REACH

NEXT GENERATION DATA INCUBATOR

EXPLORE PHASE TECHNICAL SPECIFICATIONS

11/05/2023

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

DATA PROVIDERS

1 ANNEX I. Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** The mock-up solution is suitable and correctly addresses the challenge/theme selected over the REACH dataset/s. The Big Data solution architecture proposed is adequate to tackle the data management issues associated to the solution in mind. "To what extent does the applications handle the data provided?"

The proposed system (iChain platform) was designed and set up to manage events, structured according to the EPCIS standard by GS1. Initially dedicated to implement services dedicated to improve or introduce collaboration along the supply chain, it can be used to manage events in any other observable and manageable context, and Energy management is not an exception. Data can be fed into the platform manually or automatically by fetching data directly from sensors or from streams coming from other available information systems. The data included in the provided data set have been transformed into the "event" format (attaching location, time and other information required by the standard) and sent to the platform, which can now handle individual events in order to design and implement the required services.

The iChain architecture is cloud based, and it can be connected to IoT devices either directly (for small, simple installations) or through an edge server installed on-premises.

2. **SELECTION OF ALGORITHMS AND TOOLS:** The indicated Data Science approach, i.e. algorithms chosen, and Big Data architecture approach, i.e. tools chosen may successfully accomplish the required data governance, processing and analysis. A clear understanding of the used REACH dataset/s is demonstrated.

The proposed challenge requires on the one hand to check the past energy expenditure by matching the evaluated cost with the energy bills received by SMAT, and on the other to develop a forecasting service providing an estimate of energy consumption and cost for the future months.

The first service (cost check) will be implemented starting from information about the evolution in time of energy consumption and the energy cost (the latter being obtained by external data sets). The second service (cost and consumption forecast) will be implemented in a similar way using an internal estimate of future energy consumption and the forecast information about the evolution of the cost of energy provided by AIEE.

For the forecast of energy consumption a comparison between conventional (regression-based) methods and more sophisticated, AI based methods will be proposed by implementing the two algorithms within the iChain platform and evaluating the respective performance for the specific case. AI models deal better with non-linear patterns, but the final selection of the algorithm will also depend on the amount of historical data available for AI / ML networks training.

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

DATA PROVIDERS

3. **TECHNICAL SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** The solution can truly cope with humongous and increasing datasets, potentially from diverse data providers, and is flexible it to adapt to other related domains.

The scalability of the proposed solution is guaranteed by the iChain architecture on several different planes: on a geographical scale, the selected Cloud architecture assures that the system can be extended as needed by simply adding computational resources in the locations of interest. On a customer scale, iChain includes the possibility to install an edge server on-premises when the number of sensors or data streams is reasonably high for the same plant / building. The edge server takes care of some of the data processing, sending only the essential / useful data to the core iChain. At the same time it provides an easy-access point for local events and for information that does not need to be sent to the central system.

Flexibility is guaranteed by the standard selected for the data format. Not only it gives the possibility to have iChain interact with other EPCIS-compliant systems (possibly to send and receive energy-related events from other domains / applications), but it also provides a homogeneous data lake eve if the data originates from heterogeneous sources ingested using dedicated software connectors.

The adaptability from an application development point of view comes from the fundamental principle of iChain to admit special purpose, dedicated software applications sitting on top of the raw data coming from the field. The possibility to apply different algorithms and develop simple or complex services is just another "vertical" software project.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Data sharing challenges, data governance and legal compliance, must be observed. The proposed solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR).

The proposed solution does not deal with personal data: the only non-public information that will be used is the data related to SMAT energy consumption, which will be kept strictly confidential within the project but does not include personal features and therefore is not targeted by the Regulation 2016/679 (General Data Protection Regulation). In any case, in line with the requirements of the Contract, Wiseside will use and process any data provided by SMAT only for the purposes of this Contract and during the length of the Contract. The proposed solution is based on a platform which will ingest the data as provided by SMAT and other data collected from public sources, and use that data for the sole purposes of the project. At the end of the contract Wiseside will delete the data to which it was granted access, unless a different request comes from the Dat a Provider (SMAT)

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Feasible and credible quality process followed for the final product generation. The potential risks in all the phases of the project (design of the solution, development, testing, deployment…) are identified and convincing mitigation plans put in place.

The platform developed by Wiseside - that will be customized to provide the requested solution for SMAT - is based on a cloud architecture, so that all the risks related to unavailability of the service are mitigated. As far as the solution development goes, the proposed approach includes the development and test of the required forecast module based on different algorithms, in order to identify the best cost/performance balance. The basic linear regression algorithm is well known and we do not expect any development risks, so we are confident to be able to achieve the basic (yet least interesting) result. If the amount of available data is large enough, we will also try out more sophisticated AI-based models, for which we may encounter issues that are not easy to identify today. A mitigation plan in this case comes from the operational model of Wiseside, which operates in a network of academic and research organisations granting access to highly qualified support in a wide range of topics, one of which is energy management and forecast. The energy production and forecast is anyway a widely studied topic, and a vast literature is available which makes us believe that a significantly accurate and sophisticated solution to the proposed challenge will be found and implemented.

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

DATA PROVIDERS

REACH

NEXT GENERATION DATA INCUBATOR

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

DATA PROVIDERS