http://www.everest-h2020.eu

dEsign enVironmEnt foR Extreme-Scale big data analyTics on heterogeneous platforms

D1.2 – Initial Data Management Plan

The EVEREST project has received funding from the European Union's Horizon 2020 Research & Innovation programme under grant agreement No 957269

Project Summary Information

Project Title	dEsign enVironmEnt foR Extreme-Scale big data analyTics on heterogeneous platforms
Project Acronym	EVEREST
Project No.	957269
Start Date	01/10/2020
Project Duration	36 months
Project website	http://www.everest-h2020.eu

Copyright

Num.	Partner Name	Short Name	Country
1 (Coord.)	IBM RESEARCH GMBH	IBM	CH
2	POLITECNICO DI MILANO	PDM	IT
3	UNIVERSITÀ DELLA SVIZZERA ITALIANA	USI	CH
4	TECHNISCHE UNIVERSITAET DRESDEN	TUD	DE
5	Centro Internazionale in Monitoraggio Ambientale - Fondazione CIMA	CIMA	IT
6	IT4Innovations, VSB – Technical University of Ostrava	IT4I	CZ
7	VIRTUAL OPEN SYSTEMS SAS	VOS	FR
8	DUFERCO ENERGIA SPA	DUF	IT
9	NUMTECH	NUM	FR
10	SYGIC AS	SYG	SK

Project Coordinator: Christoph Hagleitner – IBM Research – Zurich Research Laboratory

Scientific Coordinator: Christian Pilato – Politecnico di Milano

Disclaimer

Deliverable Information

Work-package	WP1
Deliverable No.	D1.2
Deliverable Title	Initial Data Management Plan
Lead Beneficiary	NUM
Type of Deliverable	ORDP: Open Data Research Plan
Dissemination Level	Public
Due date	01/04/2021

Document Information

Delivery date	20/04/2021
No. pages	21
Version | Status	0.4 | Final
Responsible Person	Fabien Brocheton (NUM)
Authors	Fabien Brocheton (NUM), Antonella Galizia (CIMA), Katerina Slaninova (IT4I), Radim Cmar (SYG), Guido Rusca (DUF), Riccardo Cevasco (DUF)
Internal Reviewer	Antonio Parodi (CIMA)

The list of authors reflects the major contributors to the activity described in the document. All EVEREST partners have agreed to the full publication of this document. The list of authors does not imply any claim of ownership on the Intellectual Properties described in this document.

Revision History

Date	Ver.	Author(s)	Summary of main changes
01/12/2020	0.1	Fabien Brocheton (NUM), Antonella Galizia (CIMA), Katerina Slaninova (IT4I), Radim Cmar (SYG) Guido Rusca (DUF)	Initial Draft
22/03/2021	0.2	Christian Pilato (PDM)	Added some comments and revisions
31/03/2021	0.3	Fabien Brocheton (NUM)	Added text on Open Data Management plan and FAIR principles

Quality Control

Approved by internal reviewer	15/04/2021
Approved by WP leader	20/04/2021
Approved by Scientific Coordinator	20/04/2021

Table of Contents

1 Executive summary

The document describes the current status of the Data management plan (DMP) at M6, specially in regard to research data available outside the project.
Indeed, EVEREST project has chosen to be part of the Open Research Data (ORD) pilot of the H2020 program[1]. The ORD Pilot aims to improve and maximize access and re-use of research data generated by Horizon 2020 projects and considers the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. The ORD pilot applies:

- Primarily to the data needed to validate the results presented in scientific publications.
- Other data can also be provided by the beneficiaries on a voluntary basis

There are two main pillars in the Pilot:

- Develop (and keep up to date) a Data Management Plan (DMP)
- Provide open access to research data (i.e., *implement* the DMP):
  - Deposit our data in a "research data repository".
  - Ensure third parties can freely access, mine, exploit, reproduce, and disseminate our data.
  - Provide related information and identify (or provide) the tools needed to use the raw data to validate our research.

Allowing data to be Findable, Accessible, Interoperable and Reusable corresponds to the **FAIR data concept** requested by the ORD pilot.
This document describes the current decisions and the plans for the next months, and the partners plan to update it regularly during the project when new data will come in.

1.1 Structure of the document

The document is organized as follows:

- Sections 2 to 4 describe the research data managed in the three pilot use cases that we plan to make available for research outside the project.
- Section 5 presents some general rules on data management policies.

1.2 Related document

D2.1 – More details on the EVEREST use cases
D2.3 – More details on the data managed inside the project.

[1] All details are available on https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2 AIR QUALITY USE CASE DATA MANAGEMENT PLAN

In this section, the air quality pilot data management plan will be described.

2.1 Data Summary

The context of the use case is the forecast of air pollution peak (or no peak) due to the emission of an industrial site in order for this site to manage its impact and prevent such pollution event on the population. The objective is to improve the quality of the forecast by avoiding false pollution peaks (financial loss for the site) or miss pollution peaks (health impact on population).
This use case relies on three main steps for each air quality forecast simulation:

- Step 1: Compute deterministic (short range) and/or probabilistic (nowcasting) meteorological forecast with the WRF model. During each simulation, an assimilation procedure in order to force computation by observations is activated. This part is executed typically on HPC server or can be executed also on FPGA-accelerated platform for selected kernels.
- Step 2: Combine this deterministic (short range) and/or probabilistic (nowcasting) meteorological with another weather forecasts and local measurement in order to obtain, by machine learning approach, a better weather forecast. This part is executed on a cloud server.
- Step 3: Compute deterministic air quality forecast with the ADMS model, based either on the weather forecast from the step 1 or the step 2. This part is executed on Windows cloud server.

Step 1 requires the following data sets:

- Global forecast dataset (GFS) produced by NCEP (US National Center for Environmental Prediction) or Integrated Forecasting System (IFS) produced by ECMWF (European Centre for Medium-Range Weather Forecasts)
- For the assimilation and validation procedures:
    - Surface weather observation data (hydrometeorological variables, e.g. temperature, water vapour, wind speed and direction) provided by authoritative (e.g. ECOMET) and personal weather (e.g. underground) stations network.
    - Other hydrometeorological variables (e.g. reflectivity, soil moisture, land surface temperature, sea surface temperature, integrated water vapor content) retrieved from ground-based (radar) and spaceborne remote sensing system (Sentinel data from the European Copernicus service).

The outputs of the step 1 are meteorological parameters for the simulated domain on a 3D grid.

Step 2 requires also to download external input data:

- Another dataset of weather forecast: it will be forecasts produced by NUMTECH at different scales (Europe/France), extension of the GFS

dataset used at step 1 (use of another cycle forecast), eventually forecast from Meteo France.
- For the learning phase, local weather surface observation at industrial site.

The outputs of the step 2 are meteorological parameters at the location of the industrial site (where local observation data is provided).
For the step 3, the outputs of steps 1 and 2 are intermediate data used as input data. Another input data are emission data for the forecast period. The outputs are 2D (at surface) grid of pollutants concentrations.

Most of the datasets describe above are open-source data available freely, except:
- **NUM** forecast data which are commercial ones, used freely in the framework of this research project, but not possible to share publicly.
- Local weather observation data at industrial site which are private data (to keep confidential inside the EVEREST project).
- Some third-party observation data (radar data or personal surface station) which are commercial ones. In the framework of the EVEREST project, these datasets could be used freely, but not possible to share publicly.

Most of the output data relevant to the air quality pilot are proprietary model output. Categories of the research outputs produced by the use case are listed as follows:
- The WRF model output from step 1 described above.
- The ADMS model output from step 3 described above.

2.2 FAIR data

Making data findable, including provisions for metadata
The data will be stored on data repositories with digital object identifiers. We will choose in priority public repositories as long as they allow us to comply with the constraints on the datasets access. Datasets will all have a metadata description, and, in the case of datasets with access restrictions, their metadata will be publicly available. A semantic versioning scheme will be used to track versions of the datasets. The partner responsible for generating that data will be the point of contact for requesting an access to the data.

Making data openly accessible
For the air-quality use case, there will be no restrictions on the use of dataset for only research activity in any domain. No control on the use of the dataset will be done, except at the downloading step where applicants in order to use dataset must declare for which research activity they want to do it, and accept to quote EVEREST in acknowledgements in case of publication.
In order to make data interoperable and increase the data re-use, the research data will be provided on an open scientific format largely used in the meteorology and air-quality applications, and not in a proprietary model format. The format

is NetCDF (Network Common Data Form) devoted for storing multidimensional scientific data including inside its metadata description. Open-source tools are available in order to read easily this format.

2.3 Plan of the outputs

The research outputs produced by the use case are:

- The WRF model output: External researchers will have access to a new source of local weather forecast over France for their own applications in order to evaluate the impact of assimilation and procedures applied in the framework of EVEREST, without any limitation on the application domain except the list of weather parameters provided in the dataset. The ADMS model output: External researchers will have access of real temporal and spatial variations of atmospheric impacts of some industrial sites which can be the input of another applications as for example health impact assessment protocols or design measuring devices.

For each dataset, a specific table provides additional information.

Table 1 – Research dataset for AIR QUALITY use case.

ID	ITEM	DESCRIPTION
D1	Dataset name and reference	WRF output (reference will be updated)
	Dataset description	2D and 3D meteorological fields, over different locations (France and Italy) at hourly temporal resolution, for a selected period.
	Standards, format and metadata	NetCDF format
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	During the project: Outputs for the selected periods will be stored for the duration of the EVEREST project, with a focus on parameters used for air quality application. After the project: Outputs will be stored on the CIMA storage system.

		Sample data may be archived for future research (e.g., Zenodo like repository, and Earth System Science Data journal - www.earth-system-science-data.net).
D2	Dataset name and reference	ADMS output (reference will be updated)
	Dataset description	2D air-quality concentrations over different locations (France and Italy) at hourly temporal resolution, for a selected period.
	Standards, format and metadata	NetCDF format
	Is dataset confidential? Must be encrypted?	Spatial coordinates will be "anonymized" (relative coordinates) in order to not have a link to an industrial site
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	During the project: Outputs for the selected periods will be stored for the duration of the EVEREST project, with a limited set of pollutants. After the project: Outputs will be stored on the NUMTECH storage system. Sample data may be archived for future research.

3 RENEWABLE ENERGY PRODUCTION USE CASE DATA MANAGEMENT PLAN

This section describes the DMP for the renewable energy production pilot.

3.1 Data Summary

The context of the use case is the prediction of the renewable energy produced by a wind farm reducing the risks related to severe meteorological ramp-up/down events. The objective is to improve the quality of the forecast by reducing the related uncertainty minimizing the prediction error for renewable energy trading avoiding false production peaks (financial loss) or miss production peaks (wastefulness of natural resources).
The following table shows the wind farms dispatched by **DUF** that will be considered in this study:

Wind Farm code	Electricity Zone	Power capacity MW	Yearly production [MWh]
UP_SARD	SARD	29,75	55.000
UP_SICI	SICI	33,15	66.000
UP_CSUD	CSUD	17,85	26.600
UP_SUD	SUD	18,00	47.827
UP_CALA	CALA	34,00	80.000

This use case relies on three main steps for each energy production prediction simulation:

- Step 1: Compute deterministic (short range) and/or probabilistic (nowcasting) with the WRF model. Data assimilation procedures are applied to force computation through atmosphere observations in order to improve weather prediction. Due to the high computational and memory requirements, HPC resources are mandatory to run the simulations. Some execution parts can be executed also on EVEREST FPGA-based systems.
- Step 2: Concerns with deterministic prediction of hourly energy generation, in the site-specific meteorological conditions forecasted by WRF model. The generation is updated in real-time and has a timescale suitable for intraday and day ahead markets.
- Step 3: Compute energy production forecast, based either on the results obtained from the step 1 or the step 2 and with machine learning approach for implementation of site-specific data

Step 1 requires different input data the main ones are:

- Global forecast dataset (GFS) produced by NCEP (US National Center for Environmental Prediction) or Integrated Forecasting System (IFS) produced by ECMWF (European Centre for Medium-Range Weather Forecasts)

- Some surface observation data (sea surface temperature, humidity of soils, etc.).
- For the assimilation procedure:
    - Additional Surface weather observation data (hydrometeorological variables, e.g., temperature, water vapours, wind speed and direction) provided by authoritative (e.g., ECOMET) and personal weather (e.g., underground) stations network.
    - Another hydrometeorological variables (e.g., reflectivity, soil moisture, land surface temperature, sea surface temperature, integrated water vapor content) retrieved from ground-based (radar) and spaceborne remote sensing system (Sentinel data from the European Copernicus service).

The outputs of the step 1 are meteorological parameters for the simulated domain on a 3D grid.

Step 2 requires site specific and technical additional data for energy production estimation (e.g., turbine power curve)

For the step 3, the outputs of steps 1 and 2 are intermediate data used as input data. The outputs are horizontal wind field in the lowest part of the atmosphere on a 2D grid of spacing around 2-3 km and hourly power generation predicted by deterministic model.
In this step site specific observation data and historical data are provided by **DUF**.

Most of the dataset describe above are open-source data available freely, except:

- **DUF** site specific historical and observation data (e-g. hourly basis generation, wind speed...), confidential and to not share publicly.
- **DUF** forecast data, if necessary, to model training or assessment, which are commercial ones, used freely in the framework of this research project, but not possible to share publicly.
- no implicit or explicit references to the wind farm or the owner companies must be published
- Local weather observation data at the wind farm location site which are private data (to keep confidential inside the EVEREST project).
- Some third-party observation data (radar data or personal surface station) which are available only for research purposes within the scope of the project and available after approval of the owner (e.g., Italian Civil Protection Department). In the framework of the EVEREST project, these datasets could be used freely, but not possible to share publicly.

Most of the output data relevant to the energy production pilot are proprietary model output. Categories of the research outputs produced by the use case are listed as follows:

- The WRF model output from step 1 described above.
- The energy prediction model output from step 3 described above.

3.2 FAIR data

Making data findable, including provisions for metadata

The data will be stored on data repositories with digital object identifiers. We will choose in priority public repositories as long as they allow us to comply with the constraints on the datasets access. Datasets will all have a metadata description, and, in the case of datasets with access restrictions, their metadata will be publicly available. A semantic versioning scheme will be used to track versions of the datasets. The partner responsible for generating that data will be the point of contact for requesting an access to the data.

Making data openly accessible

For the energy production use case, there will be no restrictions on the use of dataset for only research activity in any domain. No control on the use of the dataset will be done, except at the downloading step where applicants which want to use a dataset must declare for which research activity and accept to quote EVEREST in acknowledgements in case of publication.

In order to make data interoperable and increase the data re-use, the research data will be provided on an open scientific format largely used in the meteorology applications, and not in a proprietary model format. The format is NetCDF (Network Common Data Form) devoted for storing multidimensional scientific data including inside its metadata description. Open-source tools are available in order to read easily this format.

3.3 Plan of the outputs

The research outputs produced by the use case are:

- The WRF model output: External researchers will have access to a new source of local weather forecast over Italy for their own applications in order to evaluate the impact of assimilation and procedures applied in the framework of EVEREST, without any limitation on the application domain except the list of weather parameters provided in the dataset.
- The energy production model output: External researchers will have access of real energy production from windfarms which can be input requested by another application models, such as for example models to determine when maintenance could be applied, etc.

For each category, a specific table provides additional information.

Table 2 – Research dataset for ENERGY PRODUCTION use case

ID	ITEM	DESCRIPTION
D3	Dataset name and reference	WRF output (reference will be updated)

	Dataset description	2D and 3D meteorological fields, over different locations at hourly temporal resolution, for a selected period.
	Standards, format and metadata	NetCDF format
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	During the project: Outputs for the selected periods will be stored for the duration of the EVEREST project, with a focus on parameters used for air quality application. After the project: Outputs will be stored on the CIMA storage system. Sample data may be archived for future research, e.g., Zenodo like repository, and Earth System Science Data journal - www.earth-system-science-data.net
D4	Dataset name and reference	Energy prediction output (reference will be updated)
	Dataset description	Energy production prediction over different locations (wind farms set in Italy) at hourly temporal resolution, for a selected period.
	Standards, format and metadata	NetCDF format
	Is dataset confidential? Must be encrypted?	Spatial coordinates will be "anonymized" (relative coordinates) to avoid any link to the wind farm sites
	Data sharing/access inside EVEREST	Yes

	Data sharing/access outside EVEREST for research	Yes
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	During the project: Outputs for the selected periods will be stored for the duration of the EVEREST project. After the project: Outputs will be stored on CIMA storage system. Sample data may be archived for future research, e.g., Zenodo like repository, and Earth System Science Data journal - www.earth-system-science-data.net).

4 TRAFFIC MODELING USE CASE DATA MANAGEMENT PLAN

This section describes the DMP for the traffic modelling use case.

4.1 Data Summary

Traffic modelling and prediction is a critical component for smart cities to build their intelligent traffic management system (ITS). The goal is to find a true traffic data model representation of the city, which is used for providing precise traffic predictions.
Our computation ecosystem starts with reading big raw sensory data, both real-time and long history records. Traffic simulator subsequently converts the sensory data into a traffic model as well as into rich training sequences for prediction model training. Next, a traffic prediction model is learnt from the training data set, finally being exploited by route calculation service.
The use case consists of three computationally and data intensive steps:

- Step 1: Compute traffic data model from FCD data and O/D matrix. The computation is typically executed on the HPC server.
- Step 2: Train and regularly update the prediction model based on traffic data model. The computation is typically executed on HPC server, Microsoft Windows Cloud platform, or can be executed also on FPGA-accelerated platforms.
- Step 3: Online routing utilizing the trained prediction model. The computation is typically executed on HPC server or can be executed also on EVEREST FPGA platforms.

For Step 1, the data input is a provisioned origin-destination matrix (O/D) and a large historical data set of floating car data (FCD). FCD is represented by geo-positions and the raw and noisy speeds of vehicles sensed approximately each 5 seconds from navigation devices, that is from millions of devices every day over the period of several years worldwide. However, our model will operate on selected cities only (like Vienna) counting thousands of vehicles daily and with the data from the time window of a limited period.
The calculated traffic model is represented by speed profiles with metadata for main roads in a given city. The profiles are organized in the form of aggregated speeds over 10- or 15-minute intervals across the week period. Additional dimensions are a seasonality attribute (month of the year) and weather condition factors.

For Step 2, the speed profiles with metadata calculated in the first step can be used as input sequences for training the prediction models. The large amount of vector samples can be obtained for each road element under prediction. The envisioned number of road elements typically relate to the number of main crossings in a city. As an example, the city of Vienna counts several thousand

crossings, resulting in on average four times more of road elements leading to that many of independent prediction models as a result.

For Step3, the trained prediction models will be incorporated into the online routing **SYG** platform/simulator. A set of experiments based on different scenarios will be performed, e.g., online routing on the level of cities (smart-city routing) and routing on the level of country. The collected data will be provided in a raw format as well as a post-processed statistical overview including the used methodology.
The datasets used in our use case are with respect to the access scheme classified as follows:

- O/D destination matrix is a purchased data under license terms, thus are private data with the possibility to keep confidentially available to EVEREST project.
- Historical weather data.
- SYGIC FCD data used in this use case will be available to EVEREST project, and in a limited scope can be publicly available as open source.
- Calculated traffic model in the form of speed profiles can in the limited scope be shared publicly to the research projects.
- Learned Prediction model with its test dataset results can be provided as publicly available benchmarks to research projects.

4.2 FAIR data

Making data findable, including provisions for metadata
The data will be stored on data repositories with digital object identifiers. We will choose in priority public repositories as long as they allow us to comply with the constraints on the datasets access. Datasets will all have a metadata description, and, in the case of datasets with access restrictions, their metadata will be publicly available. A semantic versioning scheme will be used to track versions of the datasets. The partner responsible for generating that data will be the point of contact for requesting an access to the data.

Making data openly accessible
For the traffic modelling use case, there will be no restrictions on the use of dataset for research activity. The only restriction might be the purchased data (e.g., origin-destination matrix and historical weather data), which will follow the license term of the provider. No control on the use of the dataset will be done, except the downloading step where applicants must declare for which research activity they want to use the dataset, and accept to quote EVEREST in acknowledgements in case of publication.
In order to make data interoperable and increase its reusage, the research data will be provided in an easy usable format ready for download and immediate use, typically vector data in CSV format.

4.3 Plan of the outputs

There are five research outputs produced by the use case as described in the following table.

Table 3 – Research dataset for TRAFFIC MODELING use case

ID	ITEM	DESCRIPTION
D5	Dataset name and reference	FCD data sample
	Dataset description	Floating car data (speeds on GPS locations) for a small geographical region (part of a city) for a given time interval (e.g., one month)
	Standards, format and metadata	CSV
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	Data sample will be archived on EVEREST premises during project duration and for the future use after the project end possibly moved to **SYG** premises (e.g., Zenodo like repository, and Earth System Science Data journal - www.earth-system-science-data.net)
D6	Dataset name and reference	Traffic model profiles
	Dataset description	Annotated speed labels and other metadata on a sample grid road network

		(part of a city on a higher-class roads)
	Standards, format and metadata	CSV
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	Data sample will be archived on EVEREST premises during project duration and for the future use after the project end possibly moved to SYGIC premises (e.g., Zenodo like repository, and Earth System Science Data journal - www.earth-system-science-data.net)
D7	Dataset name and reference	Traffic prediction training sequence
	Dataset description	Data sample for ML learning of prediction algorithm
	Standards, format and metadata	CSV
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation	Data sample will be archived on EVEREST

	(including storage and backup)	premises during project duration and for the future use after the project end possibly moved to SYGIC premises (e.g., Zenodo like repository, , and Earth System Science Data journal - www.earth-system-science-data.net)
D8	Dataset name and reference	Simulated probabilistic speed profiles
	Dataset description	Simulated probabilistic speed profiles for the selected city (e.g. Vienna or Prague).
	Standards, format and metadata	CSV
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes (Free Access)
	Is dataset reusable?	Yes
	Archiving and preservation (including storage and backup)	Archiving and preservation will follow rules of the selected service for the data storage (e.g., Zenodo like repository and Earth System Science Data journal - www.earth-system-science-data.net)
D9	Dataset name and reference	Benchmark dataset for the simulation of routing in Smart City
	Dataset description	The configuration of the traffic simulator and the setting of input parameters for the simulation of routing in a selected city (selected use cases).

	Standards, format and metadata	CSV
	Is dataset confidential? Must be encrypted?	Not confidential
	Data sharing/access inside EVEREST	Yes
	Data sharing/access outside EVEREST for research	Yes
	Is dataset reusable?	Yes (research activity only)
	Archiving and preservation (including storage and backup)	Archiving and preservation will follow rules of the selected service for the data storage (e.g. Zenodo like repository and Earth System Science Data journal - www.earth-system-science-data.net)

5 Access to computational resources on data centres

At this date, the EVEREST architecture can include:

- Servers shared between partners such as potentially **IBM** and **IT4I** servers and private partner's servers (**NUM**, **CIMA**, **IT4I**, **SYG**).
- Roles of each servers between pre- and post-treatment, execution, visualizations, data storage (temporary, permanent, ….).

IT4I operates Barbora and Salomon supercomputers, a special system for artificial intelligence computation NVIDIA DGX-2, and a petascale system KAROLINA (under operation from Q2/2021). User access to IT4I supercomputing services is based on projects — membership in a project provides the access to granted computing resources. Computational resources may be allocated via several allocation mechanisms. For the EVEREST project[2] are relevant Open Access Competitions. More detailed information is available at the complete documentation[3].

[3] https://docs.it4i.cz

References

None