

SEBASTIEN

M7 – Report on released database of sectoral indicators/indices

Milestone Lead	CMCC
Milestone due date	2024/07/31
Status	FINAL
Version	V1.0
Project	SEBASTIEN



DOCUMENT INFORMATION

Title	Milestone 7 - Report on released database of sectoral indicators/indices
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	CMCC
Milestone Description	Report on released database of sectoral indicators/indices
Means of verification	Report on released database of sectoral indicators/indices is shared with the agency
Contributors	Alessandro D'Anca (CMCC), Valentina Scardigno (CMCC), Paola Nassisi (CMCC)
Requested deadline	M31
Reviewer	Marco Milanese (UNITUS)

Table of content

1. Introduction.....	4
2. Sebastien datasets database	5
3. Indices/Indicator produced by Sebastien	7
3.1. Production indices.....	7
3.2. THI - Temperature Humidity Index	9
3.3. Pasture	11
3.4. SCC – Somatic Cell Count variation	14
3.5. Blue Tongue	16
3.6. IoT - Animal sensor data	18
4. Conclusions.....	19

1. Introduction

This document describes the database produced within the SEBASTIEN project in order to present the sectoral indicators produced.

These indicators are produced on a daily or one-time basis from different data sources.

In the case of indicators produced on a one-time basis, input data were collected and adapted in order to feed the different production pipelines based on ML procedures (as reported in Milestone 4). The data produced were then organised in the datalake and exposed via appropriate APIs.

The indicators that are updated/verified daily, are based on input data that are updated/verified once a day, in particular:

- short-term weather forecasts available from the Mistral portal,
- Sentinel data available on the Copernicus portal that are verified daily and updated at varying intervals.

At each update, the corresponding pipeline (again based on ML procedures) is activated producing the corresponding indicators. These data are also stored and organised in the data lake and exposed through appropriate APIs in order to feed the Sebastien service portal.

The indicators produced are collected within the datalake using different formats depending on their use: netcdf file, basic forma for the dataset produced, and png image or geo-json description to ensure visualisation on the Sebastien service portal.

2. Sebastien datasets database

The set of data produced by the project is shown in the following table.

Indices	Set	Estimation	Time period	Related service	Metadata url	Output format	Update frequency
Percentage variation of milk yield	Production indices	Indoor/Outdoor	Short term	1 - Estimation of production decline based on climate variables	https://sebastien-datalake.cmc.c.it/api/v2/datasets/pi	Netcdf, GEO-json file	Daily
Percentage variation of protein content							
Percentage variation of fat content							
Percentage variation of milk yield	Production indices	Indoor/Outdoor	Long term	1 - Estimation of production decline based on climate variables	https://sebastien-datalake.cmc.c.it/api/v2/datasets/pi-long-term	Netcdf, png image	None
Percentage variation of protein content							
Percentage variation of fat content							

Temperature-Humidity index	THI	Indoor	Short term	2 - Estimation of the environmental conditions of the stables	https://sebastien-datalake.cmc.c.it/api/v2/datasets/thi	Netcdf, GEO-json file	Daily	
		Indoor	Long term	2 - Projection of the environmental conditions of the stables in the long term		Netcdf, png image	None	
Pasture data	Pasture	Outdoor	Short term	3 - Pasture monitoring	https://sebastien-datalake.cmc.c.it/api/v2/datasets/pasture	Netcdf, GEO-json file	~10 days	
Somatic cell count variation	SCC	Indoor/Outdoor	Short term	4 - Maps on the risk of infection with long-term communicable diseases (2050)	https://sebastien-datalake.cmc.c.it/api/v2/datasets/pi	Netcdf, GEO-json file	Daily	
			Long term			https://sebastien-datalake.cmc.c.it/api/v2/datasets/pi-long-term	Netcdf, png image	None
			Long term			https://sebastien-datalake.cmc.c.it/api/v2/datasets/blue-tongue	Netcdf, png image	None
Probability of developing blue-tongue disease	Blue tongue	Outdoor	Long term					
IoT animal sensor data	IoT		Short term	My Sebastien - Sensors	https://sebastien-datalake.cmc.c.it/api/v2/datasetd/iot-animal	Netcdf, GEO-json file	Hourly	

In the remainder of the document, the different datasets will be detailed following the subdivision by Set to which they belong.

3. Indices/Indicator produced by Sebastien

3.1. Production indices

This category includes indices useful to study how climate change will affect the production capacity of farms, by evaluating its effects on the quantitative and qualitative characteristics (percentage of fat and protein) of the milk.

In particular, they concern three aspects of milk production:

1. Quantity (loss expressed in Kg);
2. Protein (loss expressed in percentage);
3. Fat (loss expressed in percent);

These indices are calculated on the short term and the long-term period. In particular, the short-term evaluations concern forecasts up to 2 days of production loss. These forecasts are updated daily when new weather forecasts up to 2 days are available.

The long-term evaluations represent the median of the differences between the predicted phenotype values for the future (from 2021 to 2050 considering the RCP 8.5 - high emissions - scenario) compared to the actual values evaluated for the past (from 1989 to 2018). The statistics are calculated for the summer months: June, July, and August, chosen because production suffers the greatest loss in these months due to changing climatic conditions. These forecasts are generated on a one-time basis and published on Sebastien's portal.

In addition, these indices are calculated both inside the stables and outside from the respective expected climatic conditions (short and long term).

For each of the 3 indices, therefore, we have a total of 4 possible combinations for which it is assessed:

1. Short term + indoor
2. Short term + outdoor
3. Long term + indoor
4. Long term + outdoor

for a total of 12 indices.

Figure 1 shows the possibility of selecting a production-related indicator via the service portal.

PHENOTYPE

- ✓ Milk Yield
- Fat Percentage In Milk
- Proteins Percentage In Milk

Outdoor ▼

PERIOD

Short-term (2 Days) ▼

CALCULATE

Figure 1. Dropdown menus to select the Phenotype, Farming mode (Indoor/Outdoor) and Period (short/long term)

The following code (in json format), on the other hand, represents an extract of the metadata associated with the “Production indices” set as obtained from the url: <https://sebastien-datalake.cmcc.it/api/v2/datasets/pi>

```

{
  "metadata": {
    "role": "public",
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "Forecast Production Indices",
  "id": "outdoors",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "Production Indices derived from MISTRAL COSMO-2I data",
      "contact": {
        "name": "Data Deliver System Support Team",
        "email": "dds-support@cmcc.it",
        "webpage": "https://www.cmcc.it/research-organization/research-
divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
      },
      "label": "Production Indices over Italy",
      "image": null,
      "doi": null,
      "update_frequency": null,
      "license": null,
      "publication_date": "2023-09-15",
      "related_data": null,
      "id": "pi"
    },
    [...]
  },
  "fields": {
    "fat": {
      "units": "unknown",
      "description": "Fat"
    }
  }
}

```



```

},
"milk": {
"units": "unknown",
"description": "Milk"
},
"proteins": {
"units": "unknown",
"description": "Proteins"
},
"scc": {
"units": "unknown",
"description": "SCC"
}
}
}

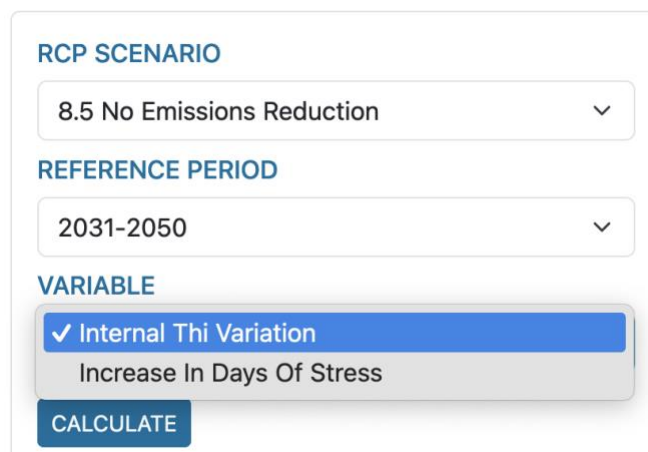
```

3.2. THI - Temperature Humidity Index

This category includes indices related to the estimation of the temperature-humidity index (THI) inside stables and outside based on external environmental conditions. These indices are evaluated short and long term. Specifically, in the short-term period, the THI is assessed for the next 2 days. These forecasts are updated daily when new weather forecasts are available.

The long-term forecasts, on the other hand, are calculated considering 2 different scenarios: RCP 4.5 (intermediate emissions) and RCP 8.5 (high emissions). For each of these scenarios, 4 different forecast periods can be considered: 2021-2040, 2031-2050, 2041-2060 and 2051-2070. For each of the previous combinations, it is possible to calculate the change in the internal THI of the stables and the increase in the number of stress days compared to the 1981-2000 reference period. In total, therefore, there are 16 different indicators as the scenario, the forecast period and the type of index change.

Figure 2 shows a possible indicator selection for a given period and RCP scenario.



The image shows a user interface for selecting parameters. It features three dropdown menus: 'RCP SCENARIO' with '8.5 No Emissions Reduction' selected, 'REFERENCE PERIOD' with '2031-2050' selected, and 'VARIABLE' with 'Internal Thi Variation' selected. Below the variable menu, 'Increase In Days Of Stress' is also visible. A blue 'CALCULATE' button is positioned at the bottom of the selection area.

Figure 2. Dropdown menus to select the Indicator, Reference Period and RCP scenario.

The following code (in Json format), on the other hand, represents an extract of the metadata associated with the 'THI' set as obtained from the url: <https://sebastien-datalake.cmcc.it/api/v2/datasets/thi>

```
{
  "metadata": {
    "role": "public",
    "filters": [
      {
        "name": "date",
        "user_defined": "T",
        "label": "Date"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "Hourly Thermohygrometric Indices",
  "id": "hourly",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "Thermohygrometric Indices derived from MISTRAL COSMO-2I
data",
      "contact": {
        "name": "Data Deliver System Support Team",
        "email": "dds-support@cmcc.it",
        "webpage": "https://www.cmcc.it/research-organization/research-
divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
      },
      "label": "Thermohygrometric Indices over Italy",
      "image": null,
      "doi": null,
      "update_frequency": null,
      "license": null,
      "publication_date": "2023-06-19",
      "related_data": null,
      "id": "thi"
    },
    "products": {
      "hourly": {
        "role": "public",
        "filters": [
          {
            "name": "date",
            "user_defined": "T",
            "label": "Date"
          }
        ]
      },
      "catalog_dir": "/catalog/Catalog/external/",
      "description": "Hourly Thermohygrometric Indices"
```

```

    },
    "projection-ext": {
      "role": "public",
      "catalog_dir": "/catalog/Catalog/external/",
      "description": "Projection External Thermohygrometric Indices"
    },
    "projection-int": {
      "role": "public",
      "catalog_dir": "/catalog/Catalog/external/",
      "description": "Projection Internal Thermohygrometric Indices"
    },
    "forecast": {
      "role": "public",
      "filters": [
        {
          "name": "scenarios",
          "user_defined": "T",
          "label": "Scenarios"
        },
        {
          "name": "period",
          "user_defined": "T",
          "label": "Period"
        }
      ],
      "catalog_dir": "/catalog/Catalog/external/",
      "description": "Thermohygrometric Indices Forecasts"
    }
  },
  [...]

```

3.3. Pasture

This indicator represents the evaluation of fresh and dry biomass of pastures in near real-time, using satellite data. It is calculated periodically based on the availability of satellite observations (updated every 10-15 days).

Figure 3 shows an example of a list of pastures saved by the user via the Sebastien service portal and for which fresh and dry biomass information is available.



Pasture monitoring i





SAVED AREAS		↻	
Pasture 01	✓		
Pasture 02	✓		

Figure 3. Example of list of user's pastures

The following code (in json format) represents an extract of the metadata associated with the "Pasture" set as obtained from the url: <https://sebastien-datalake.cmcc.it/api/v2/datasets/pasture>.

```
{
  "metadata": {
    "role": "public",
    "filters": [
      {
        "name": "pasture",
        "user_defined": "T",
        "label": "Pasture"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "Latest Pasture TQ",
  "id": "tq-latest",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "Pasture TQ and dry substance derived from Remote Sensing Indices ",
      "contact": {
        "name": "Data Deliver System Support Team",
        "email": "dds-support@cmcc.it",
        "webpage": "https://www.cmcc.it/research-organization/research-divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
      },
      "label": "Pasture TQ and dry substance",
      "image": null,
      "doi": null,
      "update_frequency": null,
      "license": null,
      "publication_date": "2023-11-22",
      "related_data": null,
      "id": "pasture"
    }
  }
}
```

```
},
"products": {
  "tq-latest": {
    "role": "public",
    "filters": [
      {
        "name": "pasture",
        "user_defined": "T",
        "label": "Pasture"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Latest Pasture TQ"
  },
  "ss-latest": {
    "role": "public",
    "filters": [
      {
        "name": "pasture",
        "user_defined": "T",
        "label": "Pasture"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Latest Pasture SS"
  },
  "tq-sum": {
    "role": "public",
    "filters": [
      {
        "name": "pasture",
        "user_defined": "T",
        "label": "Pasture"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Aggregate Pasture TQ"
  },
  "ss-sum": {
    "role": "public",
    "filters": [
      {
        "name": "pasture",
        "user_defined": "T",
        "label": "Pasture"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Aggregate Pasture SS"
  }
}
}
```

3.4. SCC – Somatic Cell Count variation

This indicator represents the estimation of the climate impact, both external and internal to the barn, in the short and long term, on somatic cell count, an indirect indicator of udder health. In the same way of the indicators belonging to the Production indices set, it is evaluated both inside the barn and outside, and the forecasts are either two days ahead or long term.

The short-term forecasts are updated daily on the basis of new weather forecasts of up to 2 days. The long-term evaluations represent, as in the case of production indices, the median of the differences between the predicted SCC values for the future (from 2021 to 2050 considering the RCP 8.5 - high emissions - scenario) compared to the actual values evaluated for the past (from 1989 to 2018). The statistics are calculated for the summer months: June, July, and August. These forecasts are generated on a one-time basis and published on Sebastien's portal.

Figure 4 shows the dropdown menus available on the portal to select the farming mode and period.



Figure 4. Dropdown menus to compute the risk of mastitis

The following code (in json format) represents an extract of the metadata associated with the “SCC” set. Having the same structural characteristics as the Production Indices set, it has been grouped in the same metadata set. The metadata set referring to the long-term evaluation is presented here: <https://sebastien-datalake.cmcc.it/api/v2/datasetd/pi-long-term>

```
{
  "metadata": {
    "role": "public",
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "Forecast Daily Production Indices outdoors",
  "id": "outdoors",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "Production Indices derived from MISTRAL COSMO-2I data",
```

```
"contact": {
  "name": "Data Deliver System Support Team",
  "email": "dds-support@cmcc.it",
  "webpage": "https://www.cmcc.it/research-organization/research-
divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
},
"label": "Production Indices over Italy",
"image": null,
"doi": null,
"update_frequency": null,
"license": null,
"publication_date": "2023-09-15",
"related_data": null,
"id": "pi-long-term"
},
"products": {
  "outdoors": {
    "role": "public",
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Forecast Daily Production Indices outdoors"
  },
  "stable": {
    "role": "public",
    "catalog_dir": "/catalog/Catalog/external/",
    "description": "Forecast Daily Production Indices in stable"
  }
}
},
"data": {
  "domain": {
    "crs": {
      "name": "latitude_longitude",
      "semi_major_axis": 6371229,
      "semi_minor_axis": 6371229,
      "inverse_flattening": 0,
      "longitude_of_prime_meridian": 0
    },
    "coordinates": {
      "latitude": {
        "min": 36.336,
        "max": 49.616,
        "units": "unknown",
        "axis": "LATITUDE"
      },
      "longitude": {
        "min": 2.972,
        "max": 20.352,
        "units": "unknown",
        "axis": "LONGITUDE"
      }
    }
  }
},
"fields": {
```

```
"fat": {
  "units": "unknown",
  "description": "Fat"
},
"milk": {
  "units": "unknown",
  "description": "Milk"
},
"proteins": {
  "units": "unknown",
  "description": "Proteins"
},
"scc": {
  "units": "unknown",
  "description": "SCC"
}
}
}
```

3.5. Blue Tongue

The Blue Tongue indicator represents the prediction from July to November of the higher or lower probability of developing Blue Tongue disease in sheep in Sardinia in the future (from 2021 to 2050) compared to the past (from 1989 to 2018). This indicator, therefore, was developed only for long-term predictions and is calculated on a one-time basis. The related data were stored in the datalake.

The following code (in json format) represents an extract of the metadata associated with the 'Blue Tongue' set as obtained from the url: <https://sebastien-datalake.cmcc.it/api/v2/datasets/blue-tongue>

```
{
  "metadata": {
    "role": "public",
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "Blue Tongue monthly data from July to November",
  "id": "monthly",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "Blue Tongue monthly data from July to November",
      "contact": {
        "name": "Data Deliver System Support Team",
        "email": "dds-support@cmcc.it",
        "webpage": "https://www.cmcc.it/research-organization/research-divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
      },
      "label": "Blue Tongue monthly data",
      "image": null,
      "doi": null,
      "update_frequency": null,

```



```
"license": null,
"publication_date": "2023-09-15",
"related_data": null,
"id": "blue-tongue"
},
"products": {
"monthly": {
"role": "public",
"catalog_dir": "/catalog/Catalog/external/",
"description": "Blue Tongue monthly data from July to November"
}
}
},
"data": {
"domain": {
"crs": {
"name": "latitude_longitude",
"semi_major_axis": 6371229,
"semi_minor_axis": 6371229,
"inverse_flattening": 0,
"longitude_of_prime_meridian": 0
},
"coordinates": {
"latitude": {
"min": 38.756,
"max": 41.356,
"units": "unknown",
"axis": "LATITUDE"
},
"longitude": {
"min": 8.032,
"max": 9.892,
"units": "unknown",
"axis": "LONGITUDE"
},
"time": {
"min": "2023-07-31T00:00:00.000000000",
"max": "2023-11-30T00:00:00.000000000",
"units": "days since 2023-07-31",
"axis": "TIME",
"time_unit": "day",
"time_step": 30
}
}
},
"fields": {
"blue_tongue": {
"units": "unknown",
"description": "Blue Tongue"
}
}
}
}
```

3.6. IoT - Animal sensor data

In this category are a series of indicators related to animal health, for instance: environmental temperature and relative humidity, GNSS position (lat/lon), heart rate, acceleration, etc. This information is collected individually for each animal in real time and stored in the Sebastien datalake.

Figure 5 shows an example of information related to the status of an animal and displayed via the relevant section within the Sebastien service portal.

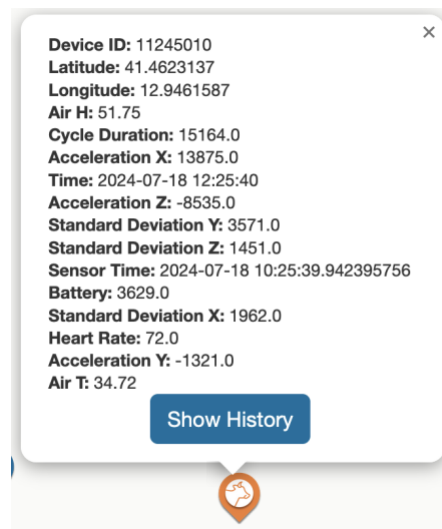


Figure 5. Information from sensors (IoT) concerning the current state of an animal.

The following code (in json format) represents an excerpt of the metadata associated with the “IoT” set as retrieved from the url: <https://sebastien-datalake.cmcc.it/api/v2/datasets/iot-animal>

```
{
  "metadata": {
    "role": "public",
    "filters": [
      {
        "name": "user",
        "user_defined": "T",
        "label": "User"
      }
    ],
    "catalog_dir": "/catalog/Catalog/external/"
  },
  "description": "IoT Animal data for CMCC SEBASTIEN project",
  "id": "latest",
  "dataset": {
    "metadata": {
      "catalog_dir": "/catalog/Catalog/",
      "description": "IoT Animal data for CMCC SEBASTIEN project",
      "contact": {
        "name": "Data Deliver System Support Team",
        "email": "dds-support@cmcc.it",
```

```

    "webpage": "https://www.cmcc.it/research-organization/research-
divisions/advanced-scientific-computing-division#1553329820238-2055494b-9aa6"
  },
  "label": "IoT Animal data for CMCC SEBASTIEN project",
  "image": null,
  "doi": null,
  "update_frequency": null,
  "license": null,
  "publication_date": "2023-06-19",
  "related_data": null,
  "id": "iot-animal"
},
[...]
```

```

  "fields": {
    "AirH": {
      "units": "unknown",
      "description": null
    },
    "AirT": {
      "units": "unknown",
      "description": null
    },
    "HeartRate": {
      "units": "unknown",
      "description": null
    },
    [...]
  }
}

```

4. Conclusions

The database produced in the context of the SEBASTIEN project has the aim to collect the sectoral indicators produced in order to organise them in terms of functionality and way to access and final use (e.g., by the SEBASTIEN service portal).

Many indicators have been produced to explain different aspects of the livestock sector and its relationship with weather and climate conditions. Each of them is either updated daily, following the most up-to-date weather forecasts or newly available satellite data, or generated once on the basis of climate future scenarios (from 2030 to 2070). The SEBASTIEN service portal represents a useful tool for selecting and visualizing them in an intuitive and user-friendly way.