Challenges

Guide for Applicants

Launch of the Challenges Wednesday, 26 May 2021 at 4:00 PM (CEST)

Submission deadline, Thursday, 1 July 2021 at 6:00 PM (CEST)

1. What is Fiware4Water?

FIWARE4Water is a European research and innovation project aimed at creating more efficient water management processes and bringing together communities of innovation suppliers and end-users. The project links the water sector to the smart solution platform 'FIWARE' via demonstration projects in the water cycle. FIWARE makes it possible to combine and process various data streams, such as water quality and meteorological information, into a standardised output. The water sector can then use this output, for instance, to improve operational management, create maps and build AI systems.

Fiware4Water is moving ahead diving into the digital transformation of the water sector and demonstrating FIWARE capabilities and the potential of its interoperable and standardized interfaces for both the water sector and end users such as cities, water utilities, water authorities, solution providers and citizens, allowing cross-domain cooperation and data exchange.

The development of modular smart applications using FIWARE and open API architecture for the real-time management of water systems will connect the end-users and innovation supplier communities for water compliant interfaces and data models on a comprehensive cross-domain platform associated with 4 demo cases for smart water applications and 3 demo networks.

Fiware4Water, is led by the International Office for Water (France) and counts with a consortium formed by 14 European partners. The project results will be validated in real scenarios in the city of Athens (Greece), Cannes (France), Amsterdam (The Netherlands) and Cranbrook (United Kingdom)

Fiware4Water consortium:

Office International de l'Eau, Easy Global Market SAS, National Technical University of Athens, University of Exeter, Deutsche Verein des Gas- und Wasserfaches, Centre national de la recherche scientifique, Etaireia Ydreyseos Kai Apochetefseos Proteyoysis Anonimi Etaireia, FIWARE, Waternet, SUEZ Smart Solutions, Business Development Group, South West Water

For more information, visit Fiware4Water website. You can also contact us by email challenges.support@fiware4water.eu. More details on how to contact us can be found in section 15 of this document.

2. What does Fiware4Water offer?

Fiware4Water is bringing water into cross domain applications, using standardised interfaces, models and methods, also to increase interoperability. We are boosting innovation in the water domain helping to take the most of data and generate significant benefits for European SMEs. The economic consequences of the implementation of the FIWARE methodology will provide outstanding opportunities for local businesses and especially SMEs to become not only passive receivers of innovation but proactive catalysts for future developments.

The main benefits offered by Fiware4Water are:

- support SMEs and developers in creating innovative services.
- make these services as generic, interoperable and replicable as possible
- exploit those services under the FIWARE marketplace/platform
- adapt new devices (water quality sensors and smart meters) to be FIWARE-compliant and thus highly interoperable

We are supporting solutions to increase internal and external interoperability of water services within the smart city, kick-starting the creation of a new innovation ecosystem for water sector services.

We are offering an open architecture that can be used by water utilities as well as third parties to develop solutions and applications. By providing representative data of specific fields (smart meter readings, time series of flow/pressure of water distribution networks, etc.) in a standardized form, third parties can develop new ideas of how to use this data to solve specific problems (e. g. leak detection).

Fiware4Water is about creating a reference architecture for how FIWARE can be used in Water Management and test it in real pilots. We are launching a number of challenges where we provide the FIWARE platform and ask developers/SMEs/startups to come with proposals of solutions. The challenges are targeted to the definition of Machine Learning (ML) models to be integrated in a FIWARE stack that includes a framework for AI/ML services we are building within the project.

3. Fiware4Water Challenges

This competition challenges data scientists and data analysts to show how real time data can be used to support better water management services through the data generated by digital meters and providing feedback to consumers and water utilities. Evidence through data is critical to address the many challenges that the water sector is facing. Recent advances in machine learning (ML) and data analytics (DA) technologies have provided the opportunity for more efficient use of the vast amount of data generated by sensors, meters and other devices.

The aim is to improve the management of water data, detecting anomalies in water quality, sensors out of range to evaluate their impact and possible actions by implementing ML/AI algorithms. It is necessary to know how to detect false positives, and other anomalies in the interpretation of the data (processing with ML/AI) to be able to differentiate a non-significant anomaly from those that require action. In short, the solutions to the FIWARE4Water challenges should help to:

- To be able to monitor the quality of the water to be treated in real time and anticipate water treatments based on this data
- To improve the quality of service Improve efficiency to avoid extra costs in water treatment
- To Reduce decision-making times

The details for the reference architecture and the implementation required to address the challenges can be found in detailed document - ML as a service specifications

3.1 Challenge 1 - Sofiyska voda AD

Proposed by: Sofia, Sofiyska voda AD

Title: Analysis of the Sewer Network's behavior (with perspective of "Real time control of sewer network")

Summary: Sofiyska voda AD searches for a solution that can combine both data from sensors' measurements in the sewerage network (level meters, low cost temperature sensors data for Combined sewer overflow monitoring etc.) and raingauge data, to be able to perform a cross-analysis of the measured parameters knowing if they happen in dry weather or during rain events. The prospect of the solution will be a real-time control of the behavior of the sewer system.

Description: For several years since now, Sofiyska voda AD has been measuring the water levels of the sewerage network of the city of Sofia. The measurement is performed with Level Meters and controlled through the SCADA system (see Figure 1).

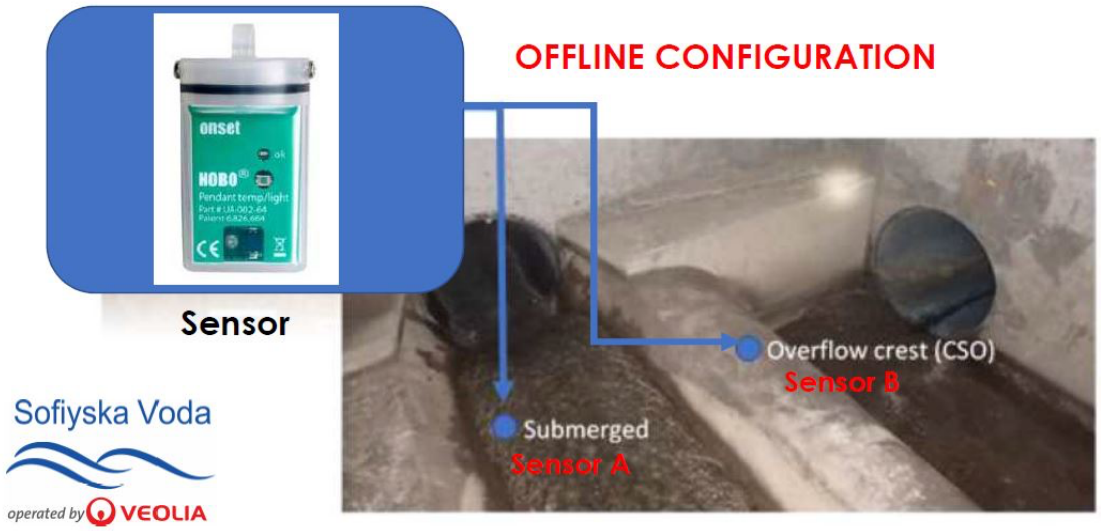

Figure 1: Installation of the water sensor

From 2019, the utility takes part in the Digital Water City Project, as a demonstration city for Low-cost Temperature sensors, monitoring the frequency of overflow of the Combined Sewer overflows (CSOs) (see Figure 2). 20 Offline Temperature sensors are already installed in 10 selected CSOs, 40 other CSOs are going to be monitored by Online Temperature sensors.

Figure 2: Installation of the water sensor

In order to calibrate the Detailed hydraulic model of the sewerage network, Sofiyska Voda AD maintains a network with 18 rain gauge stations (see Figure 3). The rain gauges transmit the information via GPRS connection, and are visualized through a WEB platform.

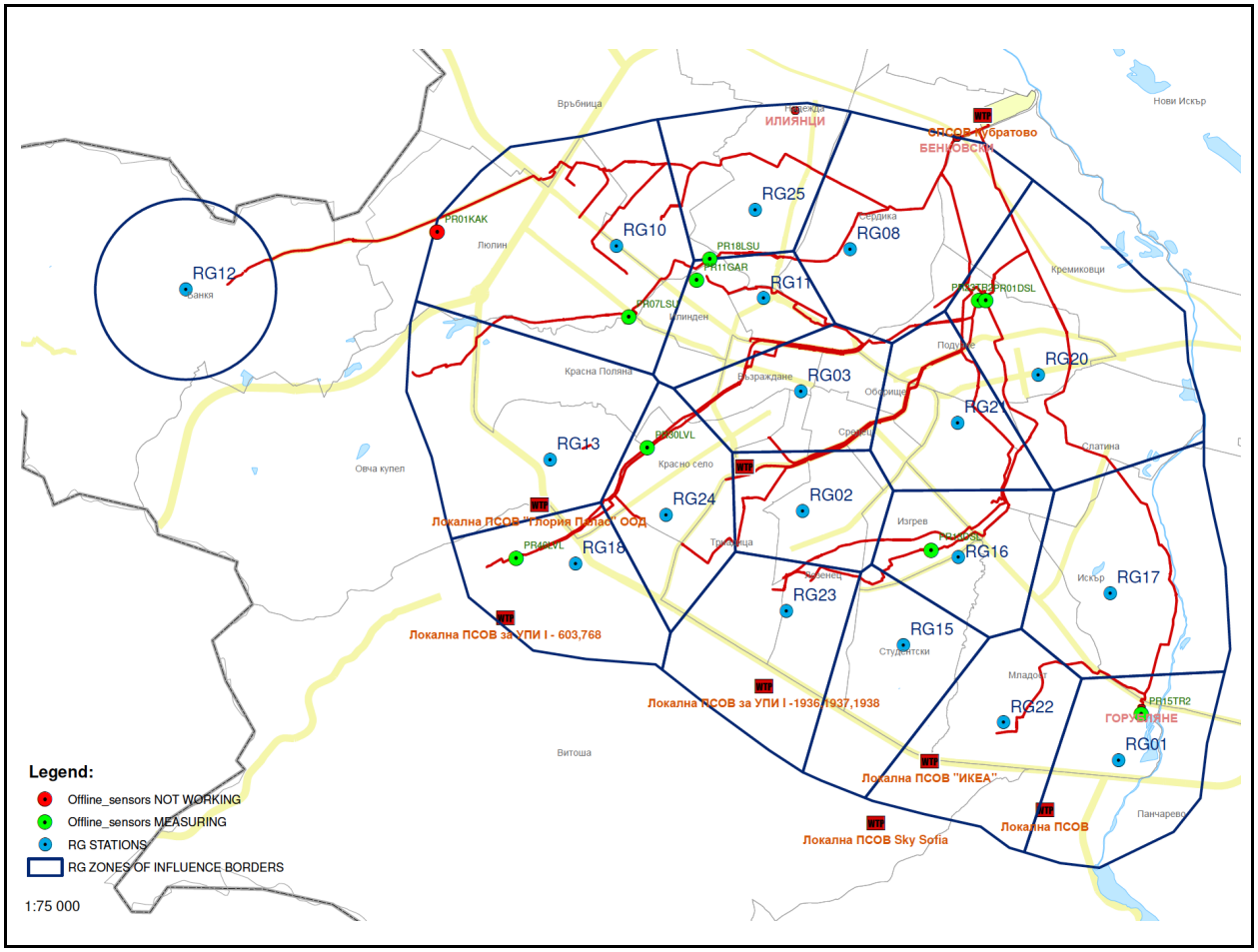

Figure 3: Location and areas of influence of the rain gauge stations in Sofia

We are searching for a solution that can combine both data (Rain Gauge metrics and the temperature of the low cost sensor in the CSO), and tell us if the overflow happens by just using temperature and rain data.

Expected Outcome:

- A predictive model capable of predicting if an overflow is happening in a specific time.
- The predictive model should predict as fast as possible the overflow.

3.1.1 Description of the data

The objective of the competition is to identify the overflow data due to rainfall or blockage of the network and make predictions. Probably the timestamps of the different measurements are not aligned, therefore some kind of aggregation should be put in place to compare different datasets.

Additionally, keep in mind that you should consult the CSO occurrence data that are within the temperature temporal range data of the sensor; in the example related to sensor PR07LSU, from 30.01.21 to 12.04.21. This way you can see if there is a relationship between the trend of the sensor temperature values and the CSO occurrence. We can assume that the lower time interval between the occurrence measurements (e.g. in PR07LSU would be 05.11.20) and the lower time interval of the sensor's temperature measurement (e.g. in PR07LSU would be 30.01.21), during that interval the occurrence data would not be relevant as we do not have temperature data to be able to analyse that behaviour.

On the other side, taking into account the upper temporal interval between occurrences (e.g. in PR07LSU would be 03.04.21) and the upper interval of temperature measurements (e.g. in PR07LSU would be 12.04.21), we can say that no CSO events occurred in that interval. However, I would not discard temperature values in this range because they may help to detect that the CSO ended on 03.04.21 as the temperature values "return to normal" (it is assumed that when a CSO occurs the temperature decreases drastically and then returns to pre-CSO values).

Keep in mind that the use of the CSV Agent facilitates the translation of the data as well as discard not needed columns.

Files

The files are separated in different folder depending on the temperature sensor id:

- <Sensor ID>, folder with the information of the Temperature sensor specified by its sensor id (e.g.: PR07LSU):
    - <Sensor ID>_SelectedVariables_temperature.xlsx - The temperature sensor data provided by the sensor (e.g.: PR07LSU) in the crest in the interval 30.01.2021 – 12.04.2021

        - Column 1: Device S/N, device serial number of the corresponding sensor. It is the same value for all the cells.
        - Column 2: Date, calendar date in format YYYY-MM-DD and time in ISO 8601-1:2019 extended format (e.g.: T08:23:05) (e.g.: 2021-04-13T08:23:05).
        - Column 3: Temperature Value ºC, the temperature measured in Celsius degrees (e.g.: 13.558).
    - <Sensor ID>_CSO_Occurrence.xlsx – The overflow occurrences detected in the system between 05.11.2020 and 03.04.2021.
        - Column 1: Date, date and time in the format MM-DD-YYYY, hh:mm:ss AM/PM (e.g.: 1/8/2021, 2:11:00 AM).
        - Column 2: Integer number, duration of the overflow event measured in minutes. A value equal to 0 indicates that the overflow event had a duration inferior to 1 minute.
    - RG<number>_01_01_2021_to_01_05_2021.csv – The rain gauge data corresponding to the area in which the <Sensor ID> is located in the time interval between 01.01.2021 and 01.05.2021.
        - Column 1: Date (Month, Day, Year) (e.g.: 2/14/21)
        - Column 2: Time (Hour, Minutes, Seconds AM/PM) (e.g.: 12:58:00 PM)
        - Column 3: Generic Output (mm/h) rainfall Intensity (mm/h) (e.g.: 6.0)

Note: the PR13DSL sensor is a little bit complex, being downstream in the Sofia's catchment, it can be also influenced by rain in the RG23, RG15 and RG02, but RG16 is the main one (see figure 1).

NGSI-LD payloads

The use of the corresponding Agent (f4w-challenge-sofia) facilitates communication with the FIWARE Context Broker through the creation of the corresponding publication NGSI-LD requests. The payload of those requests are represented in JSON-LD format, as can be seen in the following examples.

```
{
  "id": "urn:ngsi-ld:Device:RG10",
  "type": "Device",
  "category": {
    "type": "Property",
    "value": [
      "sensor"
    ]
  },
  "controlledProperty": {
    "type": "Property",
    "value": [
      "precipitation"
    ]
  },
  "dateCreated": {
    "type": "Property",
    "value": {
      "@type": "DateTime",
      "@value": "2021-01-01T00:00:00.000000000"
    }
  },
  "value": {
    "type": "Property",
    "value": 0.0,
    "unitCode": "H67"
  },
  "name": {
    "type": "Property",
    "value": "RG10"
```

```
        },
        "@context": [
          "https://smartdatamodels.org/context.jsonld"
        ]
}
```

Table 1: JSON-LD data for rain gauge sensor

```
       {
        "id": "urn:ngsi-ld:Device:10078399",
        "type": "Device",
        "dateCreated": {
          "type": "Property",
          "value": {
            "@type": "DateTime",
            "@value": "2021-01-30T00:04:43.000000000"
          }
        },
        "category": {
          "type": "Property",
          "value": [
            "sensor"
          ]
        },
        "controlledProperty": {
          "type": "Property",
          "value": [
            "temperature"
          ]
```

```
        },
        "serialNumber": {
          "type": "Property",
          "value": 10078399
        },
        "value": {
          "type": "Property",
          "value": 7.682,
          "unitCode": "CEL"
        },
        "placement": {
          "type": "Property",
          "value": "CSO"
        },
        "name": {
          "type": "Property",
          "value": "PR07LSU"
        },
        "@context": [
          "https://smartdatamodels.org/context.jsonld"
        ]
    }
```

Table 2: JSON-LD data for temperature sensor

```
{
  'id': 'urn:ngsi-ld:Device:device-005A',
  'type': 'Device',
```

```
'dateCreated': {
  'type': 'Property',
  'value': {
    '@type': 'DateTime',
    '@value': '2021-01-30T03:51:00.000000000'
  }
},
'category': {
  'type': 'Property',
  'value': [
    'sensor'
  ]
},
'controlledProperty': {
  'type': 'Property',
  'value': [
    'overflow'
  ]
},
'value': {
  'type': 'Property',
  'value': 55,
  'unitCode': 'MIN'
},
'name': {
  'type': 'Property',
  'value': 'PR13DSL
'},
'@context': [
  'https://smartdatamodels.org/context.jsonld'
]
```

```
}
```

Table 3: JSON-LD data for occurrence overflows

Data Explorer

1 MB (zip format)

Link

3.1.2 Validation

Outcomes validation:

Predictive model needs and priorities:

- The top priority of the model is to detect if an overflow occurs.
- False positives and false negatives should be minimized.
- The model needs to detect in real time the overflow.
- The overflow should be predicted as soon as possible.

Offline validation: How proposed solutions will be validated

The delivered models should predict taking into account all possible errors and be able to generalize (including not having overfitting). A set of metrics will be used to evaluate the Machine Learning models.

In case of classification predictions, the metrics will be:

- Accuracy: The proportion of correct predictions among the total number of observations.
- Precision: The fraction of positive predictions correctly predicted.
- Recall: The fraction of the positive observations successfully predicted.
- F-score: A weighted average of the precision and the recall.

Finally, if the problem requires anomaly detection, not only classification metrics will be used, but also specific AD metrics. The objective is to evaluate a set of conditions:

- The anomaly must be predicted as soon as possible.
- False positives should strictly be penalized.
- Predicting the same anomaly multiple times is not important.

Predictive model evaluation method:

- As a final step, to identify the model generalization capabilities, an evaluation set (not included in the database training data) will be used.
- The metrics used to evaluate the model are associated with the confusion matrix and the priorities already introduced.
- The evaluation will consist of evaluating if the model is capable of predicting the overflows correctly.
- Early predictions of overflow will also be considered during the evaluation.

To offline validate the results of the different solutions, participants need to deliver a docker image of the model and the pipeline of the data in a tar.gz.

The docker container must be executed using the command: docker run <image_name> <path_to_evaluation_set_file>

The evaluation set file will have the same format as the general dataset given at the beginning of the competition.

The docker execution needs to output a csv file to the same directory where the evaluation set is stored. This csv file must contain the predictions of the model (only when an anomaly is detected) in a specific format:

- Timestamp: A column indicating the timestamp of the prediction. This timestamp must match with the input register of the time series.
- Prediction: The true prediction of the model. If the anomaly to predict has different types, this should be a string specifying the type of anomaly, e.g. 'Type 1'. This column must only contain anomalous detections.
- Format example: **Head**: Timestamp, Prediction **Row**: 2020-02-23, **'blockage'**.

Online validation: - Demo day

Short demonstration of the integration of the model in the architecture. For this purpose you can use the docker-compose file provided to you for testing the model.

3.1.3 Help Kit

Help Kit: it contains the material that is necessary to create your development environment and to validate your trained machine learning models:

- Scripts to inject validation data into a context broker (and more specifically https://github.com/flopezag/f4w-challenge-sofia/tree/d15702e3c15eab2091a8d713930bdcbb4ed214e5), also see data
- Docker compose files for Stellio and Orion-LD
- Scripts to easily interact with a context broker
- Documentation on actions to be performed to deploy and validate a machine learning model packaged into BentoML

3.2 Challenge 2 - CAP

Proposed by: CAP

CAP is a large public company, which owns and manages 60 small to large wastewater treatment plants (WWTPs), serving about 2.5 million people in the peri-urban area of Milan. The WWTPs are often situated in agricultural areas.

Title: Correlation analysis of real wastewater data and laboratory data.

Domain: Wastewater treatment plant, water quality assessment.

Summary: The challenge proposed aims to find a concrete solution to explore real-time the quality of sensor data. Data quality assessment should be obtained employing periodic laboratory analyses as reference.

Description: CAP is steadily increasing the volume of wastewater monitored through online sensors. To date, within each plant, several sensors and analysers are installed. However, it is crucial to improve not only the quantity of available data but also their quality for process, safe water reuse, and compliance purposes.

Therefore, it is essential to establish a defined and continuously applied method to validate the massive amount of data with the periodic laboratory analyses. The latter would guarantee a real-time check, which would also show malfunctioning and build the bases for an optimized maintenance program.

We also believe that this challenge's results will support the deployment of several Digital Solutions developed in the Milan case of the Digital-water.city project focused on safe water reuse. Indeed, the reliability of an Early warning system for safe water reuse and the efficiency of the Match-making tool between water demand and availability heavily rely on the accuracy of the data.

Expected outcomes: The participants can work in a clustering solution to profile the different correlations observed between the real data and the laboratory data. Therefore, this task has an analytical implication (analysis of the multiple profiles and conclusions), it is expected to get an analytical report of the different profiles/clusters found and to justify the predictions of their algorithm/algorithms.

3.2.1 Description of the data

The objective of the competition is to calculate the Quality of Sensor Data based on a data lab and real data sensors. The challenge aims to understand whether the sensors could replace the lab data. Thus it seeks to check their reliability. From this standpoint, when we talked about validation and quality control of sensor data, we meant that we would like to find a correlation between them. The participants have absolute freedom to decide the best correlation methods to understand if sensor data match the lab data properly or not and create the corresponding model to analyse this information in real time. We used the index of agreement, the correlation index, t-test and other methods but it will also be ok if other solutions are suitable.

Due to the measures having different timestamps, it is a problem if we want to make a direct data analysis. Our suggestion is to make a daily average of sensors data (from 10 a.m. to 10 a.m.) in order to compare them with lab data. This is because the lab data are collected with a sampler, which gathers wastewater for 24 hours every day; therefore, taking the daily average looked to be the best approach.

Files

- lab analyses/lab_2019.xlsx, lab analyses/lab_2020.xlsx – information about the measures obtained in the laboratory. Laboratory data is the reference to monitor quality of real time data.
    - Column 1: PLANT text field with the description of the wastewater plant (e.g.: AAA).
    - Column 2: SAMPLING POINT, text field, always DISCHARGE value.
    - Colum 3: DATE, data and time of the sample in the format dd.mm.yyyy HH:MM:SS (e.g.: 10.01.2019 09:00:00).
    - Column 13: COD (Chemical Oxygen Demand) numeric values measured in mg/l. Can have measures like '----' to indicate that there is no measure or like '<15' to indicate that the measure is below 15mg/l. For the hypothesis of the analysis, we consider '<15' = MEAN(0, 15) = 7,5mg/l
    - Column 14: Total suspended solids (tss) numeric values measured in mg/l. Can have measures like '----' to indicate that there is no measure or like '<5' to indicate that the measure is below 5mg/l. For the hypothesis of the analysis, we consider '<5' = MEAN(0, 5) = 2,5mg/l
    - Column 16: Total ammonia nitrogen (as NH4) (NH4) numeric values measured in mg/l. Can have measures like '----' to indicate that there is no measure, like '<0.5' to indicate that the measure is below 0.5mg/l, and '<1' to indicate that the measure is below 1mg/l. For the hypothesis of the analysis, we consider '<0.5' = MEAN(0, 0.5) = 0,25mg/l and '<1' = MEAN(0, 1) = 0,5mg/l
    - Column 18: Nitrate (asN) (NO3) numeric values measured in mg/l. Can have measures like '----' to indicate that there is no measure, like '<1' to indicate that the measure is below 1mg/l, and '<3' to indicate that the measure is below 3mg/l. For the hypothesis of the analysis, we consider '<1' = MEAN(0, 1) = 0,5mg/l and '<3' = MEAN(0, 3) = 1,5mg/l
    - Column 21: Phosphate (P-PO4) (PO4) numeric values measured in mg/l. Can have measures like '----' to indicate that there is no measure, like '<0.2' to indicate that the measure is below 0.2mg/l, and '<0.6' to indicate that the measure is below 0.6mg/l. For the hypothesis of the analysis, we consider '<0.2' = MEAN(0, 0.2) = 0,1mg/l and '<0.3' = MEAN(0, 0.3) = 0,15mg/l
    - Column 4 – Column 12, Column 15, Column 17, Column 19, Column 20 - do not use the data from them.
- <Plant>_<Property>_real time_2019-2020.csv, measures of the sensors from the different Plants (e.g.: AAA, BBB, CCC, DDD, EEE, FFF, and GGG) and different Properties (e.g.: COD, NH4, NO3, PO4, and SST). Keep in mind that the SST Property corresponds to the Total suspended solids represented by tss.

- Column 1: Data/Ora, date and time in the format of day, month, year hours, minutes and seconds (e.g. 16/06/2019 08:54:33).
- Column 2: Valore, property value in mg/l (e.g.: 69,607).

NGSI-LD payloads

The use of the corresponding Agent (f4w-challenge-milan) facilitates communication with the FIWARE Context Broker through the creation of the corresponding publication NGSI-LD requests. The payload of those requests are represented in JSON-LD format, as can be seen in the following examples.

```
{
   "id": "urn:ngsi-ld:WaterQualityObserved:waterqualityobserved:WWTP:AAA”,
   "type": "WaterQualityObserved",
   "dateObserved": {
      "type": "Property",
      "value": {
         "@type": "DateTime",
         "@value": “2019-05-06T09:50:38.000Z”
      }
   },
   COD: {
      "type": "Property",
      "value": 16.699,
      "unitCode": “M1”
   },
   "@context": [
```

```
            "https://smartdatamodels.org/context.jsonld"

        ]

}
```

Table 4: JSON-LD measures data for Plant AAA and property Chemical Oxygen Demand

Data Explorer

12 MB (zip format)

Data files: Link

3.2.2 Validation

Outcomes validation:

Participants develop a clustering solution to profile different correlations. This task has an analytical implication (analysis of the multiple profiles and conclude). It is expected an analytical report of the different profiles/clusters found and to justify the predictions of their algorithm.

Offline validation: How proposed solutions will be validated

To evaluate the solution, which should assign a register into a cluster or profile, the teams need to deliver in https://fiware4water-challenges.devpost.com/:

- An analytical report, explaining all the findings during the solution development and the explanation of each cluster created by the algorithm.
- A docker image, detailed below, to validate the behaviour of the algorithm.

To offline validate the results of the different solutions, participants need to deliver a docker image in a tar.gz.

The docker container must be executed using the command: docker run <image_name> <path_to_evaluation_set_file>

The evaluation set file will have the same format as the general dataset given at the beginning of the competition.

The docker execution needs to output a csv file to the same directory where the evaluation set is stored. This csv file must contain the predictions of the model (the different cluster assignation) in a specific format:

- Timestamp: A column indicating the timestamp of the prediction. This timestamp must match with the input register of the time series.
- Prediction: The cluster assignation of the register.
- Format example: **Head:** Timestamp, Prediction **Row:** 2020-02-23, 'Cluster 1'**.**

Online validation: - Demo day

Short demonstration of the integration of the model in the architecture. For this purpose you can use the docker-compose file provided to you for testing the model.

3.2.3 Help Kit

Help Kit: it contains the material that is necessary to create your development environment and to validate your trained machine learning models:

- Scripts to inject validation data into a context broker (and more specifically https://github.com/flopezag/f4w-challenge-milan/tree/6230705dbcd3594b2b821d2ea4efd1cdfd991278), also see data)
- Docker compose files for Stellio and Orion-LD
- Scripts to easily interact with a context broker
- Documentation on actions to be performed to deploy and validate a machine learning model packaged into BentoML

3.3 Challenge 3 - South West Water - United Kingdom

Proposed by: South West Water

Title: Identify short term water network events and longer-term trends in flow data using anomaly detection methods and time-series analysis.

Domain: Water Consumption

Summary: The objective of the competition is to calculate the short-term and long-term water network events in order to evaluate possible bursts pipe, leakage or any change in the customer wáter consumption to be considered. It is very important to define task forces to resolve water leakage or analyse any possible improvement in the water network distribution.

Description: The purpose of the challenge is the identification of the short-term water network events as well as any long-term trend around the provided data, using for this purpose any kind of anomaly detection methods that the participant applies over the dataset. It also means that the timeseries analysis has to be considered in the execution of the challenge. The data contains flow data measured across 50 areas between 1st April 2020 and 1st March 2021 in the South West of the United Kingdom. Areas are formally known as District Metered Areas (DMAs) and are used to measure key metrics such as water demand and leakage in small areas. In rural areas, DMAs can be as large as a town or village. Urbans areas are typically made up of many DMAs.

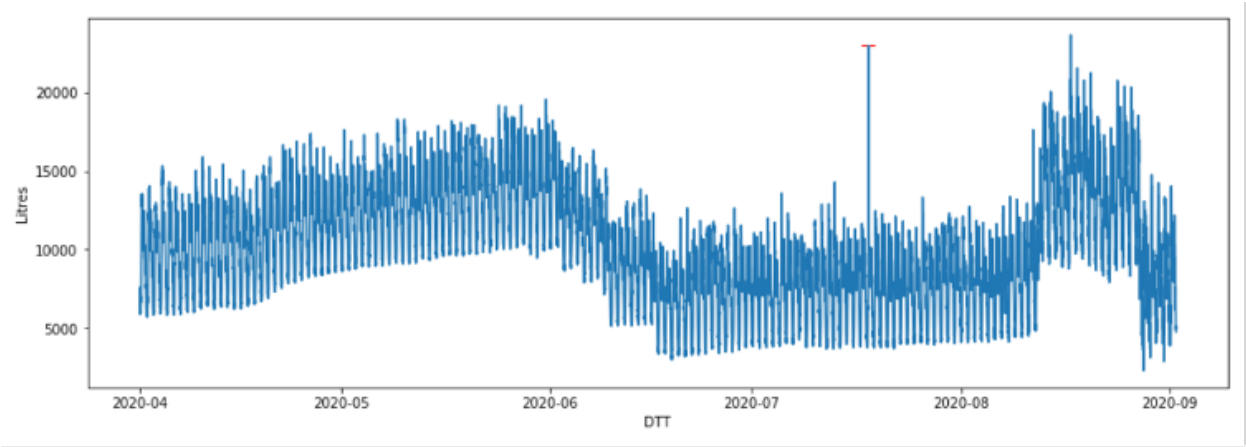

Figure 4: Flow data from District Metered Area (DMA)

Figure 1 displays a flow profile for DMA and represents the net volume of water used by the area. Water use is typically made up of domestic customers, commercial businesses or from leakage due to burst pipes.

Expected outcomes:

The expected output of this challenge is to identify:

1. Short-term events such as large pipe bursts which normally correspond to short, steep spikes in the flow profile typically lasting hours or a few days. In figure 1 this can be seen in 2020-08.
2. Longer-term trends which normally correspond to underlying leakage from small, unrepaired bursts or changes in customer water consumption such as the completion of a new housing development. This will typically last weeks to months. In figure 1 this can be seen between 2020-04 and 2020-06, where there has been a gradual increase in the flow profile followed by a gradual decline which may correspond to small burst pipes being fixed.

Note: Very short-term events (15 to 30 minutes) are likely to be data quality issues and should not be classified as bursts.

3.3.1 Description of the data

Files

Different files have been provided:

- Areas Flow Data 1 to 25, measures of water consumption
  - DMA, text value, the area of measurement (e.g.: 201D05).
  - DT, Date and Time value, Day of flow measurement (e.g.: 01/04/2020 00:00).
  - Period, numeric value, 15 minutes period of flow measurement (e.g.: 1).
  - DTT, Date and time value, datetime of flow measurement. Can be obtained from the others two columns (e.g.: 01/04/2020 00:15)
  - Litres, float number separated by ".", volume of water that flowed through the meter in a 15 minutes period (e.g.: 3576.0).
- Areas Flow Data 26 to 50, measures of water consumption
  - DMA, text value, the area of measurement (e.g.: 201D05).
  - DT, Date and Time value, Day of flow measurement (e.g.: 01/04/2020 00:00).
  - Period, numeric value, 15 minutes period of flow measurement (e.g.: 1).
  - DTT, Date and time value, datetime of flow measurement. Can be obtained from the others two columns (e.g.: 01/04/2020 00:15)
  - Litres, float number separated by ".", volume of water that flowed through the meter in a 15 minutes period (e.g.: 3576.0).
- Water Mains Burst Count By Area, total number of burst found on each DMA
  - DMA, text value, the area of measurement (e.g.: 607D03).

- BurstCount, number of bursts found in a specific DMA zone in the period of time (e.g.: 23).

NGSI-LD Payload

The use of the corresponding Agent (f4w-challenge-sww) facilitates communication with the FIWARE Context Broker through the creation of the corresponding publication NGSI-LD requests. The payload of those requests are represented in JSON-LD format, as can be seen in the following example.

```
{
   “Id”: “urn:ngsi-ld:WaterQualityObserved:<DMA>”,
   “type”: “WaterQualityObserved”,
   “dma”: {
      “type”: “Property”,
      “value”: <DMA>
   },
   “litres”: {
      “type”: “Property”,
      “value”: <Litres>,
      “observedAt”: <DTT>,
      “unitCode”: “LTR”,
      “period”: {
         “type”: “Property”,
         “value”: 900,
         “unitCode”: “SEC”
      }
   }
}
```

Table 5: JSON-LD data for DMA data

Data Explorer

11MB (zip format)

Data files: Link

3.3.2 Validation

Outcomes validation:

A successful solution will:

- Be efficient, responsive and scalable; SWW have over 1000 DMAs and need to respond as quickly as possible to burst events. Any solution might be run multiple times during the same day.
- Not classify very short flow changes (15-30 minutes) as anomalies. These are likely to be data quality issues
- Have a user-friendly interface which:
    1. Alerts users of short-term events (hours to days), providing the time of the event and the DMA. SWW would respond by issuing pipe repair work orders to the problem area.
    2. Suggests DMAs where there has been a significant change in longer term trends (weeks to months). SWW would respond by investigating the DMA using leakage detection methods.

Predictive model needs and priorities:

- The top priority of the model is to detect if a leak occurs.
- False positives and false negatives should be minimized.
- The model needs to detect in real time the leaks.
- The leak should be predicted as soon as possible.
- The leak needs to be diagnosed, classifying if it is a burst pipe or a long-term leak.

Offline validation: How proposed solutions will be validated

To validate the different predictive models developed during the challenge, the participants need to deliver them:

1. The solution must be delivered as a docker image.

2. The docker image must accept by parameter a csv file, which in this case is going to be the validation set.
3. Participants must write a csv file with two columns: “Register Index”, “Model Prediction”. “Register Index” is the register number predicted from the validation set received, and “Model Prediction” is the output of the trained machine learning model.

The delivered models should predict taking into account all possible errors and be able to generalize (including not having overfitting). A set of metrics will be used to evaluate the Machine Learning models.

In case of classification predictions, the metrics will be:

- Accuracy: The proportion of correct predictions among the total number of observations.
- Precision: The fraction of positive predictions correctly predicted.
- Recall: The fraction of the positive observations successfully predicted.
- F-score: A weighted average of the precision and the recall.

Finally, if the problem requires anomaly detection, not only classification metrics will be used, but also specific AD metrics. The objective is to evaluate a set of conditions:

- The anomaly must be predicted as soon as possible.
- False positives should strictly be penalized.
- Predicting the same anomaly multiple times is not important.

Predictive model evaluation method:

- As a final step, to identify the model generalization capabilities, an evaluation set (not included in the database training data) will be used.
- The metrics used to evaluate the model are associated with the confusion matrix and the priorities already introduced.
- The predictions will be evaluated in two steps. First, evaluate if the model is capable of predicting the leaks correctly, and second, evaluate the type of leak predicted.

- Early predictions of overflow will also be considered during the evaluation.

To offline validate the results of the different solutions, participants need to deliver a docker image in a tar.gz.

The docker container must be executed using the command: docker run <image_name> <path_to_evaluation_set_file>

The evaluation set file will have the same format as the general dataset given at the beginning of the competition.

The docker execution needs to output a csv file to the same directory where the evaluation set is stored. This csv file must contain the predictions of the model (only when an anomaly is detected) in a specific format:

- Timestamp: A column indicating the timestamp of the prediction. This timestamp must match with the input register of the time series.
- Prediction: The true prediction of the model. If the anomaly to predict has different types, this should be a string specifying the type of anomaly, e.g. 'Type 1'. This column must only contain anomalous detections.
- Format example: **Head**: Timestamp, Prediction **Row**: 2020-02-23, **'pipe burst'**.

Online validation: - Demo day

Short demonstration of the integration of the model in the architecture. For this purpose, you can use the docker-compose file provided to you for testing the model.

3.3.3 Help Kit

Help Kit: it contains the material that is necessary to create your development environment and to validate your trained machine learning models:

- Scripts to inject validation data into a context broker (and more specifically https://github.com/easy-global-market/f4w-challenges-sww/tree/d948df4ece2aaac49a4a3eb136afc08169bbdfd2) also see Data

- Docker compose files for Stellio and Orion-LD
- Scripts to easily interact with a context broker
- Documentation on actions to be performed to deploy and validate a machine learning model packaged into BentoML

4. Deadlines

- **Opening of the submission is** Wednesday 26 May, 2021 at 4:00 PM CEST
- **Submission deadline** between Tuesday, 29 June 2021 at 8:45 AM and Thursday, 1 July 2021 at 6:00 PM (CEST)
- **Preselection process** from Thursday 1 to Thursday 8 July 2021
- **Communication preselected solutions**: Thursday 8 July
- **Elevator Pitches:** Thursday, 15 July 2021 from 9:00 AM to 5:00 PM CEST
- **Consensus meeting**: Friday 16 July, 2021
- **Communication to winners**: Monday 19 July, 2021
- **31 August 2021**: Official announcement of winners during the https://www.aqua360.net/)

The Fiware4Water Challenges may also include virtual training sessions on Fiware4Water technology.

5. Eligibility Criteria

All applicants will have to abide by all general requirements described in this section to be considered eligible for Fiware4Water Challenges. Therefore, please read this section carefully.

5.1 Who are we looking for?

Participants of this Fiware4Water Challenges must be 18 years old or have reached the age of emancipation in the jurisdiction in which they are submitting the challenges. We are looking for individuals, teams or SMEs registered in any of the following countries:

- the Member States of the European Union and its Overseas Countries and Territories (OCT) or
- Associated Countries to H2020
- The United Kingdom

Participants in the DW2020 projects and other related EU projects.

All members of your team must meet the eligibility criteria described in this document. If any member of your team is ineligible or otherwise fails to comply with this Participation Agreement, the team as a whole may be disqualified in Offline or Online validation process. Each team is solely responsible for its own cooperation and teamwork. Fiware4Water will not officiate any dispute between members regarding cooperation, participation, conduct, prize sharing or intellectual property ownership.

Select one member of your team to enter the information into the registration form. The team leader who submits the solution will receive an email confirming the submission details. While one team member submits the form, all team members are required to accept the Participation Agreement as part of the registration process.

At **registration** time, the team information shall be provided through the registration system.

The registration to the Fiware4Water Challenges entails the acceptance of these guidelines and the legal notice. Participants are solely responsible for the information provided and undertake to respond to any requests for information from promoters.

5.2 How to apply

Fiware4Water Challenges registration will be open from Wednesday, 26 May 2021 at 4:00 PM (CEST) to Thursday 1 July 2021 at 6:00 PM (CEST). All proposal developed as a part of the Fiware4Water Challenges must be “fresh”, meaning that the portion that is included in the main GitHub open repository has been developed during the time period for this

Contest (Wednesday, 26 May 2021 at 4:00 PM (CEST) to Thursday 1 July 2021 at 6:00 PM (CEST)) and that all existing dependent libraries are equally available to all participants.

We will evaluate **only** proposals submitted through the online application within the submission period from **Tuesday, 29 June 2021 at 08:45 AM and Thursday, 1 July 2021 at 6:00 PM (CEST)**. Upon receipt of each proposal, we will send you a confirmation of your submission.

https://fiware4water-challenges.devpost.com

5.3 Other Eligibility Criteria

When applying to Fiware4Water challenges, please also note that:

- You have to verify the completeness of the registration form, as it will not be possible to add any further information after the deadline. After the proposal is submitted, you will be able to modify the registration form data until the deadline by contacting dpo@oieau.fr
- You can submit only one proposal to solve a specific challenge. If more than one proposal is identified, only the last proposal which has been submitted in order of time will be evaluated.
- Your proposal must be written in English in all mandatory parts in order to be eligible. Only parts written in English will be evaluated.
- All mandatory sections of your proposal must be completed. The data provided should be actual, true, complete and should allow assessment of the proposal. Additional material, not specifically requested in the online application form, will not be considered for the evaluation. We will check all the information provided in your solution during the evaluation phase.
- Your project should be based on your original work or your right to use the IPR from third parties must be clear. Going forward, any foreseen developments must be free from third party IPRs, or those third-party IPRs must be clearly stated. When using open source components please identify clearly what license they have.

6. Experimental data access rights

No GDPR protected data will be used across the challenges. The data supplied for the challenges are subject to copyright. Participants accept the responsibility of their use only for the purpose of the challenge. They commit to remove it, once the challenge has been completed. Being accepted as a participant entitles you to access the data according to the specific conditions. By accepting your participation you are accepting these conditions to access the data.

You agree to use suitable and appropriate measures to prevent persons who have not formally agreed to the rules of the Fiware4Water Challenge from gaining access to the Data Challenges. You agree not to transmit, duplicate, publish, redistribute, or make available this data to any party that is not participating in the Fiware4Water Challenge. You will be responsible for the same should they occur. if there is an unauthorized transmission or access to the data, you agree to notify us immediately upon learning and work to resolve this action.

7. How will we evaluate your proposal?

Our evaluation process is transparent, fair and equal to all participants. We will evaluate your project in two phases, first for the work offline and second for the online session.

Please pay attention to provide accurate information about your work in order to make a fair assessment of your project easier. Goals for the challenges are explained in section 3. Please answer clearly to the challenges identified and explain the method the team has used for the solution.

7.1 Experts Evaluation

In this phase, all eligible proposals will be evaluated by experts in Fiware4Water technologies, Data Models, and Challenges owners. Selected solutions will be appointed to run a demo video conference meeting Thursday, 15 July 2021. Your proposal will be evaluated within the following awarding criteria.

Off-line evaluation (70%)

Each challenge has a set of rules and metrics that are going to be used to score the machine learning models. The best machine learning model delivered will receive the 100% score, while others will have a scoring relative to the performance in comparison with the best model.

On-line evaluation (30%)

This will be assessed on the selection Day (see next section). Only those higher scoring projects will pass to the selection day.

7.2 Selection Day (Demo Day)

If your project is among the finalists, you will be invited to an online Selection Day 15th of July, where you will have the opportunity to pitch your project (demo to show the execution) in front of the Fiware4Water Challenge's Jury. After the Selection Day, we will communicate the winners of the Fiware4Water Challenges.

Public Presentation of Proposals

Each team will present its proposal with a pitch of 20 minutes in English in the 'Virtual Pitch' session plus 10 minutes for jury questions. The virtual presentation (with the subsequent evaluation) will take place on Thursday, 15 July 2021 from 9:00 to 17:00 CEST. The virtual session details will be communicated to participants in the Slack channel defined in the section 15.1.2. The Jury referred to in art. 7.3 will evaluate the proposals according to the info provided for in section 7.1.

7.3 Composition of the Jury

The final composition of the Jury will be made public by Friday, 2 July 2021 as may be modified, due to the unavailability of one or more members at any time by the promoters.

The expected members of the Jury are:

- At least a representative of FIWARE Foundation, EGM, Eurecat
- At least a representative of the entities providing with the Challenges

- Another independent member expert on water domain

The Jury will meet at the end of the presentations on Thursday, 15 July 2021 and will draw up a ranking of the projects presented for each challenge. This date can be changed depending on the number of proposals submitted to adequate the proper On-line validation during demo day. There will be 3 winning individuals/teams, one for each challenge. Nevertheless, the jury may reject any Submission that does not satisfy the objectives. The jury is sovereign and its judgment is final. It follows that no complaint can be accepted following the nomination of the winners.

7.4 What's Next? Call Agreement preparation and signature

Before the selection starts, we will ask all members participating in a team to approve their participation in the challenge.

Please do it within the deadlines that will be communicated to you. If you fail to deliver the requested documents on time, without clear and reasonable justification, we will exclude you from the further formal assessment and you will be replaced.

8. Prizes arrangement

Three winning solutions will be selected. One per each challenge. The prizes include:

1. FIWARE Foundation and EGM intend to create a consortium to submit a proposal for the IBM 2021 Call for Code - Clean Water and Sanitation. One representative (1) for each of the winning solutions will be invited to join the consortium. However, if the consortium agrees to not submit the proposal or the proposal does not get selected, FIWARE Foundation and EGM are not liable for any consequence.
2. Starting from 01.08.2021, one (1) year of free social media promotion and global visibility thanks to being featured throughout international cross-media channels of FIWARE Foundation.

3. Starting from 01.08.2021, automatic subscription to the free technical webinars provided by FIWARE Foundation. Winners can unsubscribe to the event list at any time.
4. Joining the FIWARE Marketplace to provide visibility and commercial opportunities based on the awarded solution.
5. More prizes are coming soon.

9. Intellectual property

The solutions delivered by the teams will be open source. A solution can be freely used, modified and distributed without worrying about the use of software: personal, internal or commercial. Nevertheless, the provided data during the competition is private, and belongs to each challenge promoter. Accepted licenses include GPLv2, GPLv3, EUPL, AfferoGPL, MIT, BSD, etc as examples but other licenses compliant with previous requirements will be accepted as well.

10. Personal data and Image Rights

Promoters and their members are authorised to disseminate the names of the winners, the name of the project, its purpose and description, together with extracts or images of the final product.

Each team member agrees that the promoters and their members will photograph or film him/her and disseminate to the public images that may include his/her person.

11. Right of Disqualification and Exclusion

The promoters reserve the right to disqualify anyone who does not comply with the rules, materials made available to him/her, and the premises of the Challenges.

12. Cancellation or Change of Dates

In the event of extraordinary events or other circumstances that make it necessary, the promoters may cancel the Challenge or move it to another date.

13. Other Provisions

The participant acknowledges that he/she has been informed that he/she will be held solely responsible for any inaccuracies contained in the registration format he/she has completed or for any failure to comply with the obligations contained in the regulations.

14. Applicable Law

The applicable law is the German law. In the event of any dispute concerning the interpretation or enforcement of the Regulation, the case shall be brought before the competent court in Berlin.

Promoters and partners are not responsible for damage, loss or theft of materials and equipment used by participants for the performance of their work within the Challenges.

15. Contact us

Where to find additional information:

Fiware4Water website: https://www.fiware4water.eu/smes-challenges

Hackathon platform: https://fiware4water-challenges.devpost.com

15.1 How can we help you?

15.1.1 Questions regarding the Challenges process

You can post your management questions at Helpdesk
challenges.support@fiware4water.eu

In case of any technical issues or problems, please use the corresponding Slack channel defined in the section 15.1.2

15.1.2 Technical questions

A Slack workspace has been created to allow discussing technical details with the Fiware4Water technical team in a reactive and interactive way.

To join the Slack workspace, you can use the following link: https://join.slack.com/t/f4w-challenges/shared_invite/zt-opnbyzg7-TrdHwkHizB1384euyZWJOg

Once you have joined, you can ask your questions in one of the following way:

- In the #general channel if it is a general question about the technical environment, the FIWARE ecosystem, …
- In one of the channels specifically aimed at discussing about a challenges (#sofia-challenge, #milan-challenge, #sww-challenge)
- Or you can create a private channel and invite the Fiware4Water technical team in it if you don't want to disclose your discussions

As a side note, when you want to raise a technical issue, please include the following information in your message:

- details of the specific problem (error messages you encountered, bugs descriptions, i.e. if a dropdown list isn't working, etc.)
- screenshots of the problem if it can help.
- Any other relevant information that you consider appropriate to explain the issue.