



WP5 Big Data Processing

Dr. Mihai PAUN
Vice-President
Romanian Energy Center



13-14 June 2018 First Project Review



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 731205.



CONTENT – WP5

- Objectives from Grant Agreement
- Big Data Concepts
- Big Data Platforms
- Big Data Platforms ONline Component
- Big Data Platforms OFFline Component
- Big Data Platforms used in WiseGRID Project
- Works and Development Tasks performed
- CRE Team involved in development
- Partners involved in development
- Key Content of D5.1
- General structure of MongoDB Computer Cluster
- Offline processing Computer Cluster
- Deliverables
- Lab testing started



CRE Role in WiseGRID

- **Leading** WP5 WiseGRID Big Data Processing and the two subsequent deliverables, D5.1 (report due in month 18) and D5.2 (demonstrator due in month 21)
- **Leading** T3.2 Privacy and data protection in a multi-actor environment as important part of Work Package 5. Also CRE is designated and activates as DPO within the Consortium level.
- **Contributing** with expertise, research and development activities within Work Packages 1, 2, 4, 12, 13, 14 and 15.
- **Supporting** Dissemination, Business Models, Market Analyses & Project Management activities within Work Packages 16, 17, 19, 20 and 21.



WP5 Overview

Objectives

- The configuration of the **Wisegrid Cloud-based Big Data** allowing for **Real-time Management Of Huge Volumes Of Data**
- **Big Data platform that will run like a cloud service accessible to all WiseGRID Applications**



WP5 Big Data Concepts

EVERYDAY

BIG DATA

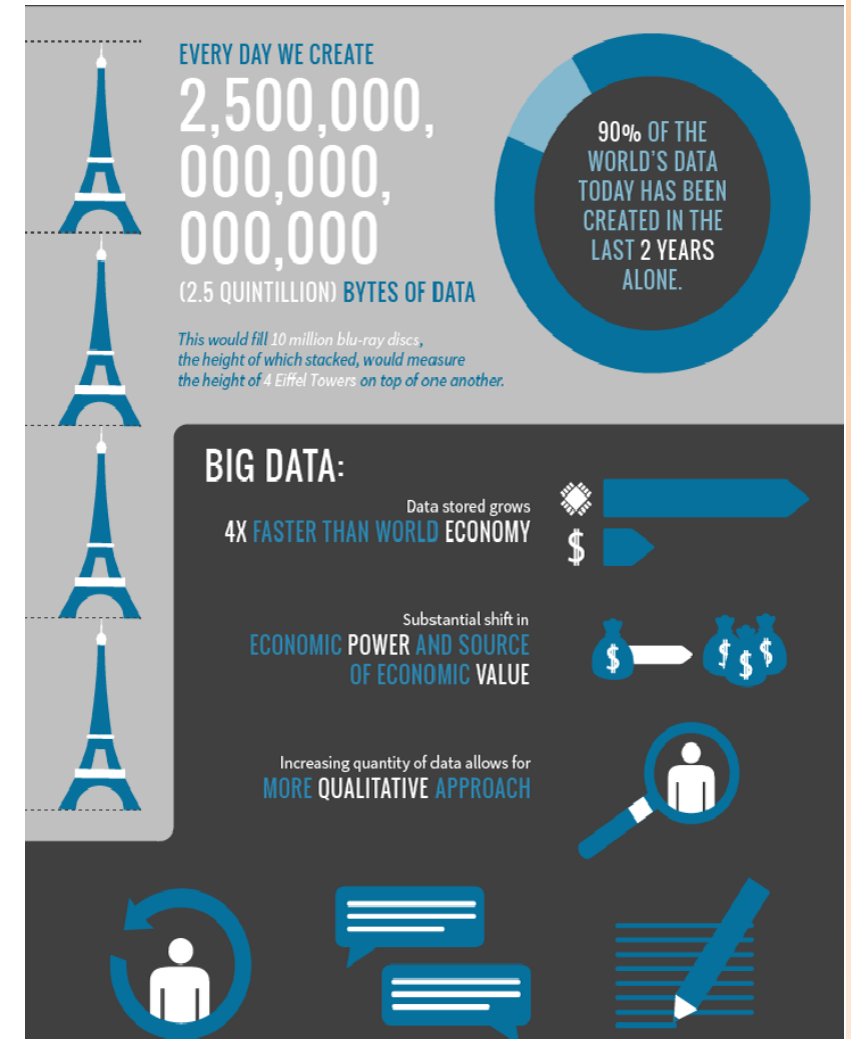


*Big data describes the collection of complex and large data sets such that it's difficult to capture, process, store, search and analyze using conventional data base systems. Its uses are shaping the world around us, offering more qualitative **insights** into our everyday lives.*



WP5 - Big Data Concepts

- There's nothing new about the notion of big data, which **has been around since at least 2001**
- Big Data is your data, every data
- It's the information owned by your company, obtained and processed through new techniques to produce value in the best way possible.
- In 2001 **Doug Laney** started defining concepts and **issued:**
- **"The Three V's" = "Volume, Velocity and Variety"**

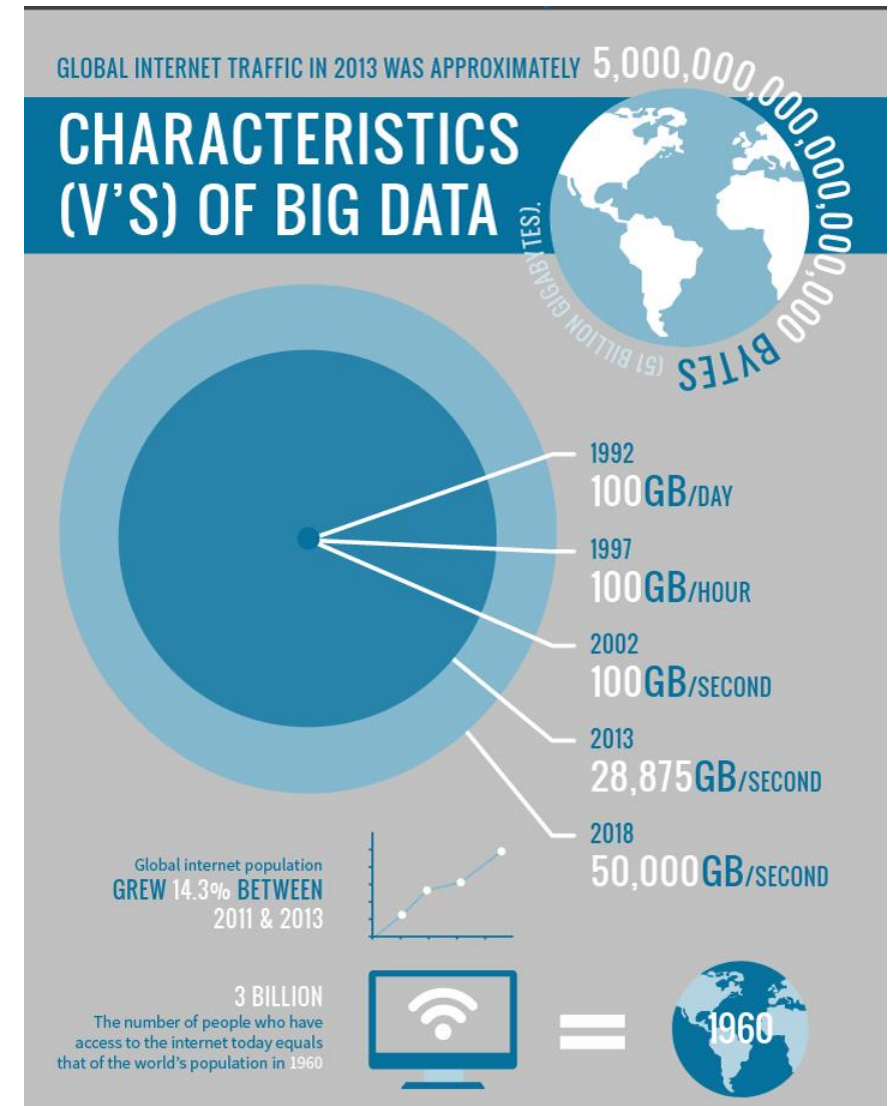




WP5 Big Data Concepts

Volume - Examples:

- A typical PC might have had **10 gigabytes** of storage in 2000.
- Facebook ingests **500 terabytes** of new data every day;
- A Boeing 747 will generate **240 terabytes** of flight data during a single flight across the US;
- The proliferation of **Smart Phones**, data they create and consume;
- **Sensors embedded** into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including videos.

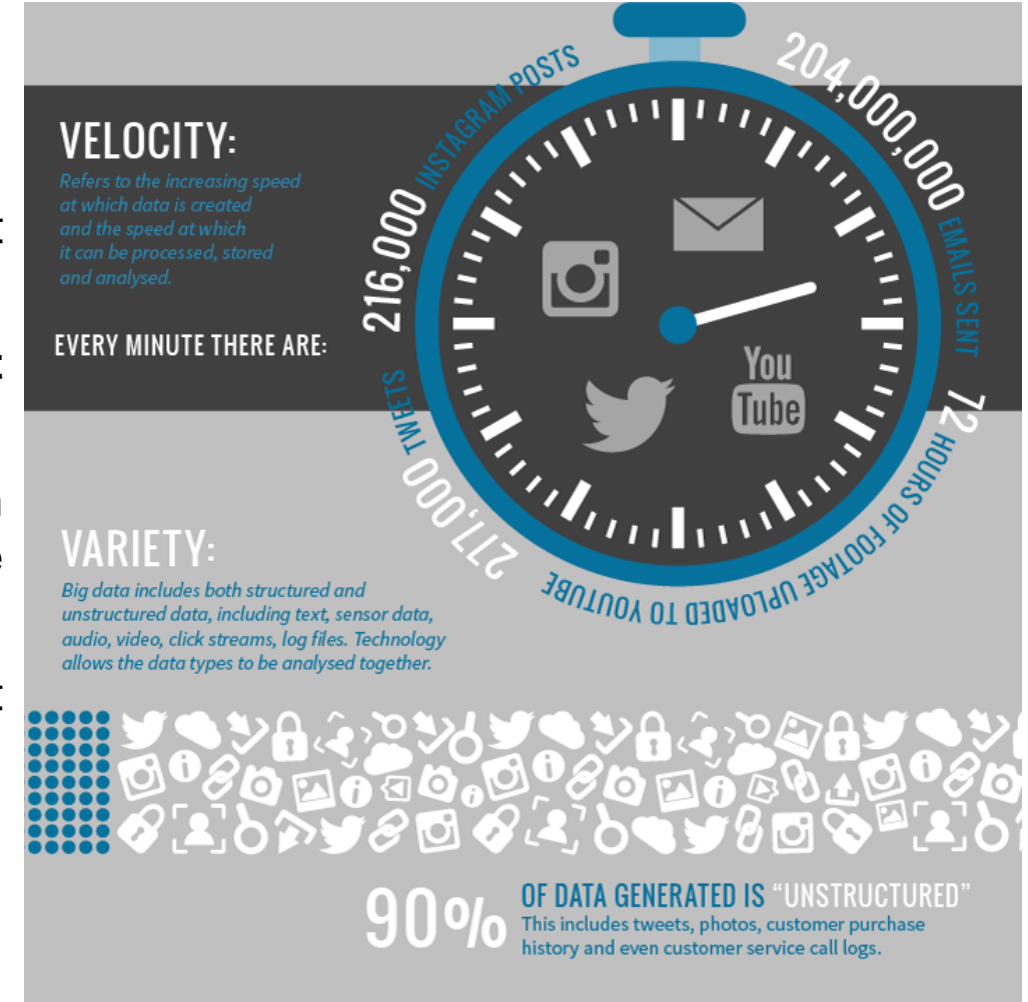




WP5 - Big Data Concepts

Velocity:

- Clickstreams and add impressions capture user behavior at **millions of events per second**;
- High-frequency stock trading algorithms reflect **market changes within microseconds**;
- **Machine to machine processes** exchange data between billions of devices; infrastructure and sensors generate massive log data in real-time;
- **On-line gaming systems** support millions of concurrent users, each producing multiple inputs per second.

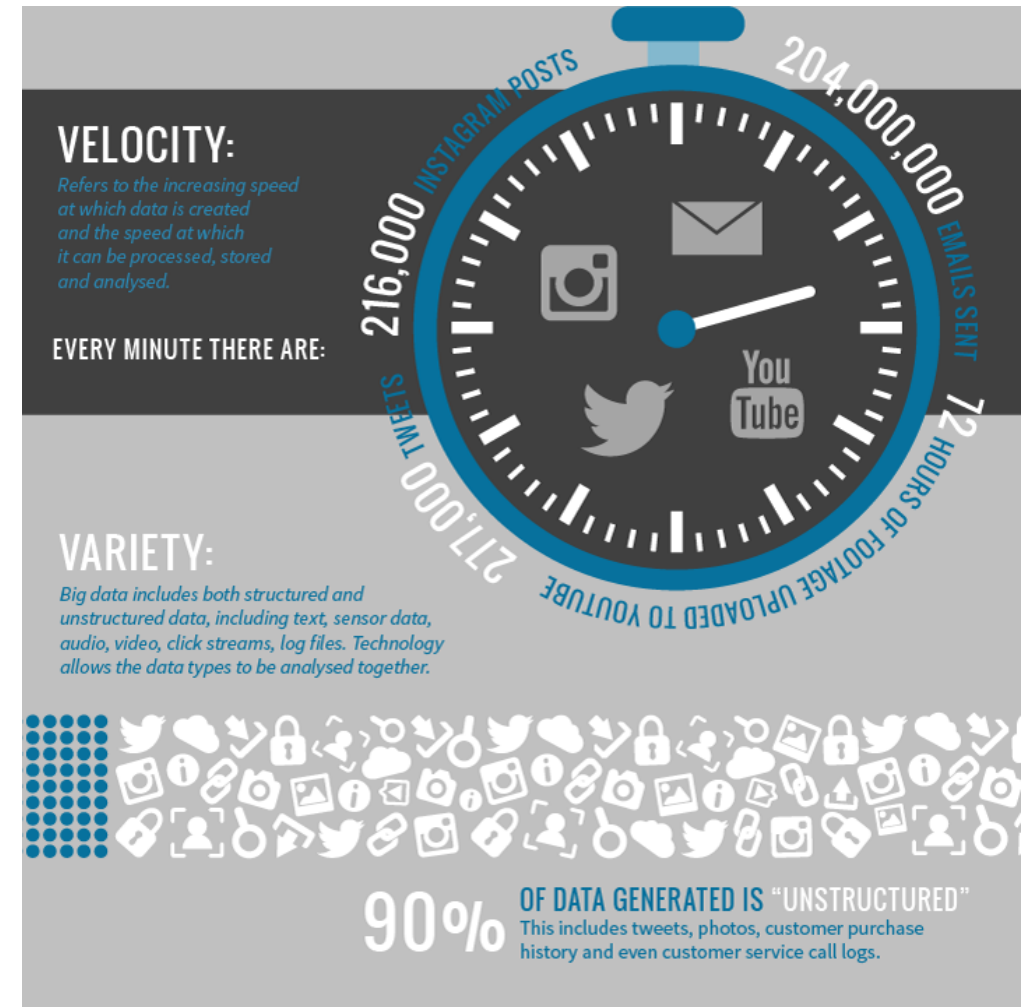




WP5 - Big Data Concepts

Variety:

- Big Data information isn't just numbers, dates, and strings;
- **Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.**
- **Traditional database systems** were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.
- **90% of data generated are UNSTRUCTURATED**





WP5 - Big Data platforms

A Big Data Platform has two main components :

- **Online data processing:** a component for storing huge amount of data in a very flexible format coming from heterogenous sources
- **Offline data processing:** a component for mining and analyzing the stored data to get new information, trends, insights from stored data.





WP5 Big Data platforms ONline Component

WG IOP does not directly access the Big Data Platform but uses **external services** (DB interaction services).

Each application will access the Big Data Platform by **direct Database engine access** through dedicated cluster computers.

In the case of WiseGRID project for the online component the **data generated** from:

- Various sensors,
- Energy smart meters,
- EVs charging stations,
- Energy production units,
- Dispatchable consumers
- Energy storage units,
- DSO data,
- WiseGRID applications
- Prosumers, & other

has to be stored in a system that is easily scalable to fulfill the project requirements





WP5 Big Data platforms OFFline Component

In the case of WiseGRID project for the offline component the data stored by the WiseGRID applications has to be **processed by a scalable application platform** in order to allow the WiseGRID applications to process huge amount of data and supply:

- Trends, statistics
- Estimations and predictions,
- Key performance indicators,
- Others.





WP5 Big Data platforms used in WiseGRID Project

For the WiseGRID project we had to build a **Big Data database and Applications platforms** that has to be **suitable for developing WiseGRID Apps**

- **Online Big Data technologies**, such as the **Database MongoDB**, ingest and store data in real-time and in an operational capacity.



- **Offline Big Data solutions** such as **Hadoop**, process data in batch for retrospective analyses that may touch most or all of a company's data. These technologies are **complementary** and may develop a complete analytics solution, you'll likely need to employ both.



CRE TEAM – EU H2020 PROJECTS



Irina NITA



Catalin CHIMIREL



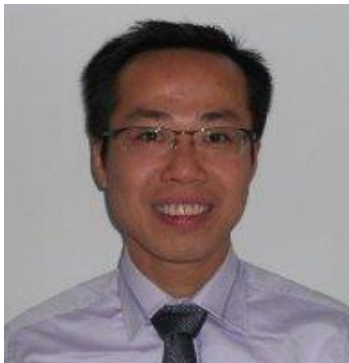
Mihai PAUN



Dan PREOTESCU



Mihai SANDULEAC



Thong VU VAN



Mihai MLADIN



Paul LACATUS



Mihai MACARIE



WP5 - CRE Team involved in Implementation



Romanian Energy Center - CRE Team, involved in the process is **led by** :

- Eng. Paul LACATUS - **ICT Manager**



And **composed by**:

- Dr. Eng Mihai SANDULEAC - Energy specialist and **Technical Manager** in WiseGRID project
- Dr. Catalin CHIMIREL - Energy specialist and **DPO** for WiseGRID
- Dr. Mihai PAUN - Energy specialist, Vice-President CRE





WP5 Other partner teams involved in development

Beside CRE, teams from other WiseGRID Consortium members were also involved with significant contributions, like:

- **ETRA :**

- Álvaro NOFUENTES
- Alberto ZAMBRANO



- **ENG:**

- Giuseppa CARUSO
- Leandro LOMBARDO



- **VS:**

- Stefan MEIR





WP5 Partners

Partner number and Short Name	Proposed Role
1 - ETRA	ETRA – Support for defining the WiseGRID cloud-based big data infrastructure
3 - ENG	ENG – Defining interfaces between IOP and Big Data system
4 - CRE	CRE – Defining the WiseGRID cloud-based Big Data infrastructure and WiseGRID Big Data Cloud-based infrastructure Testing and Refinement
11 - ICCS	ICCS – Lab testing phase , integration of Big Data infrastructure with complete system



WP5 Essential content of deliverable D5.1

The essential content of the D5.1 indicate that the **BigData platform for WiseGRID project is based on** two computer clusters:

- **MongoDB computer cluster**

- Easily horizontal scalable to fulfill requirements for the WiseGRID applications
- Linux computer cluster based on open source BigData database management platform



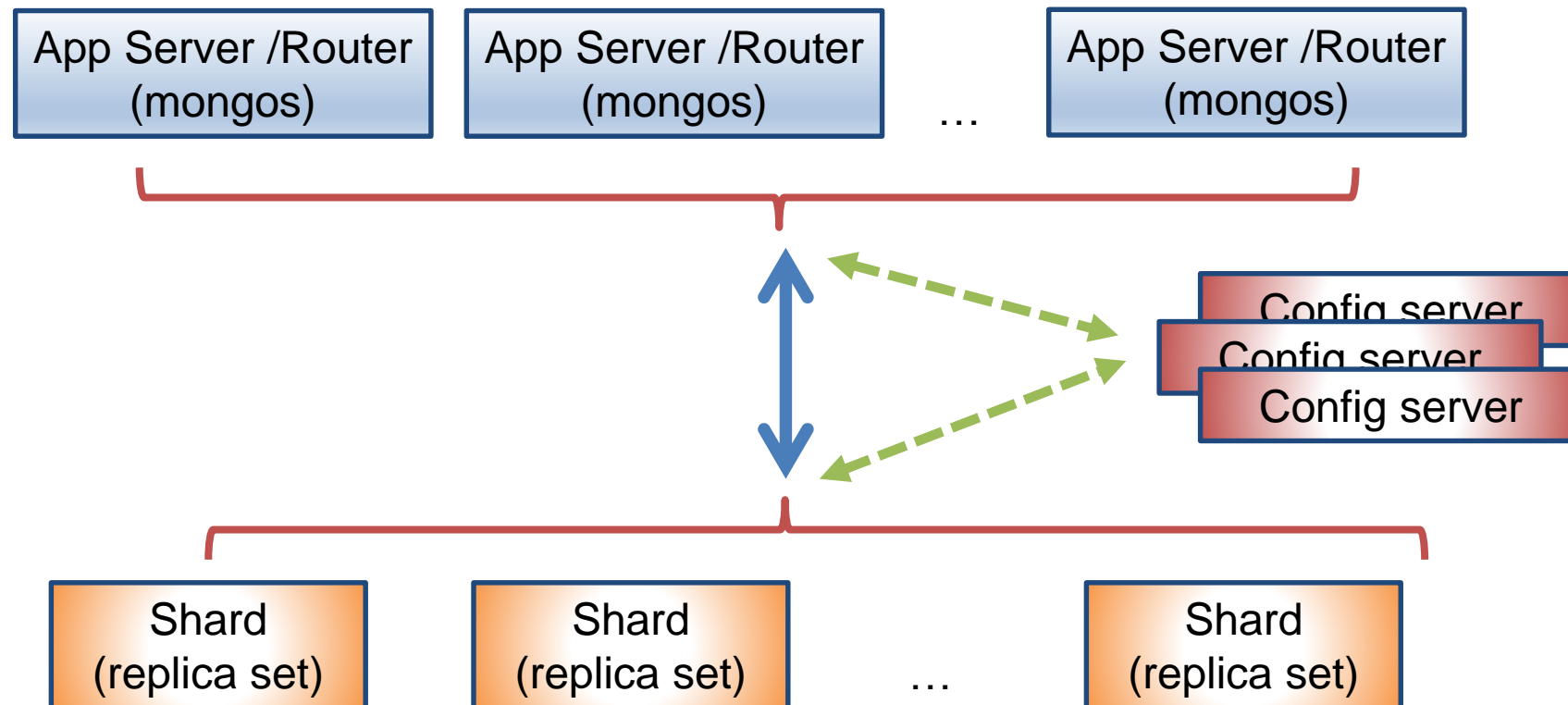
- **Apache Hadoop computer cluster**

- Linux computer cluster based on open source BigData processing platform
- Using Apache Spark application development platform and toolset





WP5 General structure of MongoDB Computer Cluster





WP5 Offline processing Computer cluster

- **Data mining and analytics should process huge amount of data in an OFFLINE process. Distributed computing platforms are providing the necessary power of processing and scalability .**
- **Apache™ Hadoop®** project develops open-source software for reliable, scalable, distributed computing.
- **Apache Spark™** is a fast and general engine for large-scale data processing.





WP5 Deliverables

Deliverable	Title	Lead Beneficiary	Type
D5.1	WiseGRID cloud-based big data infrastructure	4 - CRE	Report <i>submitted</i>
D5.2	WiseGRID Big Data Cloud-based infrastructure Testing and Refinement	4 - CRE	Demonstrator <i>in progress</i>



WP5 Lab testing started

Mongo DB initial Lab Cluster composed of:

- **6 Odroid C2 ARM 64 data nodes**
 - 3 shards of two replica nodes each with 64GB eMMC
- **5 Odroid C2 ARM 64 control nodes**
 - 3 config nodes
 - 2 application router

All mounted on 3d printed supports on a DIN rail to be mounted in a server rack





INNOVATION @ CRE - EU PROJECTS IMPLEMENTATION



Renewables in a Stable
Electric Grid



SUCCESS - Securing
Critical Energy
Infrastructures



Wide scale demonstration of
Integrated Solutions and
business models for European
smart GRID



Enabling Smart Energy
as a Service via 5G
Mobile Network
advances



CROSS BORDER management of variable
renewable energies and storage units
enabling a transnational Wholesale market



SOGNO - Service
Oriented Grid for the
Network Of the Future



EXPERIENȚE ÎMPREUNĂ.

Thank you!

For more info:

Webpage: www.wisegrid.eu

Email: wisegrid@grupoetra.com

Contact: **Eng. Paul LACATUS**
ICT Manager
CRE Romanian Energy Center
paul.lacatus@crenerg.org



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 731205.



ROMANIAN ENERGY CENTER – WHO WE ARE AND WHAT WE DO



Non-governmental and non-profit Association representing the interests of state-owned and private companies operating in the Romanian Energy Market, in relation with EU and National Institutions; Committed and implementing large European Projects in Research Development and Innovation;



Covers the entire value chain of the energy sector: Production, Transport and Distribution of Electricity, Heating, Gas, Oil, Nuclear, Coal, Renewables, Water, Energy Equipment and Services, RD&I, Legal Assistance;



Contributes to European decision-making Process with the aim of encouraging and promoting investments in low-carbon technologies and support the transition to a decarbonized energy system;



Organizes International and National Events meant to enhance the cooperation and integration of the Projects developed by the Romanian energy sector Organization into the framework designed by the European institutions.



Contributes to Energy Diplomacy, as the Coordinator of the Center for Dialogue and Cooperation on Energy Projects 16+1, China-CEEC



- **WP5 Description of work**

Outcome of the works will be the specifications of the WiseGRID Big Data Cloud-based infrastructure enabling:

- **Data Integration from a variety of heterogeneous data sources**, through the design of appropriate interfaces with the different WiseGRID components but also with different distribution grid elements (such as distributed energy resources, IoT devices and sensors, distribution grid assets, etc.),
- **Storage and Processing of huge data volumes** in a highly efficient and effective manner, through the application of Map Reduce algorithms that will be incorporated in the Big Data Platform and the utilization of NoSQL databases, enabling distributed data processing (to enhance effectiveness through the execution of parallel processing activities), prioritization, clustering and scheduling of processing activities, high data integrity and elimination of data corruption,
- **Adaptation of Service-Oriented-Architectures** (SOA) towards enhancing flexibility, scalability, re-usability, loose coupling, integration of a variety of functionalities, along with the required autonomy allowing for modifications and fine-tuning without affecting other integrated systems and components,
- **Interoperability and Interconnection with a wide range of applications** through the provision of appropriate web-service-based interfaces



WP5 Tasks Overview

Task T5.1

- **WP5 Principles of overall architecture**
 - Definition and Design of the required adapters to enable integration and **data exchange with a wide range of heterogeneous data sources**
 - Design of **appropriate data mining algorithms** for facilitating data segmentation and processing based on the Map Reduce approach
 - Utilization of **suitable NoSQL databases** and incorporation of appropriately structured KPIs regarding the platform's effectiveness



WP5 Tasks Overview

Task T5.2

- **WP5 WiseGrid cloud-based Big Data infrastructure Configuration**

This task focuses mainly on the **configuration and integration of the WiseGRID Big Data infrastructure with the WG IOP and the individual WiseGRID components**, along with the heterogeneous data sources involved in the project spanning distributed energy resources (demand, storage, EVs, renewables) and distribution grid assets.



WP5 Tasks Overview

Task T5.3

- **WP5 WiseGrid cloud-based Big Data infrastructure Testing and Refinement**

The main aim of this task is to **validate the Big Data platform and prepare for real-life deployment and operation** in synergy with the assets involved in the different pilot sites of WiseGRID. The platform will be extensively and iteratively tested and required refinements and modifications will be performed to ensure its compliance with functional specifications and operational constraints