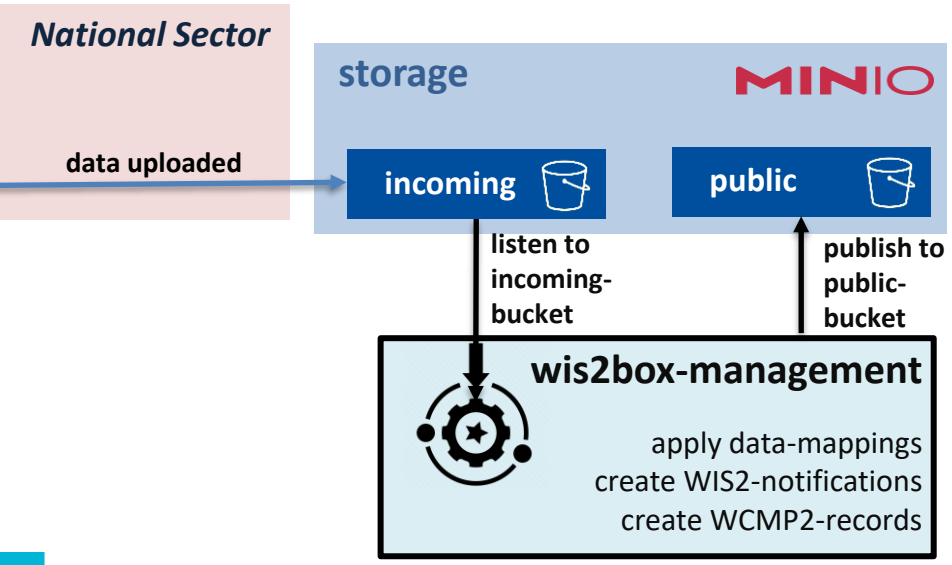
Ingesting and publishing data using wis2box

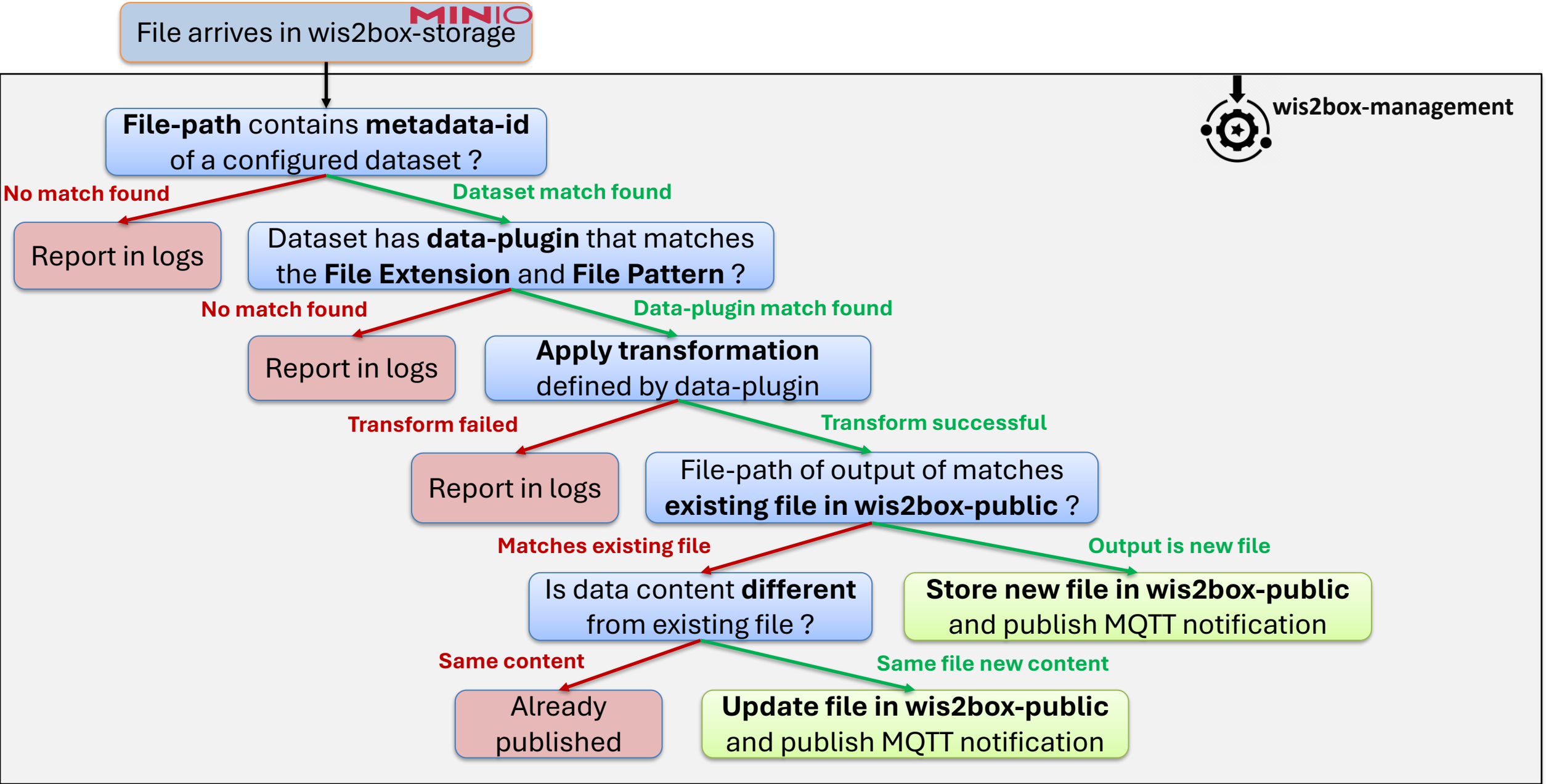
wis2box data ingest workflow



Dataset driven workflow triggered when wis2box-management receives a notification from the storage service

- wis2box-management **matches incoming data** with datasets based on **file-path**
- **data plugins** are applied to **transform & publish the data**

wis2box data mapping workflow



How to upload data to wis2box

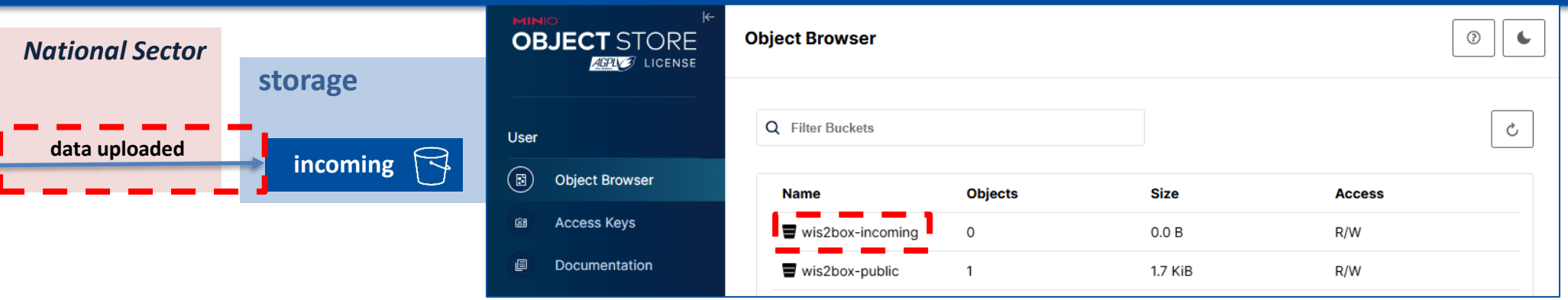
Data can be uploaded to wis2box-incoming in different ways

Optimal solution will depend on systems used in national sector

Options for ingesting data in wis2box:

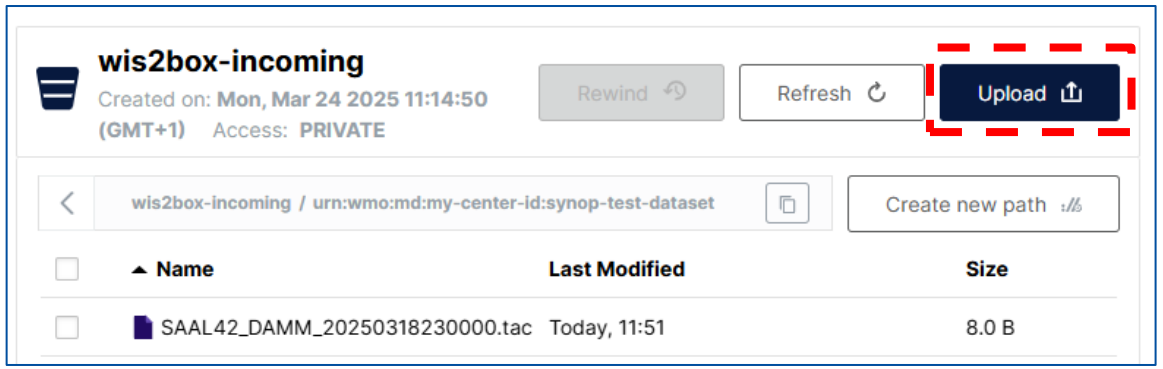
- **Manual upload using MinIO UI:** manually create path and upload data, useful for testing data-workflow
- **Interfaces in wis2box-webapp:** manual submission using FM-12 form or manual file upload interface
- **Scripts:** write code to upload data using MinIO client software
- **Using SFTP protocol:** forward data to MinIO SFTP-endpoint

MinIO User Interface



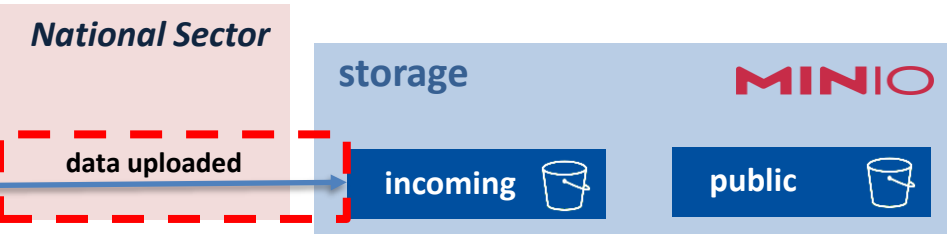
Data can be uploaded manually using the **MinIO web interface at http://<your-host>:9001**

- **Useful for testing** your data mappings before setting up an automated workflow
- Login with WIS2BOX_STORAGE_USERNAME / WIS2BOX_STORAGE_PASSWORD defined in wis2box.env
- Create new path in wis2box-incoming bucket matching metadata-id and upload your file



For security do not enable external access on port 9001 nor share wis2box-storage credentials

Data ingestion script



Data can be uploaded using a minio client library

- Useful to setup custom data workflow:
  - Cronjob running Python script uploading files from local server
  - Query local database to extract CSV and upload result
- https://min.io/docs/minio/linux/developers/minio-drivers.html

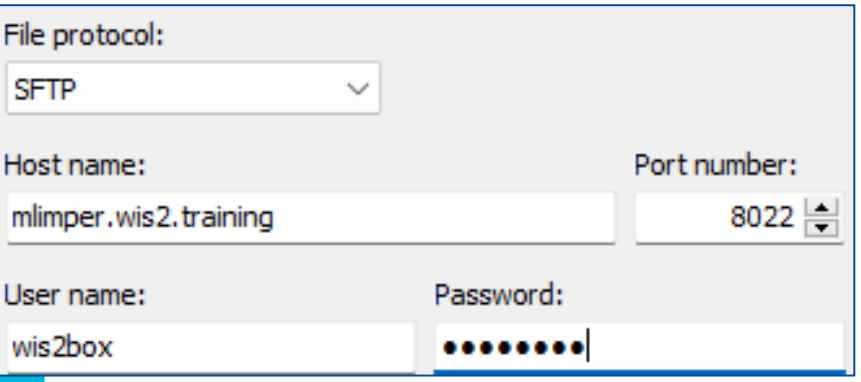
Python example to upload a file using the MinIO client module:

```
import glob
import sys
from minio import Minio

filepath = '/home/wis2box-user/local-data/mydata.bin'

endpoint = 'http://wis2box-host:9000'
WIS2BOX_STORAGE_USERNAME = 'wis2box'
WIS2BOX_STORAGE_PASSWORD = 'XXXXXXXXX'

client = Minio(
    endpoint=endpoint,
    access_key=WIS2BOX_STORAGE_USERNAME,
    secret_key=WIS2BOX_STORAGE_PASSWORD,
    secure=is_secure=False)

filename = filepath.split('/')[-1]
# path should match the metadata-id for corresponding dataset
minio_path = 'urn:wmo:md:my-center-id:synop-test-dataset'
client.fput_object('wis2box-incoming', minio_path+filename, filepath)
```

SFTP

Data can be uploaded using SFTP

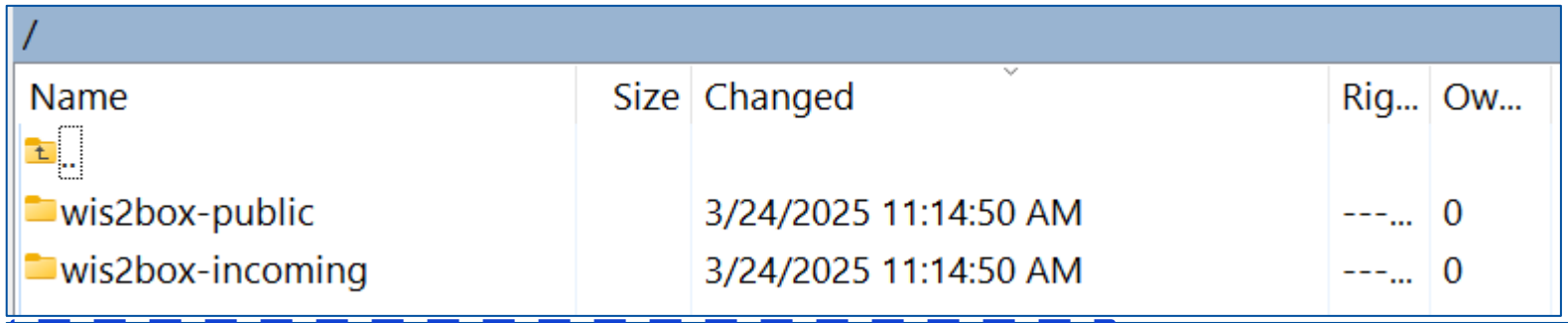
- Useful for automating data-flow from vendors supporting SFTP

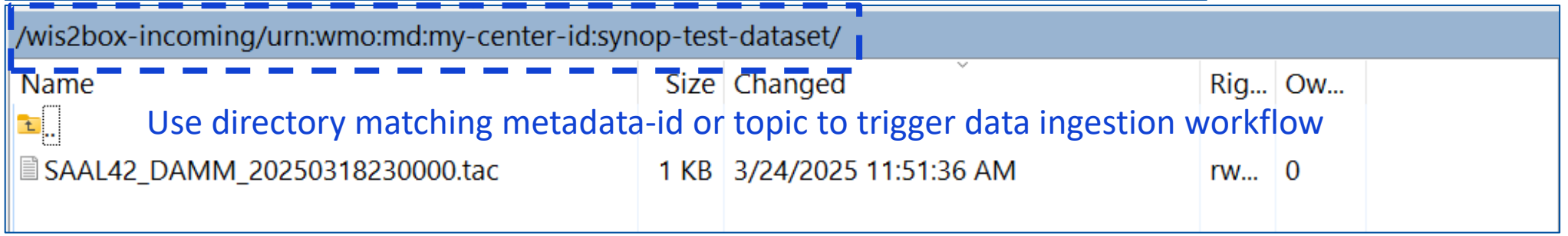


MinIO access over SFTP is enabled on port 8022

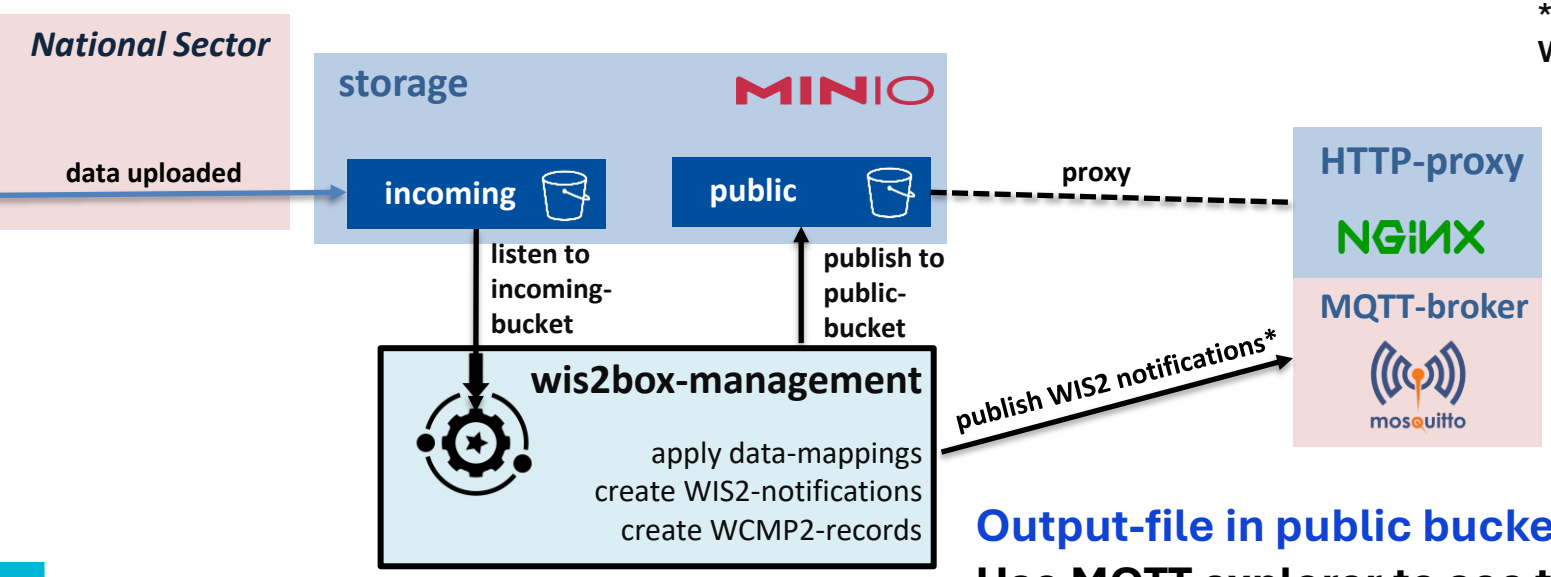
Login with WIS2BOX_STORAGE_USERNAME/WIS2BOX_STORAGE_PASSWORD

For security: restrict network access on port 8022 to known source-IP





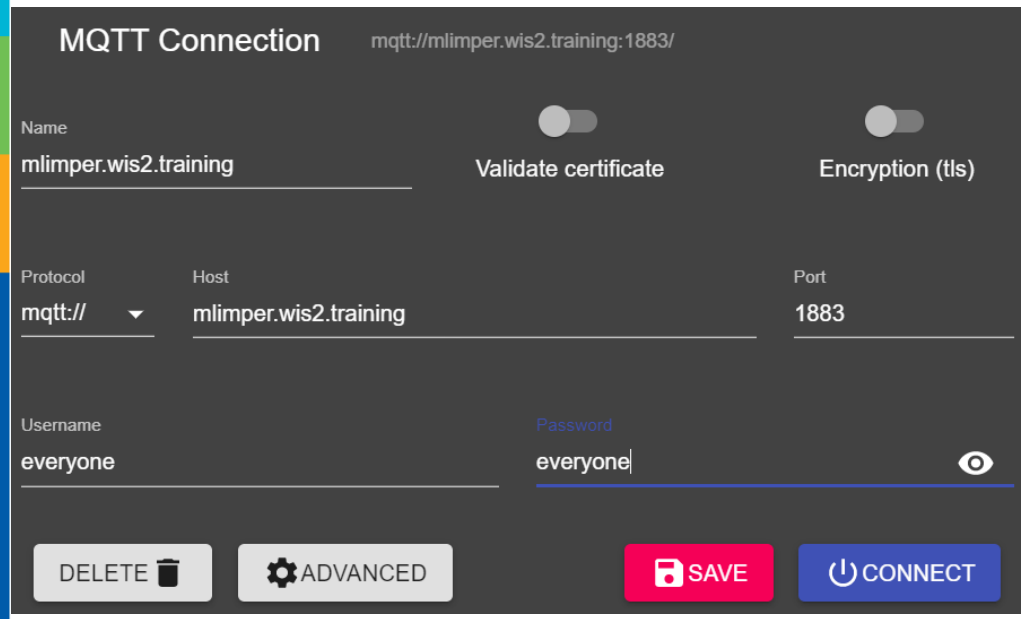
Check published result

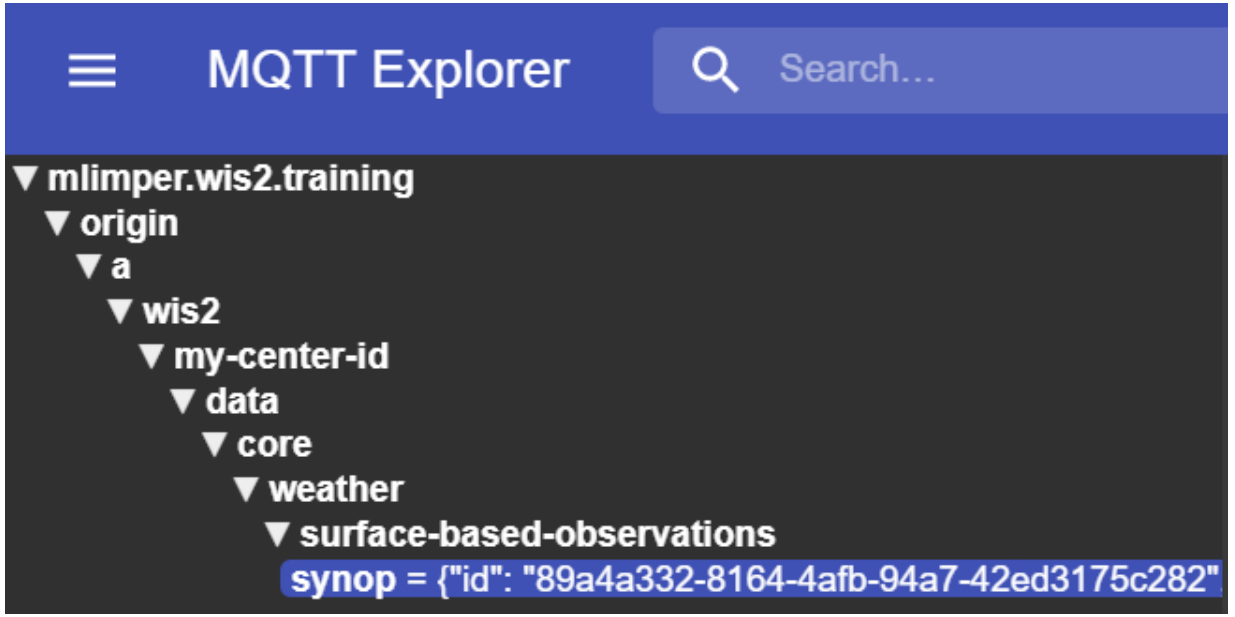


* Since the HTTP-proxy exposes wis2box-public bucket as '/data', WIS2 Notifications will use href="$WIS2BOX_URL/data/..."

Output-file in public bucket results in WIS2-notification on MQTT

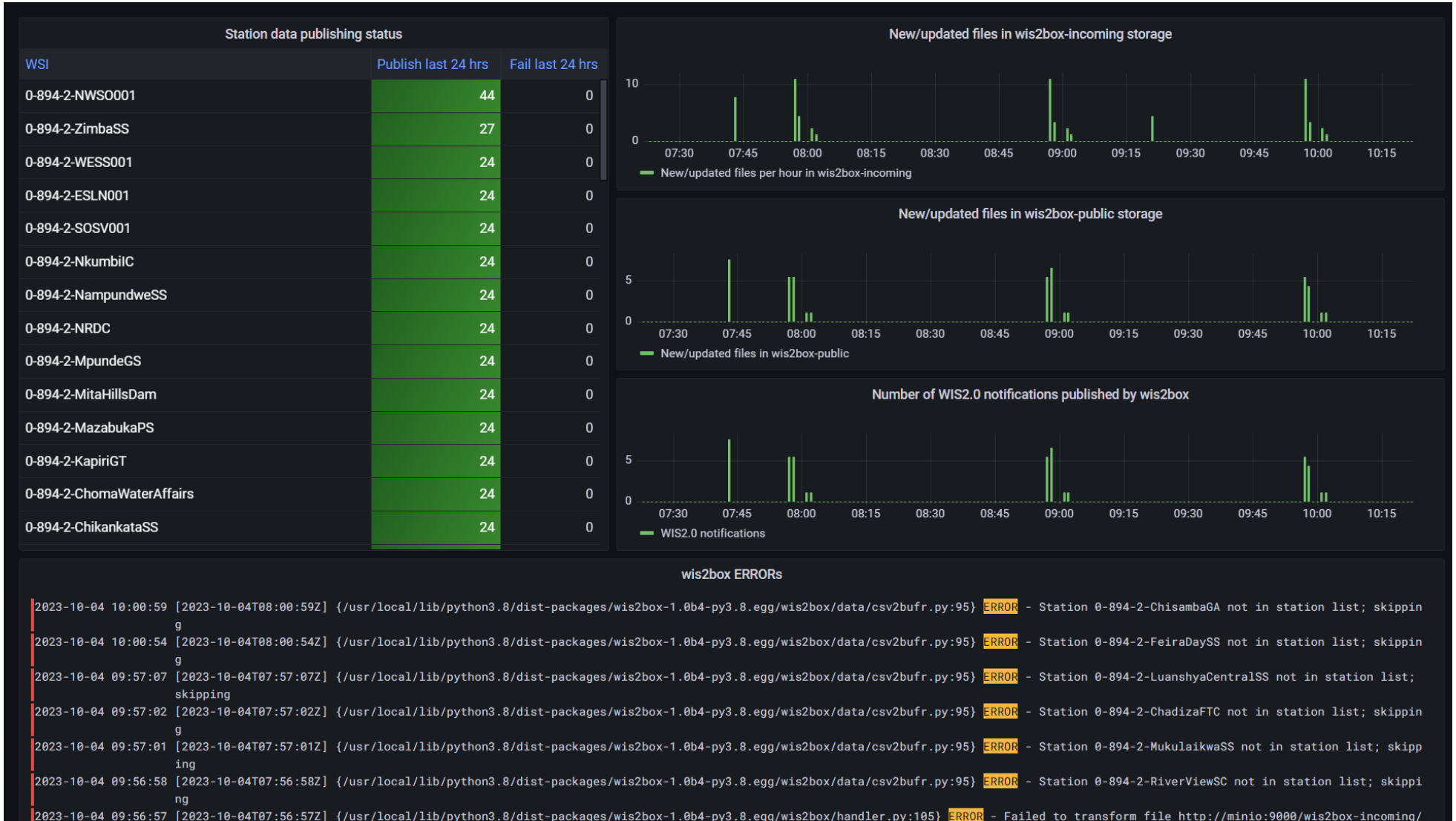
Use MQTT explorer to see the WIS2-notifications for your data





Monitoring wis2box data flow (and check errors)

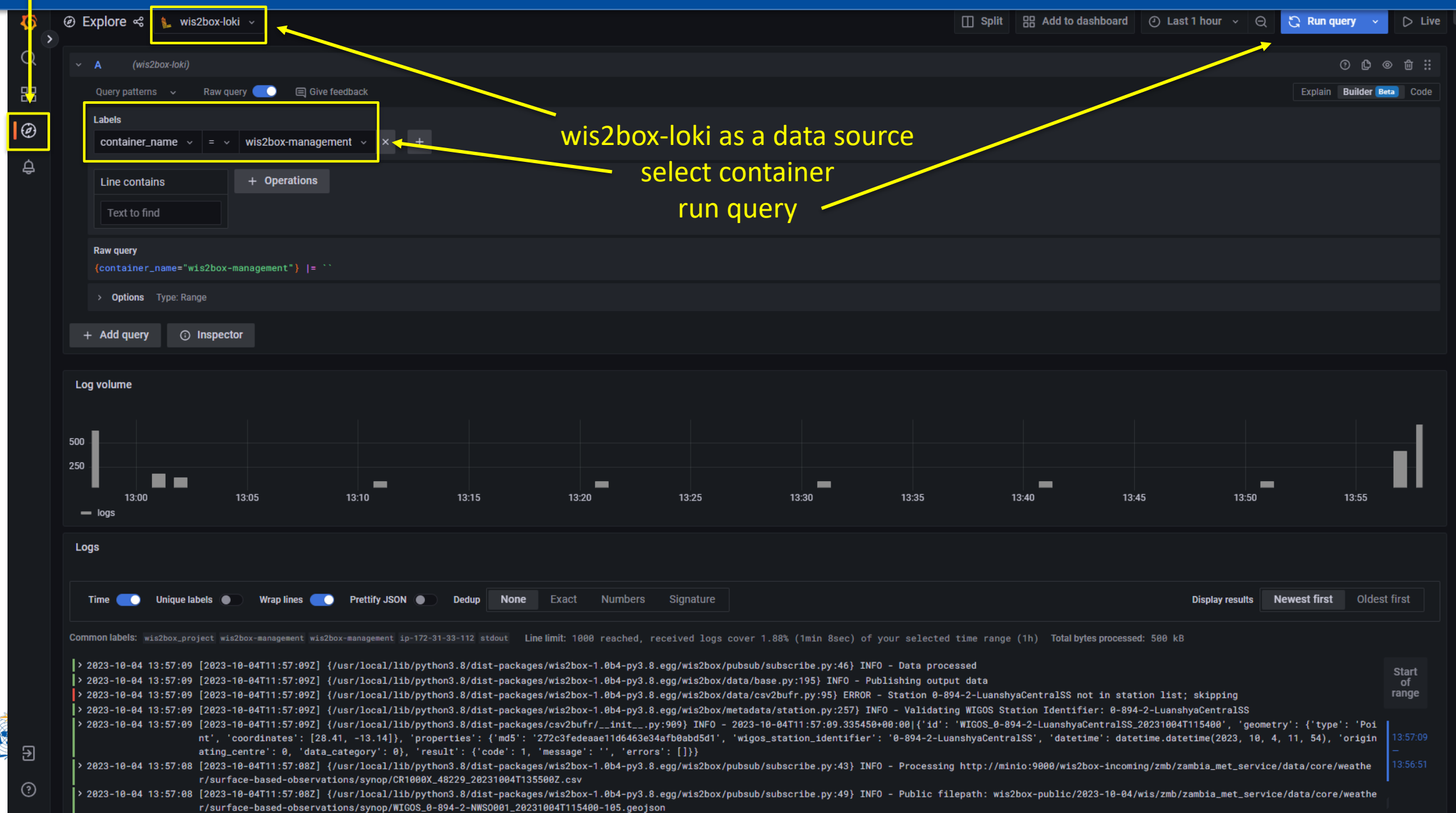
Grafana home dashboard available at http://<wis2box-host-address>:3000
Grafana data-sources: **Prometheus** (time-series metrics) and **Loki** (logs)



WORLD METEOROLOGICAL ORGANIZATION

Exploring logs stored in Loki

'explore' option In Grafana sidebar



Data Ingestion and Monitoring

Data-driven workflow triggered when data is uploaded to the MinIO service:

- **MinIO-interface** to upload data manually for **testing** the workflow
- Interfaces in wis2box-webapp: FM-12 form, manual file upload
- **Automate** data upload using **scripts** or forwarding over **SFTP**

Status of the data ingestion can be **monitored** in the **Grafana Dashboard**

After configuration, **ensure the system is maintained and remains operational** :

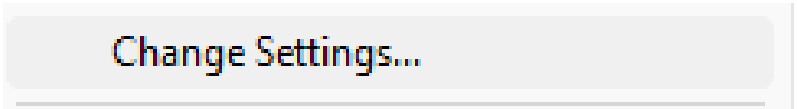
- Check for any errors (view Grafana dashboard)
- Add/update/remove datasets
- Add/update/remove stations
- Review the data content being published

Practical exercises

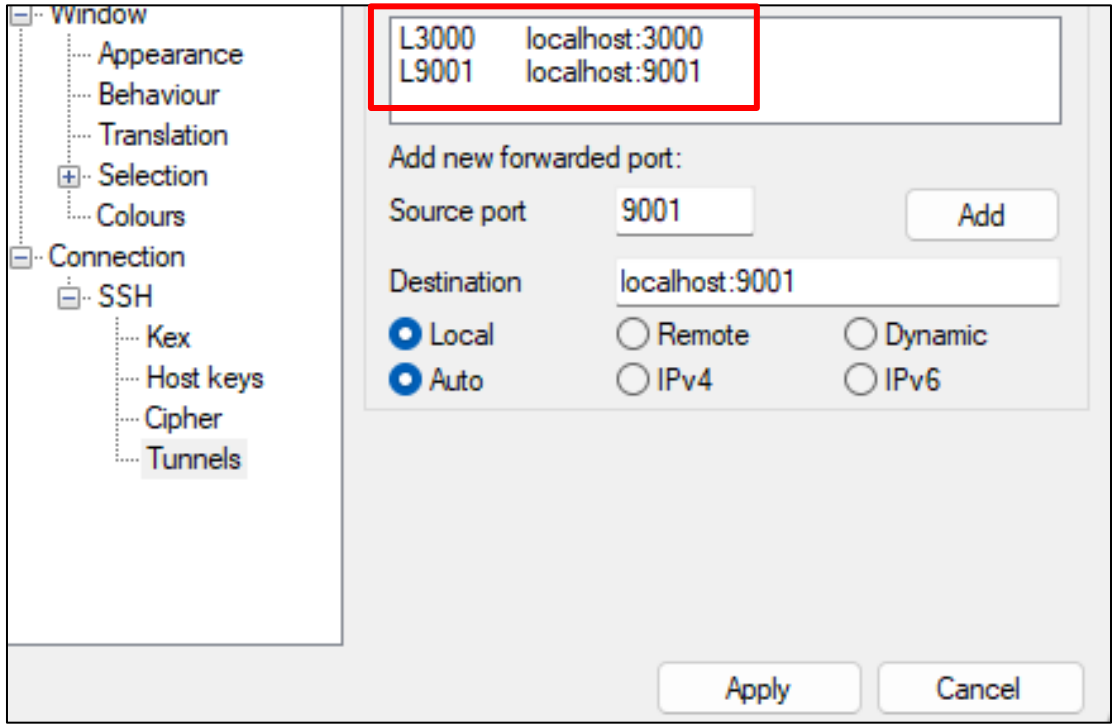
PRACTICAL EXERCISES:
https://training.wis2box.wis.wmo.int/practical-sessions/ingesting-data-for-publication/

Ports 3000/9001 for Grafana and MinIO interface are NOT available externally on European Weather Cloud, **please use SSH tunnels to access Minio and Grafana**

How to setup SSH Tunnels in PuTTY:
right click on top of window and select "Change settings"



Go to Connection->SSH->Tunnels
Add tunnels for ports 3000 and 9001
click "Apply"
Access MinIO at http://localhost:9001
Access Grafana at http://localhost:3000



WORLD METEOROLOGICAL ORGANIZATION
1950-2025 SCIENCE for ACTION

Thank you

wmo.int

WORLD METEOROLOGICAL ORGANIZATION

1950-2025 SCIENCE for ACTION