., Ottosson-Löfvenius, M., Tülp, H., Öquist, M. (2024). ETC L2 Fluxnet (half-hourly), Svartberget, 2018-12-31–2023-12-31, ICOS RI, https://hdl.handle.net/11676/xgTFH-wtBGPCJJcQwzHguokP
- Peichl, M., Nilsson, M., Smith, P., Marklund, P., De Simon, G., Löfvenius, P., Dignam, R., Holst, J., Mölder, M., Andersson, T., Jensen, R., Kozii, N., Larmanou, E., Linderson, M., Ottosson-Löfvenius, M., Tülp, H., Öquist, M. (2024). ETC L2 Meteosens, Svartberget, 2018-12-31–2023-12-31, ICOS RI, https://hdl.handle.net/11676/Owvfo6T6v9ye_6NPU5TYwxSK
- Marklund, P., Ottosson-Löfvenius, M., Smith, P. (2023). ICOS ATC Meteo Release, Svartberget (150.0 m), 2017-10-27–2023-03-31, ICOS RI, https://hdl.handle.net/11676/M838mryykcBU52fCVmromJMQ
- Marklund, P., Ottosson-Löfvenius, M., Smith, P. (2023). ICOS ATC Meteo Release, Svartberget (85.0 m), 2017-10-26–2023-03-31, ICOS RI, https://hdl.handle.net/11676/ZBSIExmUTJHsTeDxfQTT_fH5
- Marklund, P., Ottosson-Löfvenius, M., Smith, P. (2023). ICOS ATC Meteo Release, Svartberget (35.0 m), 2017-10-26–2023-03-31, ICOS RI, https://hdl.handle.net/11676/ffhIeMzOaWmVMn6jNPSMdMz1
- Svartberget Research Station (2023). Meteorological data from Svartberget, Åheden AWS, 2022 [Data set]. Swedish Infrastructure for Ecosystem Science (SITES). https://hdl.handle.net/11676.1/v0bn_ufBJ4vgq8Nen9d-Vqe5

ICOS Cities & ICOS



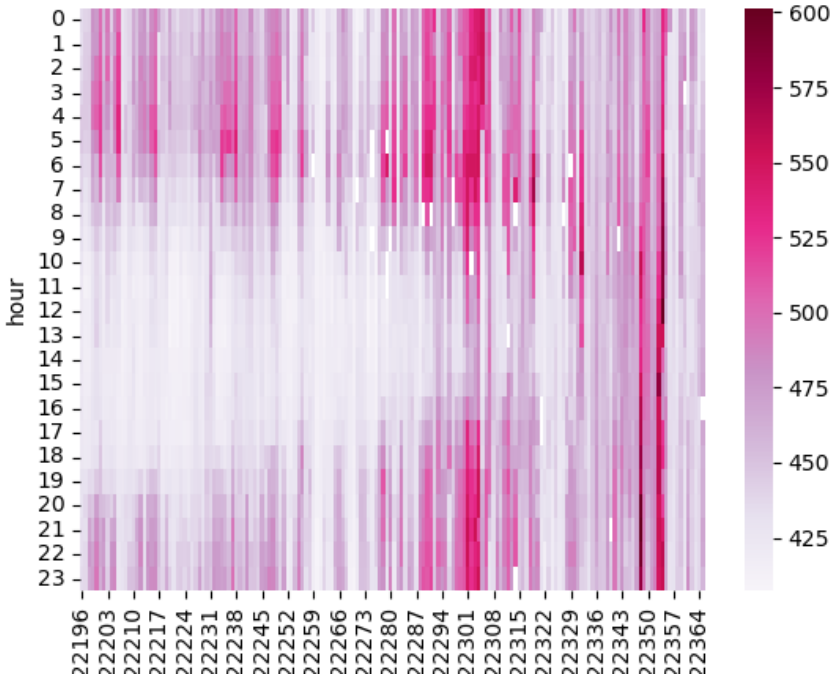
Timeseries of CO_2
from a low-cost sensor
in the city of Zurich including
a rolling mean (72h)

ICOS Cities & ICOS



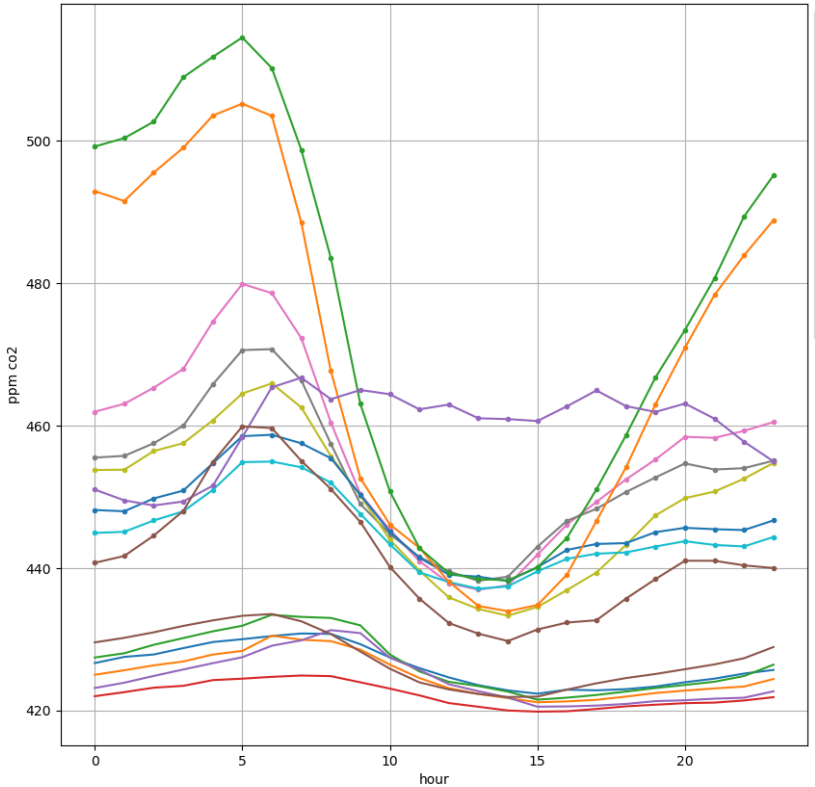
Distribution and density function of the CO_2 signal.

ICOS Cities & ICOS



Fingerprint plot:
means for all data points grouped by hours of day

ICOS Cities & ICOS



Compare many different data sets from ICOS and ICOS Cities (lines including the ‘dots’)

- ICOS_ATC_OBSPACK-Europe-L2-2023_LHW_32.0_CTS_CO2.zip
- ICOS_ATC_OBSPACK-Europe-L2-2023_BRM_72.0_CTS_CO2.zip
- ICOS_ATC_OBSPACK-Europe-L2-2023_BRM_44.6_CTS_CO2.zip
- ICOS_ATC_OBSPACK-Europe-L2-2023_BRM_212.0_CTS_CO2.zip
- ICOS_ATC_OBSPACK-Europe-L2-2023_BRM_132.0_CTS_CO2.zip
- ICOS_ATC_OBSPACK-Europe-L2-2023_BRM_12.5_CTS_CO2.zip
- zuerich_zbad_2022_co2_9442.zip
- zuerich_zbad_2022_co2_9441.zip
- zuerich_zhsf_2022_co2_5658.zip
- zuerich_sma1_2022_co2_9417.zip
- zuerich_sma1_2022_co2_9416.zip
- zuerich_leub_2022_co2_9331.zip
- zuerich_leub_2022_co2_9330.zip
- zuerich_belv_2022_co2_9299.zip
- zuerich_belv_2022_co2_9298.zip
- zuerich_belv_2022_co2_9270.zip

ICOS Cities & ICOS

- Brunner, D., Henne, S., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Laegern-Hochwacht (32.0 m), 2012-08-01–2023-01-01, European ObsPack, https://hdl.handle.net/11676/4nOWXeqNFgv5_hSm9RVdaWIT
- Leuenberger, M., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Beromunster (72.0 m), 2012-11-01–2023-03-31, European ObsPack, https://hdl.handle.net/11676/tXvPEQxxkBwIIYfX8tJqpch-
- Leuenberger, M., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Beromunster (44.6 m), 2012-11-01–2023-03-31, European ObsPack, https://hdl.handle.net/11676/ppeXm71svssuwA4NKlrlD7GE
- Leuenberger, M., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Beromunster (212.0 m), 2012-11-01–2023-03-31, European ObsPack, https://hdl.handle.net/11676/OQ2aMDJ4nbOM6wX7Xok8UPDA
- Leuenberger, M., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Beromunster (132.0 m), 2012-11-01–2023-03-31, European ObsPack, https://hdl.handle.net/11676/P7SAYE683P4v0_L-YTO_15iG
- Leuenberger, M., Atmosphere Thematic Centre, ICOS-CAL-FCL (2023). Atmospheric CO2 product, Beromunster (12.5 m), 2012-11-01–2023-03-31, European ObsPack, https://hdl.handle.net/11676/RhQHWf0KayXJ5nk55UL1j2JK
- Grange, S., Rubli, P., Emmenegger, L., 2024. ICOS Cities release of low and mid-cost CO2 data from the Zürich network for the period 1 July 2022 until 1 December 2023.
- https://doi.org/10.18160/W0AJ-4ZGE

Discussion of content

As shown in the previous slides, we found many different measurements, and probably all are valid, but more context is needed for science

- We could now talk about the scientific content and discussion would include
  - SamplingHeight
  - Horizontal Profiles
  - Sensor drift and bias
  - Instruments and calibration,
  - Sampling rate
  - etc.

- The provided sampling rate is different (half hourly, hourly, day) and needs to be harmonized / resampled to compare

- We found different names like TA and AT (the unit is the same), the overarching meta data descriptions is the same (Air Temperature)

Discussion of FAIR‘ness

- Overall, a very rich set of metadata.
- Information is structured in a complex way, and possibly hard to find.
- Easy access for humans and machines.
- Timestamp for these three repositories is harmonized to UTC. But if you look at Ecosystem measurements, it is important to take sunrise and sunset into account, hence, if you compare two sites, you need to think about time zones and photosynthesis.

Challenges

- Expert knowledge is still required
- The trap of producing pretty pictures
- Easy to compare non-comparable data
- Timestamps, all harmonized to UTC, take care if you cross timezones
- Separation of metadata and data

Opportunities

- Crossing RI's and domains made easy
- Combine data sets to find new dependencies or correlations
- Potential for machine learning by adding new features to observations
- Validate observations across domains
- Simplify workflows

Thank you!

ICOS | INTEGRATED CARBON OBSERVATION SYSTEM