

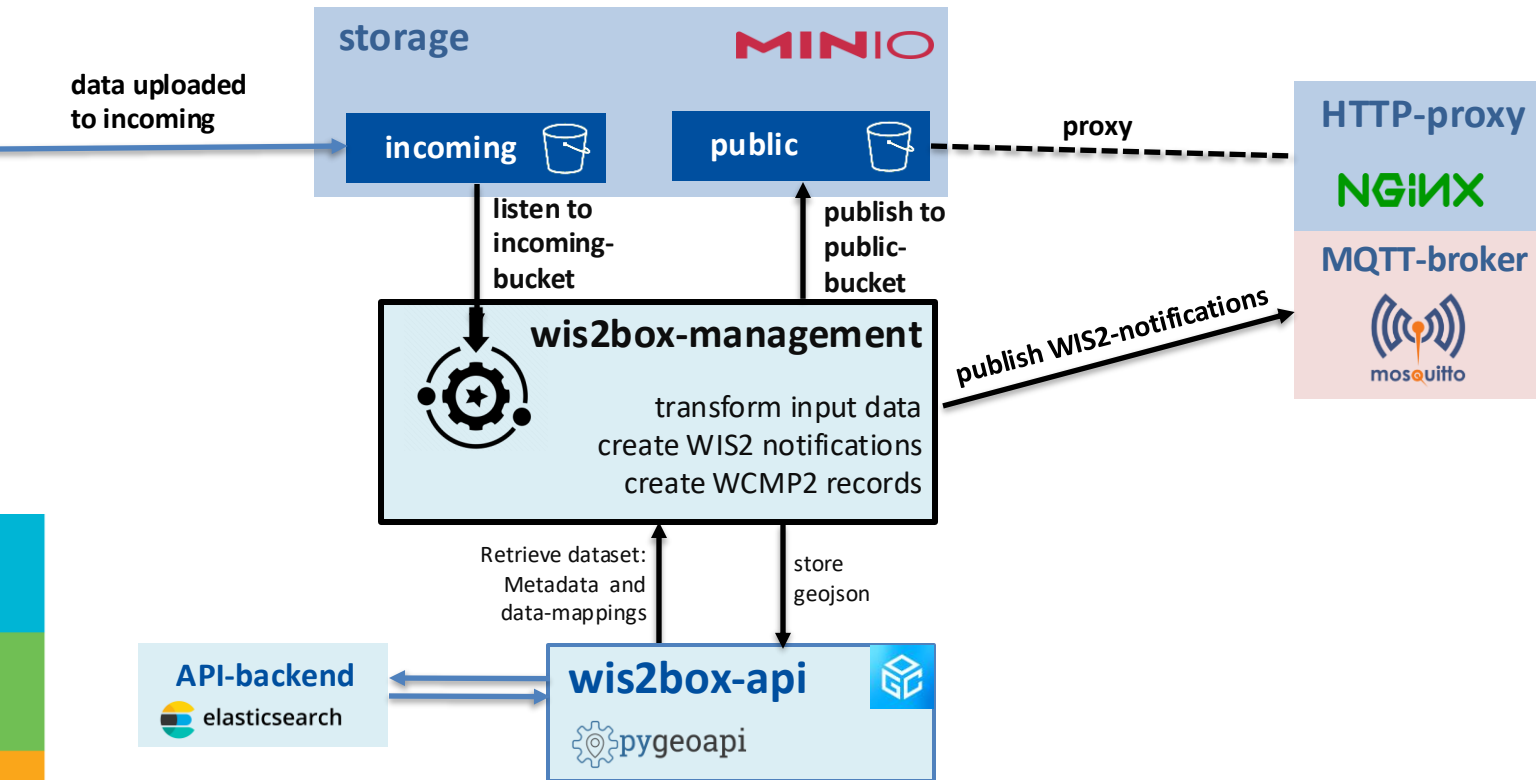
Датасеты в wis2box: метаданные и отображение данных

Подход на основе датасетов

Рабочий процесс wis2box основан на Датасетах для публикации оповещений:

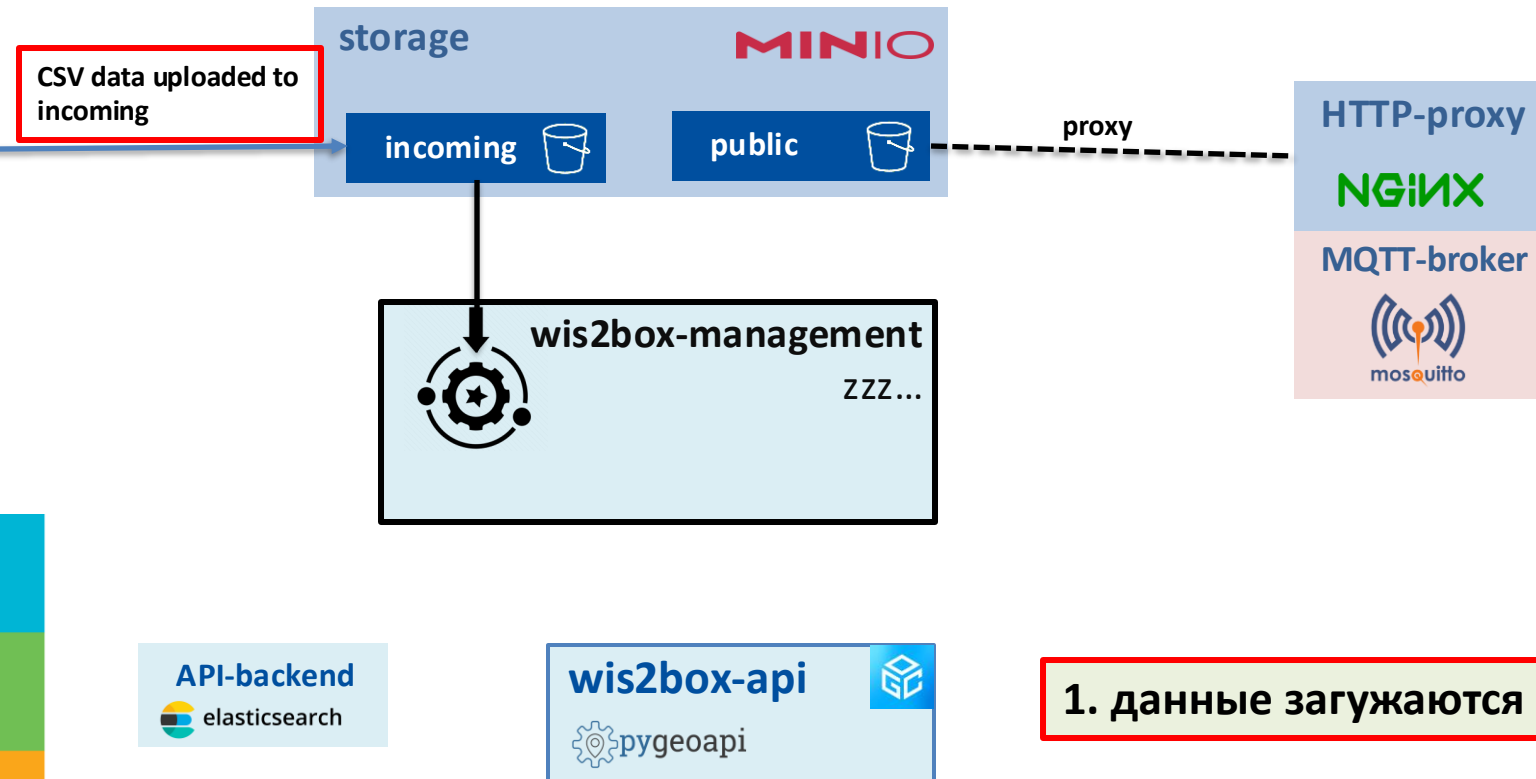
- Датасет описывается **метаданными обнаружения** в формате WCMP2
- Датасет объединяет данные и метаданные **по идентификатору метаданных**
- Датасет определяет **иерархию топиков** для публикации оповещений ИСВ2
- *wis2box: датасет имеет **плагины данных для трансформации и публикации данных***

Подход на основе датасетов



Датасеты определяют действия при загрузке файла в бакет 'incoming' хранилища

Подход на основе датасетов: шаг за шагом



1. данные загружаются в бакет incoming

Подход на основе датасетов: шаг за шагом

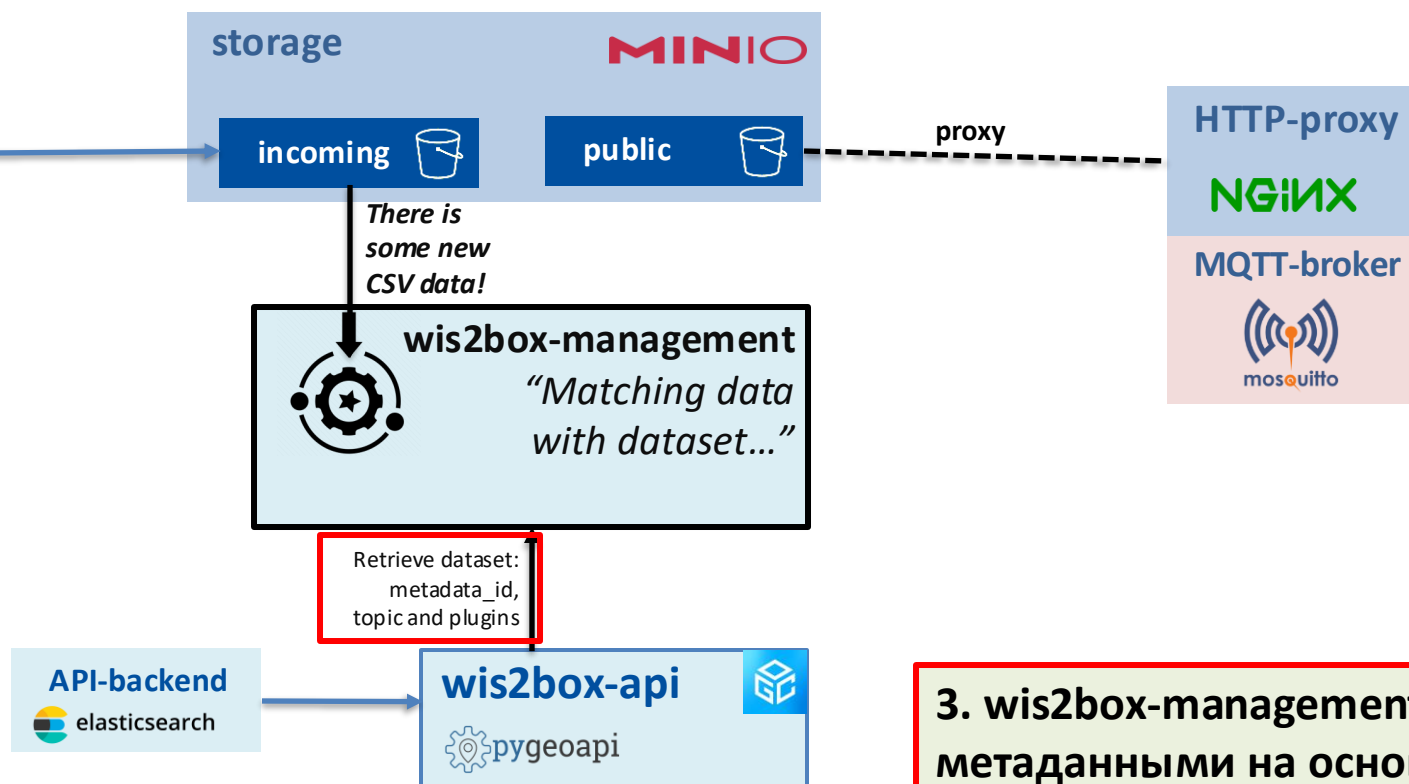


API-backend
elasticsearch

wis2box-api
pygeoapi

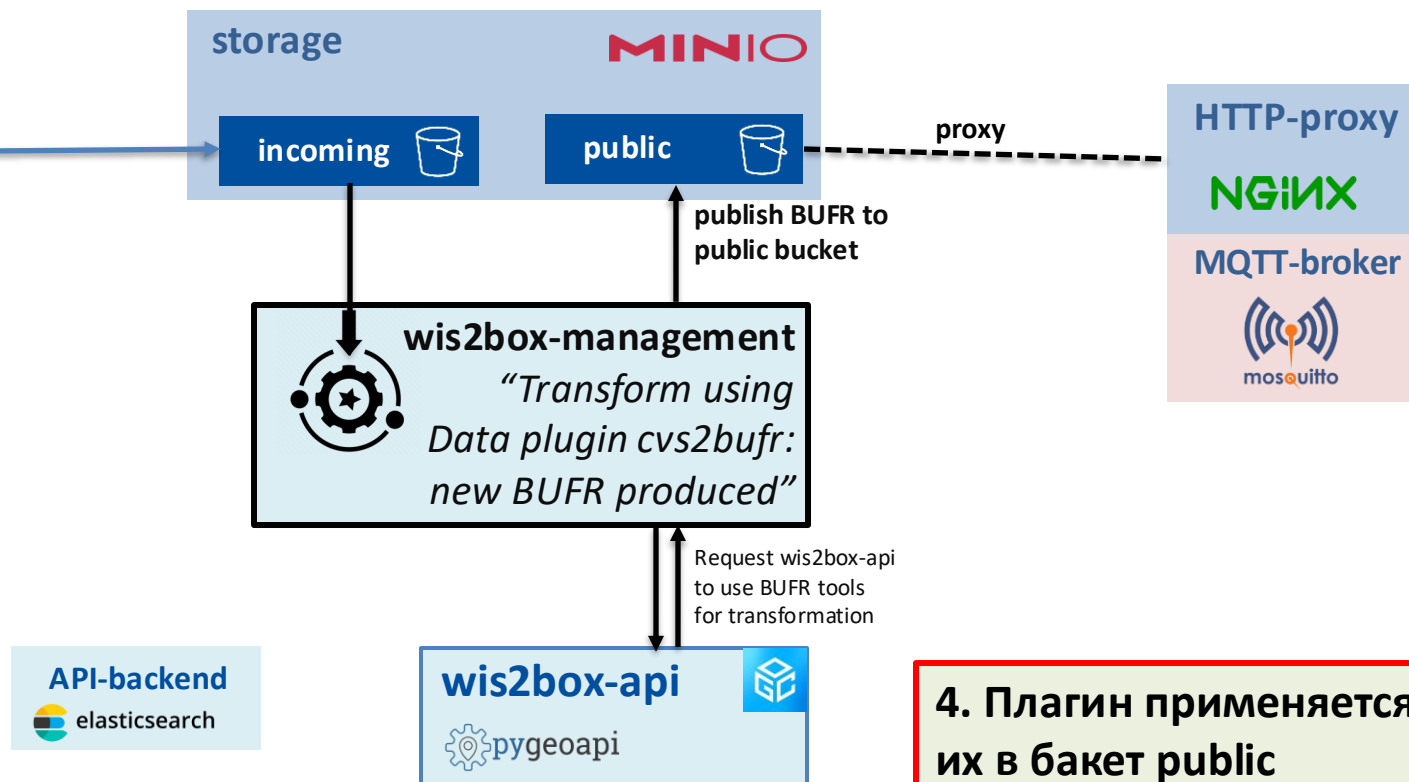
2. MinIO оповещает wis2box-management о новых данных

Подход на основе датасетов: шаг за шагом



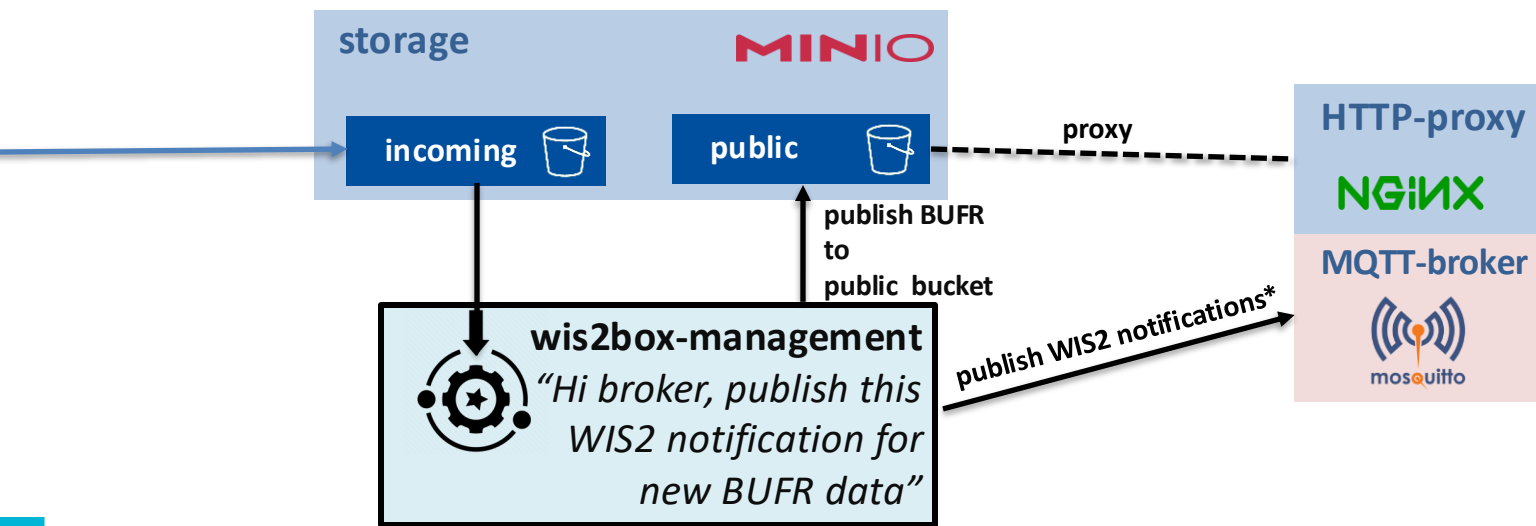
3. wis2box-management отождествляет входящие данные с метаданными на основе пути к файлу и определяет, какие плагины данных применить

Подход на основе датасетов: шаг за шагом



4. Плагин применяется для трансформации данных и публикации их в бакет public

Подход на основе датасетов: шаг за шагом



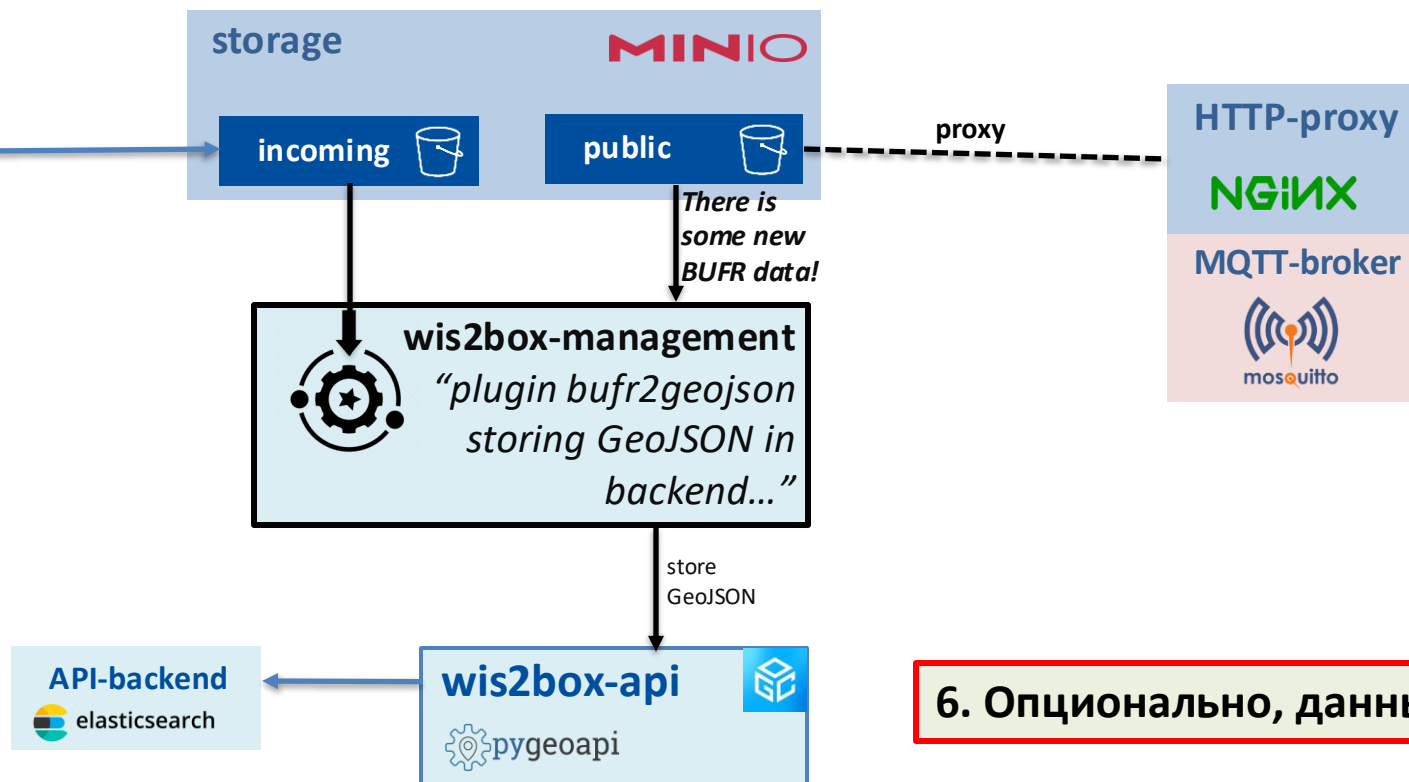
API-backend
elasticsearch

wis2box-api
pygeoapi

5. wis2box-management публикует оповещение о новых данных в соответствующий топик

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

Подход на основе датасетов: шаг за шагом



6. Опционально, данные сохраняются в GeoJSON

Датасеты в wis2box

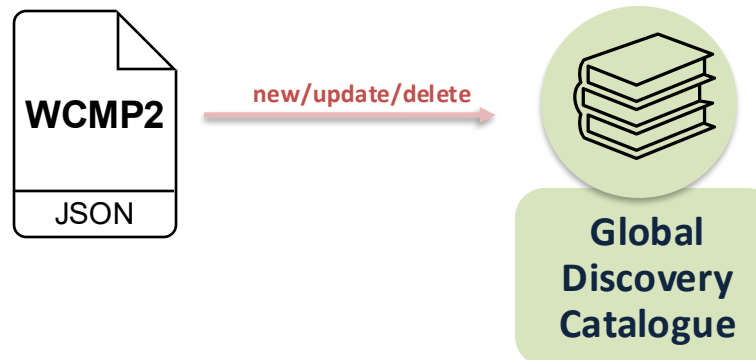
Два способа создать датасет в wis2box:

- Использовать **dataset-editor** в wis2box-webapp
- ... поместить файл MCF в контейнер wis2box-management выполнить '**wis2box dataset publish <file-path>**'

Оповещение о **новых** WCMP2 в origin/a/wis2/<centre-id>/metadata
для каждого нового опубликованного датасета

Оповещение об **изменении** WCMP2 в origin/a/wis2/<centre-id>/metadata
для каждого измененного датасета

Оповещение об **удалении** WCMP2 в origin/a/wis2/<centre-id>/metadata
для каждого удаленного датасета



Идентификатор центра (centre-id)

Для задания датасета следует предоставить centre-id вашего центра ИСВ

“Идентификатор центра (centre-id) - акроним, предложенный членом ВМО и согласованный секретариатом ВМО. Это единый идентификатор, состоящий из домена верхнего уровня (Top Level Domain (TLD)) и имени центра, и представляет издателя или распространителя данных или продукции”

Часть 1: TLD (в нижнем регистре)

uk-metoffice

br-inmet

cn-cma

id-bmkg

Часть 2: имя центра (в нижнем регистре), может содержать дефисы

Редактор метаданных в wis2box-webapp

Вас спросят задать “*Centre ID*”, который определит metadata-id и centre-id в топике

Вас спросят выбрать “Template”:

Выберите **шаблон** для инициализации датасета фиксированным топиком и значениями по умолчанию для ключевых слов и отображений данных

Выберите ‘**other**’ для произвольных типов данных и топик на основе иерархии топиков ИСВ2

New dataset: Initial Information ?

Centre ID
int-wmo-example

Template
other

- climate/surface-based-observations/daily
- ocean/surface-based-observations/drifting-buoys
- weather/surface-based-observations/ship
- weather/surface-based-observations/synop
- weather/surface-based-observations/temp

other

Dataset Editor in wis2box-webapp

Data Type = *weather/surface-based-observation/synop*

The Topic Hierarchy is fixed to *.../weather/surface-based-observations/synop*

Please enter some initial information 

Centre ID 

nl-knmi

Data Type 

weather/surface-based-observations/synop

RETURN HOME 

CONTINUE TO F

Centre ID 

it-meteoam

WMO Data Policy 

core

Discipline topic 

weather

Sub-discipline topics (choose one) 

surface-based-observations/synop

☐ experimental (free-text topic)

Topic Hierarchy 

origin/a/wis2/it-meteoam/data/core/weather/surface-based-observations/synop



1950-2025
SCIENCE for ACTION

Редактор метаданных в wis2box-webapp

Dataset Editor Form

Please choose a datasetDataset

Dataset loaded successfully.

Metadata Editor

Dataset Identification

Title

Hourly synoptic observations from fixed-land stations (SYNOP) (br-inmet)

Description

Observation data from automatic weather stations

Identifier

urn:wmo.md:br-inmet:surface-based-observations.synop

Centre ID

br-inmet

WMO Data Policy

core

Topic Hierarchy

br-inmet/data/core/weather/surface-based-observations/synop

Earth System Disciplines

Weather

Keywords (3 minimum)

+

observations

temperature

visibility

precipitation

pressure

clouds

snow depth

evaporation

radiation

wind

total sunshine

humidity

Temporal Properties

2024-06-05

End Date in UTC

☒ Dataset ongoing

Resolution

1

Unit

hour(s)

Spatial Properties

Choose an automatic bounding box (optional)

Brazil

Your country may not have an automatic bounding box

North Latitude

5.24448639!

West Longitude

-73.9872354

South Latitude

-33.7683777

East Longitude

-34.7299934

+

-



OpenStreetMap contributors

Шаг 1. Создайте и валидируйте метаданные
Шаг 2. Выберите плагины данных
Шаг 3. Подтвердите публикацию датасета

Contact Information of the Data Provider

Organization Name

WMO

URL

https://wmo.int

Country

Switzerland

Email

wis2-support@wmo.int

+

 Phone number (optional)

RESET FORM

VALIDATE FORM

Dataset Mappings Editor

Plugins in use	File extension	File pattern		
BUFR data converted to BUFR	bin	^.*\.bin\$	UPDATE	DELETE
FM-12 data converted to BUFR	txt	^.*_(\d{4})_(\d{2}).*\.txt\$	UPDATE	DELETE
BUFR data converted to BUFR	b	^.*\.b\$	UPDATE	DELETE
CSV data converted to BUFR	csv	^.*\.csv\$	UPDATE	DELETE
BUFR data converted to GeoJSON	bufr4	^*WIGOS_(\d-\d+ \d+ \w+)_.*\.bufr4\$	UPDATE	DELETE

ADD A PLUGIN

Authentication Token

wis2box auth token for 'processes/wis2box'

EXPORT AS JSON

SUBMIT

METEOROLOGICAL
ORGANIZATION

SCIENCE FOR ACTION

Датасет в формате YAML и командная строка

Датасет в wis2box может быть задан описанием в файле формата YAML.

Используя формат проекта `pygeometa` - `metadata control file (MCF)`:

```
mcf:
  version: 1.0

metadata:
  identifier: urn:wmo:md:cd-brazza_met_centre:surface-weather-observations
  hierarchylevel: dataset

identification:
  title: Surface weather observations from Republic of Congo
  abstract: Surface weather observations from Republic of Congo
  dates:
    creation: 2023-03-26
  keywords:
    default:
      keywords:
        - surface weather
        - temperature
        - observations
    wmo:
      keywords:
        - weather
      keywords_type: theme
      vocabulary:
        name: Earth system disciplines as defined by the WMO Unified Data Policy, Resolution 1 (Cg-Ext(2021), Annex 1.
        url: https://codes.wmo.int/topic-hierarchy/earth-system-discipline
  extents:
    spatial:
      bbox: [11.0937728207,-5.03798674888,18.4530652198,3.72819651938]
      crs: 4326
    temporal:
      begin: 2023-03-26
      end: null
      resolution: P1H
  url: https://example.org/malawi-surface-weather-observations
  wmo_data_policy: core
```

```
contact:
  host:
    organization: Congo National Meteorological Service (DMN)
    url: https://www.dirmet.cg
    individualname: Firstname Lastname
    positionname: Position Name
    address: P.O. Box 208
    city: Brazzaville
    administrativearea: null
    postalcode: null
    country: Republic of Congo
    email: you@example.com
```

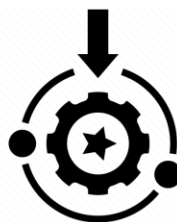
```
wis2box:
  retention: P30D
  topic_hierarchy: dz-alger_met_centre/data/core/weather/surface-based-observations/synop
  country: dza
  centre_id: dz-alger_met_centre
  data_mappings:
    plugins:
      bufr4:
        plugin: wis2box.data.bufr4.ObservationDataBUFR
        notify: true
        buckets:
          - ${WIS2BOX_STORAGE_INCOMING}
          - file-pattern: '^.*\bfr4$'
        plugin: wis2box.data.bufr2geojson.ObservationDataBUFR2GeoJSON
        buckets:
          - ${WIS2BOX_STORAGE_PUBLIC}
          - file-pattern: '^WIGOS_(\d-\d+-\d+-\w+)_.*\bfr4$'
```

MCF публикуется командой ‘wis2box dataset publish’ в контейнере wis2box-management:

```
wis2box dataset publish /data/wis2box/metadata-MCF.yml
```

При использовании MCF, вся ответственность за качество и полноту метаданных лежит на пользователе!

wis2box плагины данных



Отображение данных сопоставляет датасет с набором плагинов данных

Плагины данных используют абстрактную модель или подход для обеспечения расширяемости и переиспользования

Плагин данных определяет действия, предпринимаемые по **трансформации** и **публикации** данных

Смотри github.com/wmo-im/wis2box/tree/main/wis2box-management/wis2box/data

wis2box плагины данных

wis2box содержит следующие плагины по умолчанию:

- wis2box.data.universal.UniversalData
- wis2box.data.cap_message.CAPMessageData
- wis2box.data.bufr4.ObservationDataBUFR
- wis2box.data.synop2bufr.ObservationDataSYNOP2BUFR
- wis2box.data.csv2bufr.ObservationDataCSV2BUFR
- wis2box.data.bufr2geojson.ObservationDataBUFR2GeoJSON

Plugin Configuration

Plugin Name

File Extension

ets

Notify

Universal data without conversion

BUFR data converted to BUFR

FM-12 data converted to BUFR

CSV data converted to BUFR

BUFR data converted to GeoJSON

CAP messages

Приветствуется разработка новых плагинов для wis2box!

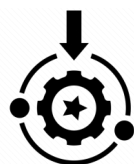
wis2box плагины данных: synop2bufr

File containing SYNOP
messages (FM-12)

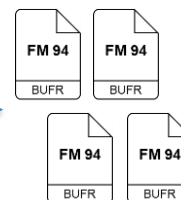


Station list

synop2bufr



One or more
BUFR files



bufr2geojson



GeoJSON

Plugin Name

- Universal data without conversion
- BUFR data converted to BUFR
- FM-12 data converted to BUFR**
- CSV data converted to BUFR
- BUFR data converted to GeoJSON
- CAP messages

plugins:

txt:

- plugin: wis2box.data.synop2bufr.ObservationDataSYNOP2BUFR

notify: true

file-pattern: '^*(\d{4})(\d{2}).*\.txt\$'

bufr4:

- plugin: wis2box.data.bufr2geojson.ObservationDataBUFR2GeoJSON

file-pattern: '^WIGOS_(\d-\d+-\d+-\w+)_.*\bufr4\$'

wis2box плагины данных: csv2bufr

Tabulated CSV data from
observing station, including
location

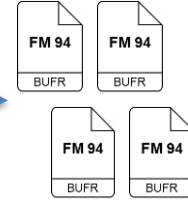


Station list

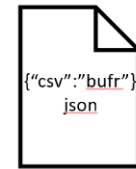
csv2bufr



One or more
BUFR files



bufr2geojson



mapping template



GeoJSON

Plugin Name

Universal data without conversion

BUFR data converted to BUFR

FM-12 data converted to BUFR

CSV data converted to BUFR

BUFR data converted to GeoJSON

CAP messages

plugins:

csv:

- plugin: wis2box.data.csv2bufr.ObservationDataCSV2BUFR
- template: aws-template.json
- notify: true
- file-pattern: '^.*\.csv\$'

bufr4:

- plugin: wis2box.data.bufr2geojson.ObservationDataBUFR2GeoJSON
- file-pattern: '^WIGOS_(\d-\d+-(\d+-(\w+))_.*\.bufr4\$'

wis2box плагины данных: bufr2bufr

File containing one or more
BUFR subsets

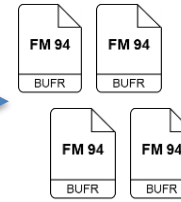
BUFR processing
(extraction of subsets)

One or more
BUFR files

bufr2geojson



Station list



GeoJSON

Plugin Name

Universal data without conversion

BUFR data converted to BUFR

FM-12 data converted to BUFR

CSV data converted to BUFR

BUFR data converted to GeoJSON

CAP messages

plugins:

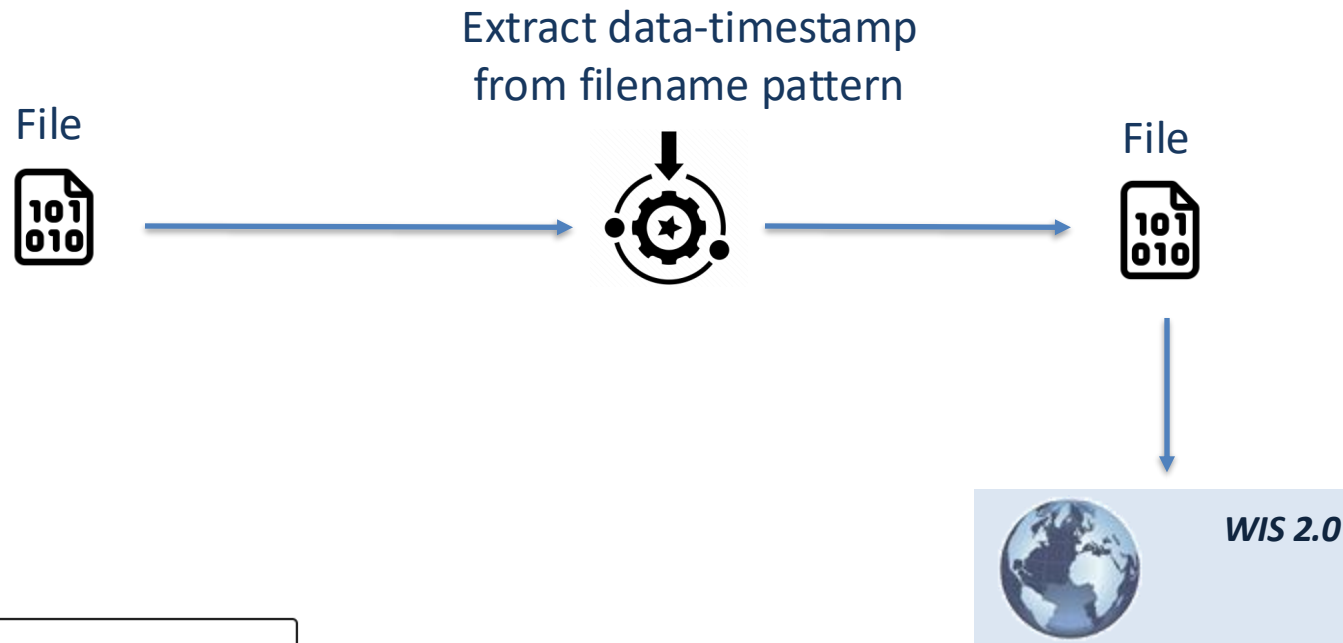
bin:

- plugin: wis2box.data.bufr4.ObservationDataBUFR
- notify: true
- file-pattern: '^.*\.bin\$'

bufr4:

- plugin: wis2box.data.bufr2geojson.ObservationDataBUFR2GeoJSON
- file-pattern: '^WIGOS_(\d+-(\d+-(\d+)-\w+)-.*\.bufr4\$'

wis2box плагины данных: universal/passthrough



Plugin Name

Universal data without conversion

BUFR data converted to BUFR

FM-12 data converted to BUFR

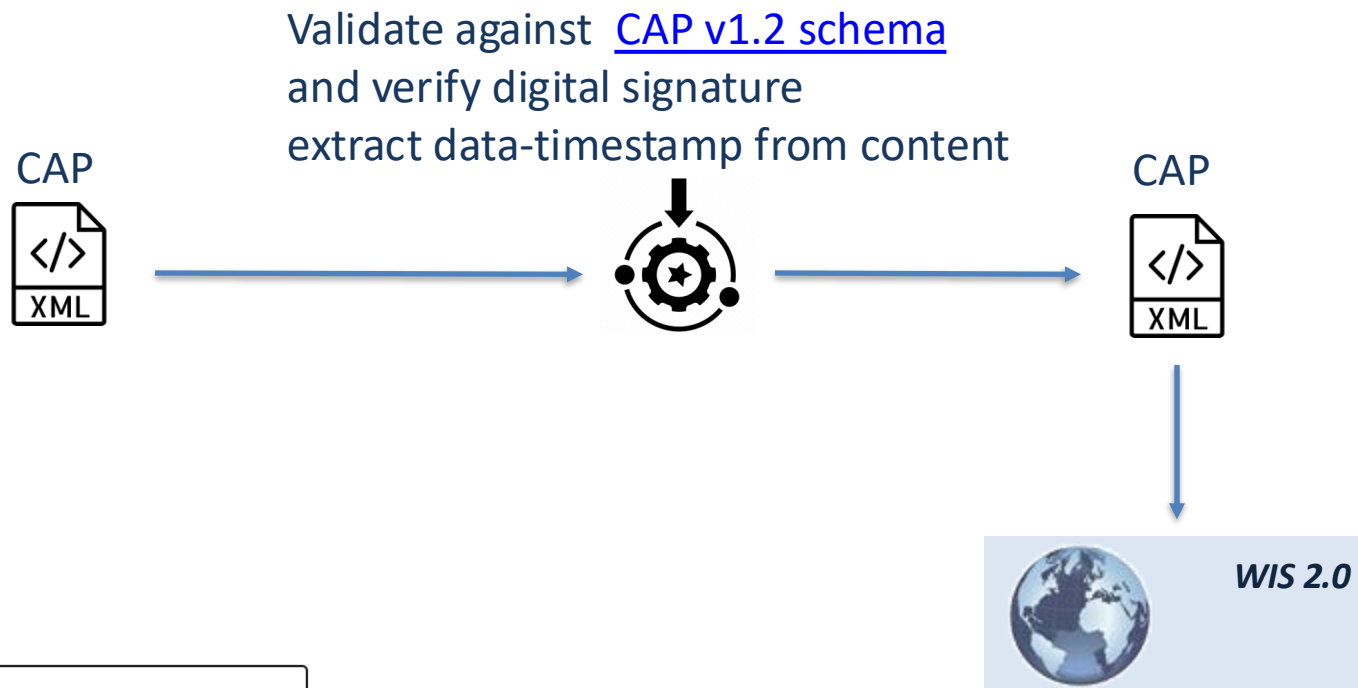
CSV data converted to BUFR

BUFR data converted to GeoJSON

CAP messages

```
plugins:
  grib2:
    - plugin: wis2box.data.universal.UniversalData
      notify: true
      buckets:
        - ${WIS2BOX_STORAGE_INCOMING}
      file-pattern: '^.*_(\d{8})\d{2}.*\.grib2$'
```

wis2box плагины данных: Common Alerting Protocol (CAP)



Plugin Name

- Universal data without conversion
- BUFR data converted to BUFR
- FM-12 data converted to BUFR
- CSV data converted to BUFR
- BUFR data converted to GeoJSON
- CAP messages**

```
plugins:
  xml:
    - plugin: wis2box.data.cap_message.CAPMessageData
      notify: true
      buckets:
        - ${WIS2BOX_STORAGE_INCOMING}
      file-pattern: '^.*\.xml$'
```

Итого

Подход на основе плагинов в wis2box-managment: должен быть настроен датасет для публикации оповещений ИСВ2 для входящих данных

Датасеты в wis2box состоят из двух компонент:

- **Метаданные обнаружения** для публикации WCMR2 в Глобальный Каталог Обнаружения
- **Отображение данных**, определяющее плагины данных для трансформации и публикации данных

Датасеты могут быть созданы в dataset-editor в wis2box-webapp

... или используя файлы YAML и командную строку в wis2box-management

Спасибо

wmo.int



WORLD
METEOROLOGICAL
ORGANIZATION

