CENAGIS geo-cyberinfrastructure as a R&D environment for a Surveyor 4.0

Dariusz GOTLIB and Kamil CHOROMAŃSKI and Bogusław KACZAŁEK, Poland

Key words: geo-cloud computing, geospatial analyses, AI cyberinfrastructure, GIS

ABSTRACT

In the middle of the 4th industrial revolution, surveyors need to use new geo-information tools effectively. There is a need for an IT cloud-based test environments to support R&D work. The answer to such demands is the CENAGIS cyberinfrastructure originally designed and developed in Warsaw University of Technology, built within the Centre for Geospatial and Satellite Analysis.

It is prepared as a "sandbox" for research with use of virtual computers and many developing technologies, such as containerization, distributed and parallel computing, GPU computing, Artificial Intelligence (AI), and big data solutions. Users have an access to rich geospatial data resources managed by Polish SDIs (from cadastral, through topographic, geological, statistical, to weather, and environmental data) in one place. Available data includes point clouds from aerial laser scanning for the entire Polish territory and satellite images from the European Space Agency. Users can also upload, compare, correct, and process their own measurement data. Extending the Repository's content with data from different countries and regions is possible in cooperation with other entities.

Users can utilize the available software packages with configured access to the data in the Repository. There is also a Geospatial Analysis Platform available to share self-developed analytical services and take advantage of solutions shared by others. There are also groupware tools that allow collaboration between users, especially between researchers and company specialists.

The paper will present the architecture and technological solutions used as well as the proposed work model. The solution is based mainly on open source technologies such as CloudStack, Apache Spark, Apache HDFS and Jupyter. Both relational and NoSQL databases are used. The user can use the infrastructure in the, PaaS (Platform as a Service), SaaS (Software as a Service) and CaaS (Containers as a Service) models.

In this paper, selected use cases related to the work of Surveyor 4.0 will be shown. These are involving the preparation, processing and analysis of a variety of large spatial data sets using modern information technologies. The research results on infrastructure performance and plans for further development will also be presented.

Key words: AI cyberinfrastructure, geo-cloud computing, geospatial analyses

CENAGIS geo-cyberinfrastructure as a R&D environment for a Surveyor 4.0

Dariusz GOTLIB and Bogusław KACZAŁEK and Kamil CHOROMAŃSKI, Poland

1. INTRODUCTION

In the middle of the 4th industrial revolution, surveyors effectively need to use new geo-information tools. There is a need for IT cloud-based test environments to support R&D work. Deloitte company report „Trendy technologiczne 2019" (Deloitte, 2019) discusses phenomena that will cause radical changes in business in the next two years. Blockchain, cognitive technologies and digital reality are mentioned as key developments. Industry 4.0 is primarily further digitalization, but more advanced than before, realized by connecting people, devices and data in advanced networks.

Industry 4.0 is associated with concepts such as Artificial Intelligence, Internet of Things, Augmented Reality, Virtual Reality, Cloud Computing, Big Data, Cybersecurity, System Integration, Blockchain, Digital Twins, Robotics, 5G, etc. Surveyors need to engage in data analysis. Increasingly, they need knowledge of artificial intelligence and growing IT expertise. Surveying instruments are becoming increasingly automated and may become IoT devices in the future. Without IT knowledge, a surveyor has limited ability to consciously use and create intelligent solutions that collect and analyse spatial data. The advantage of surveyors over other professionals is a deep understanding of spatial information acquisition, taking into account many aspects, including legal. Combined with IT knowledge, this allows for the birth of Surveyor 4.0 as an important player in the era of Industry 4.0 (Gotlib, 2020) .

However, Surveyor 4.0 needs modern tools and work environments to meet the challenges posed. As in most cases in technology development, it starts with scientific work and R&D experiments. Therefore, in Poland the challenge of building a geo-cyberinfrastructure dedicated to research and development experiments on geospatial data dedicated to surveyors, cartographers and geographers has been taken up. In 2022, a 3-year project was completed at the Warsaw University of Technology, resulting in the creation of CENAGIS cyberinfrastructure. Cyberinfrastructure is the combination of data resources, network protocols, computing platforms, and computing services that bring information and computational tools to people to conduct research and big data applications in an information-driven world. Geospatial CI (geo-cyberinfrastructure) refers to cyberinfrastructure that uses geospatial theories and geospatial information to transform the way research, development, and education are conducted within and across scientific disciplines (such as environmental and earth sciences, for example) (Yang et al., 2010).

The CENAGIS system is a sandbox IT environment where technologies and geospatial data can be tested and verified. Cyberinfrastructure is the main component of The Centre For Scientific Geospatial Analyses And Satellite Computations. The originator and contractor of the project is the Warsaw University of Technology Faculty of Geodesy and Cartography.

The project was carried out by the Warsaw University of Technology, but its idea is to serve the entire academic community dealing with geomatics in its broadest sense. It is also characterized by openness to foreign partners (we currently cooperate with 27 universitates and institutes and over ten companies).

Project was conducted in cooperation with OPEGIEKA, Hexagon and Cloudferro companies. The first phase of the CENAGIS implementation has been co-financed from the European Fund for Regional Development within the Priority Axis I "Utilisation of research-and-development activities in economy", Activity 1.1. "Research-and-development activities of research units" of the Regional Operational Programme of the Mazovia Voivodship for the period 2014-2020.

2. SYSTEM ARCHITECTURE AND TECHNOLOGIES

2.1. General Characteristics

CENAGIS is an advanced IT infrastructure (cyber-infrastructure) allowing for implementation of geospatial analyses (in areas such as spatial big data or data mining) and satellite computations. The infrastructure allows for performing analyses covering the entire country. Spatial analyses of large areas, such as the entire country opens quite new opportunities to perform new types of research works. The main objective of the computations carried out in the environment is to develop algorithms and models ready to be used by various services and operational activities carried out by different industries and public institutions.
CENAGIS consists of a scientific repository of spatial data for the area of Poland, a computing centre and a system of virtual machines. Repository contains open, vector and raster spatial data for entire Poland. Moreover, as a result of cooperation with partners, access to satellite imageries of the European Space Agency (ESA) is ensured.
A unique solution includes the possibility to configure virtual machines, according to users' demands, ensuring the required computational power, GIS software, disk space and, first of all, the configured access to relevant files spatial data. An open analytical platform plays a key role. It was developed within the project implementation and will be expanded in successive years of CENAGIS infrastructure utilisation within various scientific projects, performed by scientists cooperating with the Centre.
CENAGIS provides the convenient access to big sets of spatial data (with the structures developed especially for scientific purposes), as well as an analytical platform dedicated for scientific analyses. The Centre is by default open for cooperation in the field of many kind of geospatial resources, which are being developed in Poland (such as geodetic, road, railway, forest, geological, agricultural, meteorological, planning, architectural, environmental protection, defence data) in order to their common processing, analysing and harmonisation.
As early as at the stage of creation of the Centre integration of the developed infrastructure with two commercial nodes - OPEGIEKA Company and CloudFerro Company in Warsaw - was assumed; those companies became the partners of the Centre (consortium members). Partnership with OPEGIEKA company will leverage the disk resources and computing power available within the Research and Development GIS Centre OPEGIEKA.

As the second commercial node, the infrastructure of CloudFerro Company will be also used. It was created within the European Space Agency Programme "Innovative Platform Testbed Poland (EO IPT). That project was first of all and one of the biggest space project commissioned by the European Space Agency to Polish entities. It is also the key ESA project concerning the use of big data in the field of high resolution optical observations.
The general diagram of connections of different nodes of the IT infrastructure and planned phases of implementation are presented in Fig. 1.

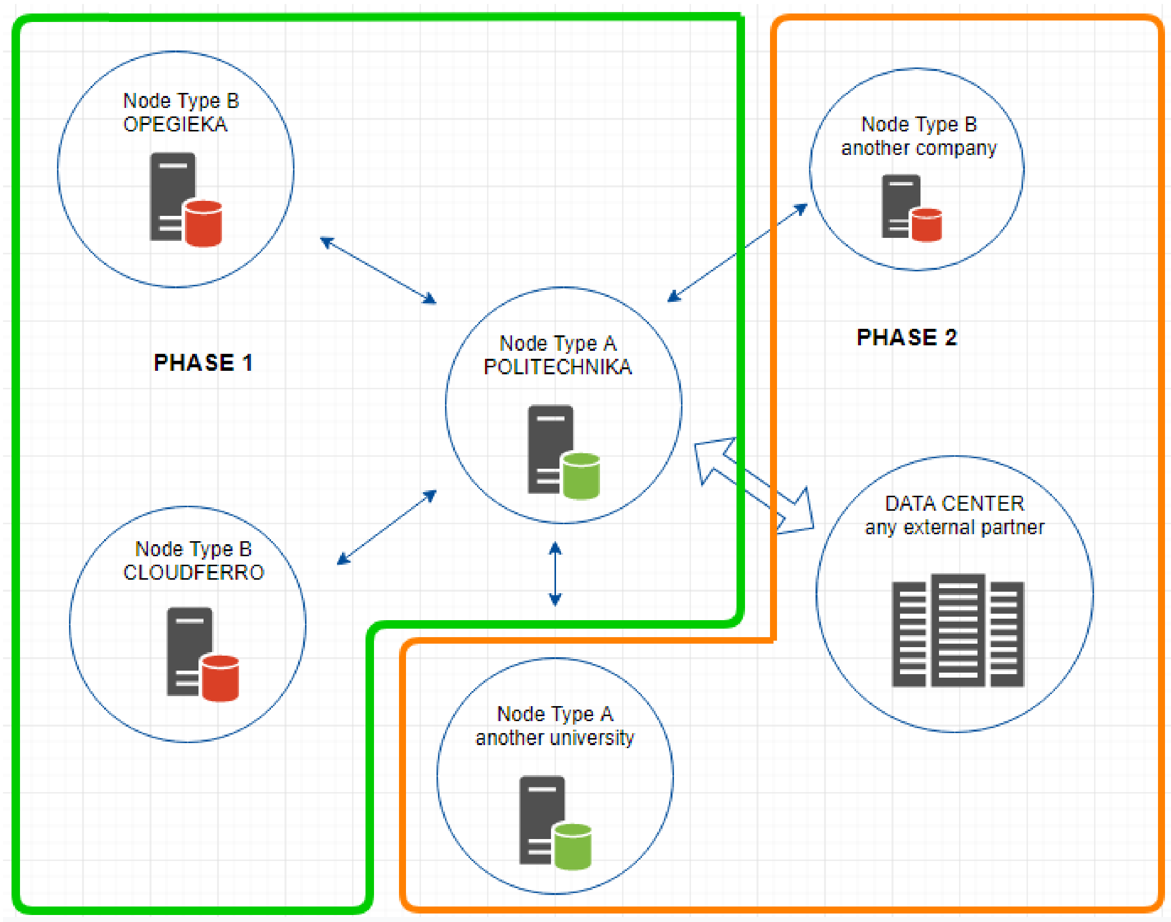


Figure 1. A conceptual diagram of connections of the CENAGIS IT infrastructure. Academic nodes (Type A) and commercial nodes (Type B).

2.2. Architecture and used information technologies

A characteristic feature of the CENAGIS Platform is that users can use a wide range of technologies, both commercial and open source. Proven and stable technologies can be utilized along with experimental ones. This principle covers different levels of the system, from system software, through database software to user applications.

There are four main subsystems visible in the CENAGIS infrastructure from the user's point of view:

1) Virtualization Subsystem
2) Computing Subsystem
3) Service Serving Subsystem
4) Data Repository

In fact, the division is slightly different and more complex. From the hardware side, there are two computer clusters MERCATOR and ROMER (named after famous cartographers - a world-famous one and one of the greatest Polish cartographers) and there is a central subsystem for handling synchronization of various services. A simplified system architecture is shown in Figure 2. In the system, users can learn about and use, among other technologies, those listed in Table 1.

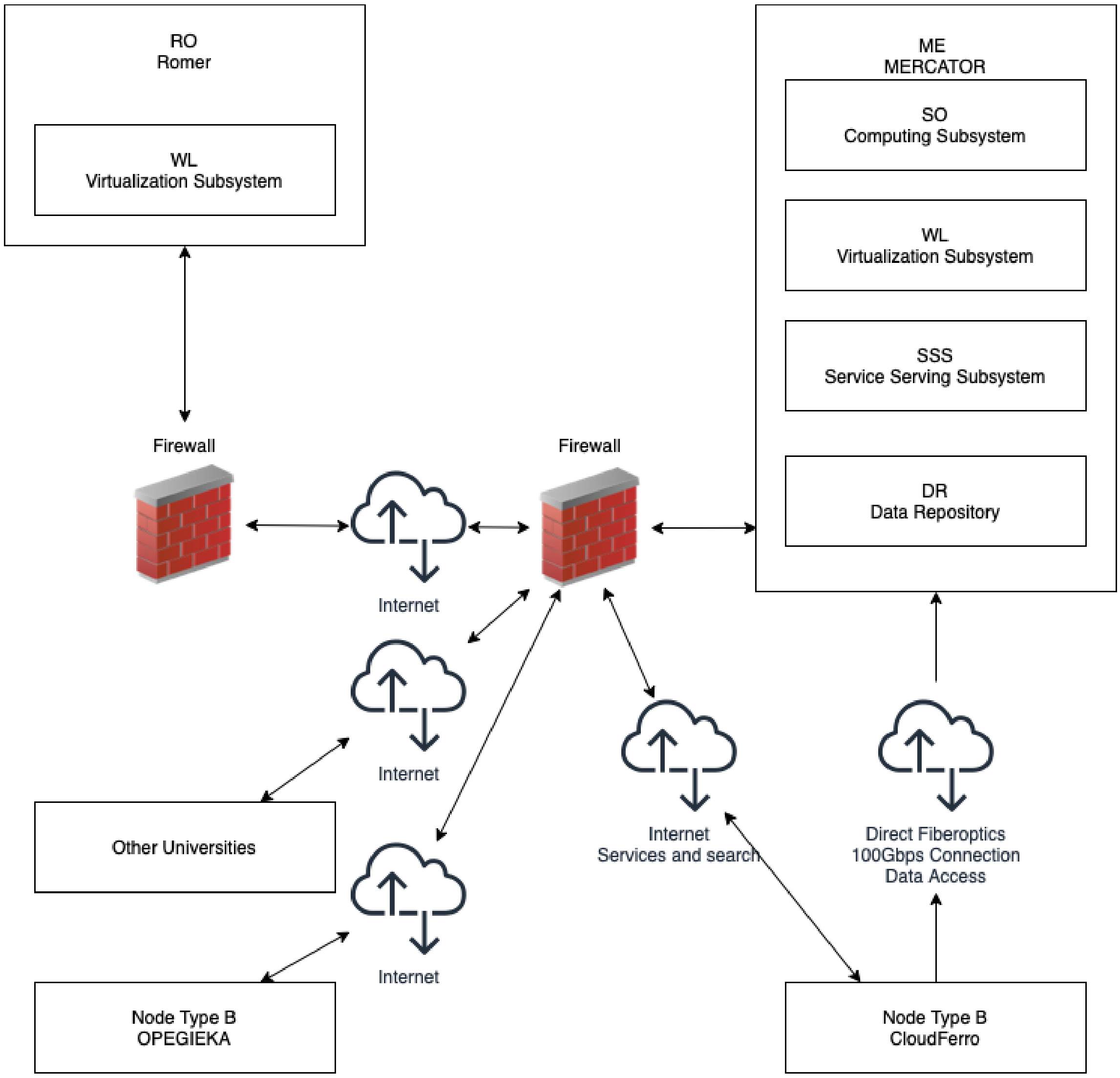


Figure 2 High-level IT architecture of the CENAGIS Platform

Tab. 1 Overview of components utilised in each of CENAGIS subsystems

Virtualization Subsystem	Computing Subsystem
Software: Hexagon (Geomedia, Erdas Imagine/Apollo, M. App. Enterprise, Geospatial Portal, Luciad), ArcGIS, FME, QGIS, PostGIS, GeoServer), Dephos (Limon) **Computer operating system**: Windows, Linux **Hypervisor**: KVM + CloudStack	**Software**: GeoMesa, RasterFrames, Dephos, Accumulo, Jupyter Lab, self-created programming libraries **Computer operating system**: Linux **Big data management technologies**: HDFS, Mesos, Spark, other.

A particularly important component is the software from Hexagon company, the project's consortium member. Fig. 2 shows a view of the M. App Enterprise system used, among other things, for interesting cartographic presentations and the creation of advanced data dashboards using Service Serving Subsystem.

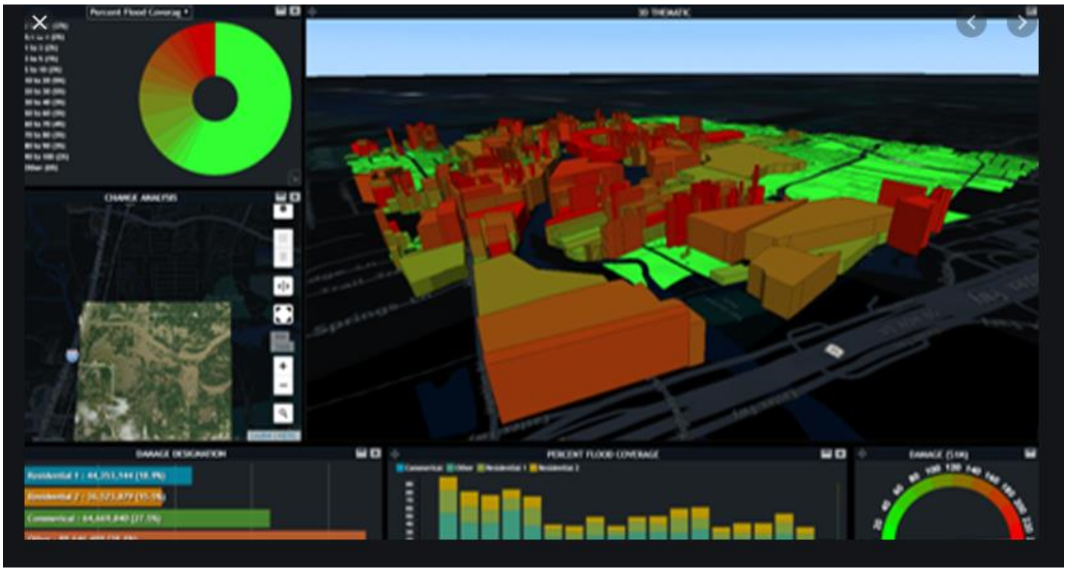


Figure 2 A view of the software available to CENAGIS users - M. App. Enterprise (Hexagon).

The CENAGIS platform was implemented in a manner that allows users to use services in 4 typical cloud models: PaaS (Platform-as-a-service), SaaS (Software-as-a-service), CaaS (Containers-as-a-Service) depending on their needs.

2.2.1. PaaS Model

The use of the PaaS model is possible primarily through the Virtualization Subsystem. User gain access to their own virtual machine or a collection of virtual machines tailored to their specific work. One can use a selected operating system and allocated resources of CPU, GPU, RAM, HDD or computers with preconfigured application software e.g. for ETL process creation (FME), visualization and spatial analysis (Hexagon, ESRI, QGIS) or database management systems (PostgreSQL) and file data repository (HDFS).

2.2.2. SaaS/CaaS Model

By choosing this model of work, the user accesses the system mostly by the JupyterLab (Kluyver el al., 2016) interface and the Service Delivery Subsystem. From the level of Jupyter, one can execute scripts written in Python, R or Julia programming languages. From the JupyterLab level, one can invoke analyses with use of Apache Spark (Zaharia et al., 2016). Spark cooperates with GeoTrellis and GeoMesa (Hughes et al., 2015) libraries to perform spatial analyses on large amounts of spatial raster and vector data. Spark has access to Apache HDFS (Hadoop Distributed File System) to perform parallel computing on large datasets stored in repository (Shvachko et al., 2010). The whole system is managed by Apache Mesos and Marathon, that control resources usage, user quota, access credentials and service continuity. The user can also use predefined analytical services available through the Access Panel – main, high-level interface of CENAGIS infrastructure.

3. FUNCTIONALITY AND METHODS OF USAGE

The infrastructure described above is primarily intended to perform the function of a Geospatial Analysis Research Centre. When designing the system, different target user groups were envisioned: from analysts without geoinformatics knowledge to advanced developers. Depending on their skill level, users benefit from different methods of access to the system and other functions (including a different model of working in the cloud).
Users can conduct work themselves as well as outsource it to employees of the centre or other users of the Platform. In general, the infrastructure is designed to conduct research and experiments in cooperation between science and industry. The basic substantive scope of R&D work that should are supported by the CENAGIS Platform includes:

- Applications of geoinformation in smart city and urban planning
- Photogrammetric methods of processing 3D terrain data
- Spatial Data Infrastructure – testing new concepts, architecture, and technologies
- Multi-criteria geospatial analysis
- Using Machine Learning methods in geospatial analysis
- Modelling of geographic phenomena
- 3D city modelling
- Cartographic modelling
- Spatial Data Mining
- Testing Location-based services (LBS) and navigation applications
- Optimization of IT infrastructure for geospatial solutions

Access to most functions of the system is possible through a specially prepared Access Panel. It allows, e.g.:

- Browsing data in the CENAGIS Repository
- Browsing and editing metadata in the Resource Catalogue
- Viewing and running Geospatial Services (Geospatial Services Node)
- Managing virtual computers

– Using the computing environment
– Using team communication tools

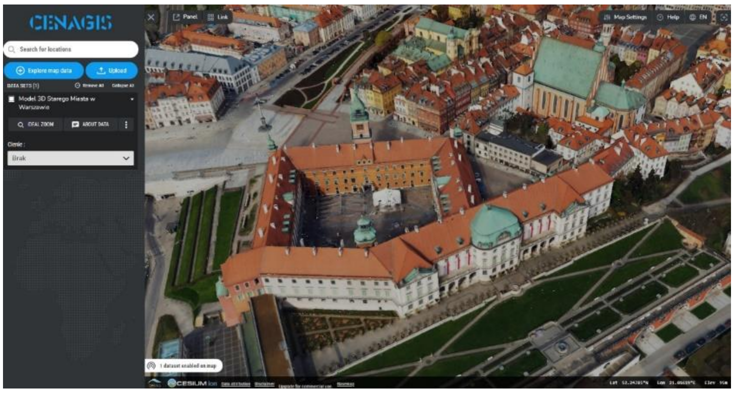
Selected detailed features of the system are shown below.

3.1. Data exploration

Users can view the data available in the System Repository in several ways:

1) Viewing data from the CENAGIS Repository, own data and OGC services using the CENAGIS MapViewer web browser (including point cloud viewing, multi-temporal comparison and 3D visualizations) - Fig. 3 a and b.
2) Browsing data stored in the CENAGIS Repository, own data and OGC services using the Hexagon Geospatial Portal (including downloading data fragments on personal CENAGIS drive)
3) Browsing the CreoDIAS Repository (access to ESA satellite imagery) in two different ways, including simplified access via CENAGIS SatExplorer (Figure 4)

A



B

Figure 3 Possibility of visualization in CENAGIS MapViewer. A) 3D models (mesh), B) LIDAR point clouds

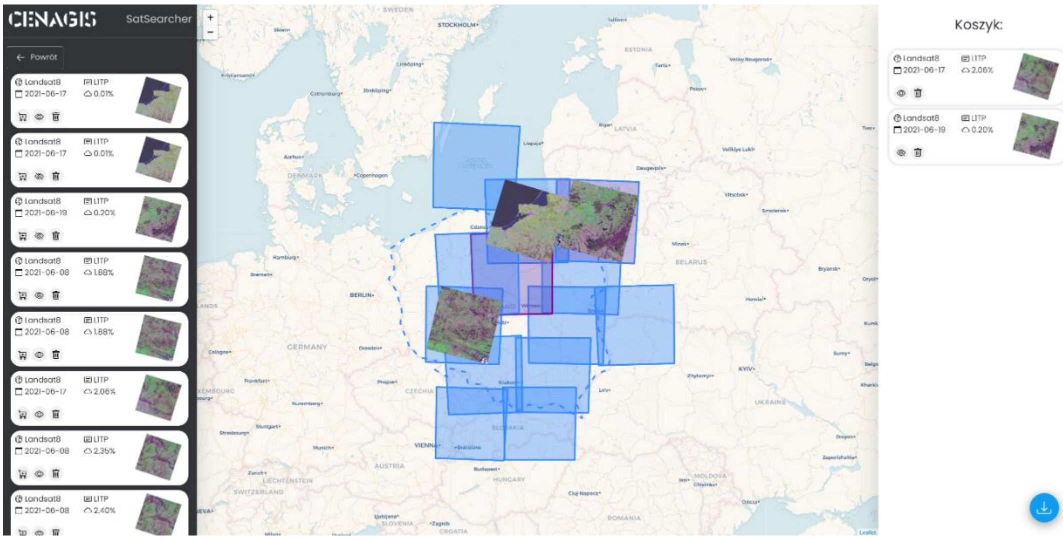


Figure 4 Search engine for ESA satellite images - CENAGIS SatExplorer

An important feature available in the system is the ability to view and edit metadata in the Resource Catalog. It allows users to:

- Search resources (data, software, work environments, projects, research teams, publications, documentation)
- Analyse metadata
- Edit users' own metadata about personal resources added to the Repository

3.2. Application serving

CENAGIS platform also serves as a Geospatial Service Node. It is a place where applications and services prepared by the system developers and users are available. The platform allows users to give access to their applications (including those developed in the M. App. Enterprise environment). It allows to present their research, prototype or production solutions and to examine opinions during its prototyping stage.

3.3. Virtual laboratories

Working on a virtual computer resembles working on a classic desktop computer. In contrast to the classic environment, one can access it from any computer with Internet access, share and display the work environments. Access to different operating systems is possible: Windows and various versions of Linux (e.g. Ubuntu, CentOS). This allows, for example, to examine data processing and visualization processes on computers with different number of CPU, RAM, GPU. Each virtual computer also provides easy access to large geospatial data resources from the CENAGIS Repository. They are connected as network drives or through a configured connection to a PostGIS database with data loaded and ready to go. Access to software (QGIS, Hexagon – number of licenses are available, ESRI – Bring Your Own License model) with configured connection to CENAGIS Repository data is also assured (Fig. 5).

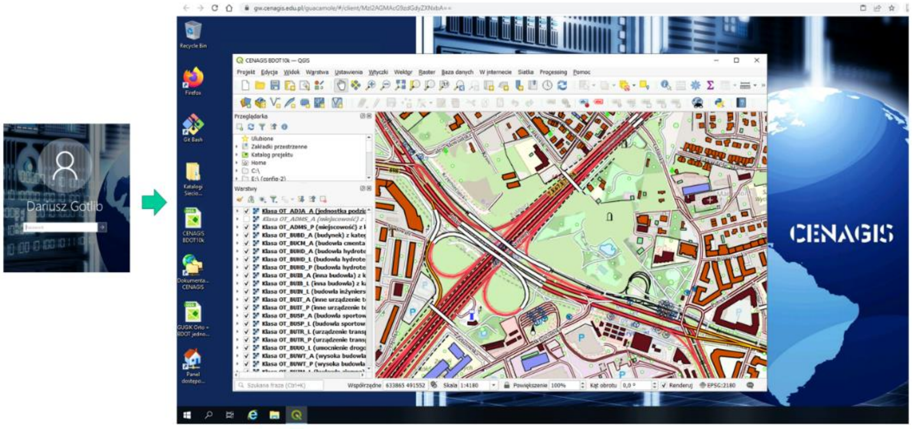


Figure 5 Virtual machine equipped with QGIS software with configured link to spatial data located in the CENAGIS Repository.

3.4. Spatial big data analyses

The most advanced users can leverage the full power of the infrastructure and capabilities of the big data environment using a distributed computing environment. The infrastructure allows to use distributed computing technology on several dozen nodes with the use of tens of GPU cards (NVIDIA TESLA T4 16GB). Users have a possibility to use containerization technology (Docker (Merkel, 2014)) and tools prepared to work with spatial big data e.g. GeoTrellis, Geomesa, Accumulo (Kepner et al., 2014). Access is realized primarily through the JupyterLab interface (Fig. 6).

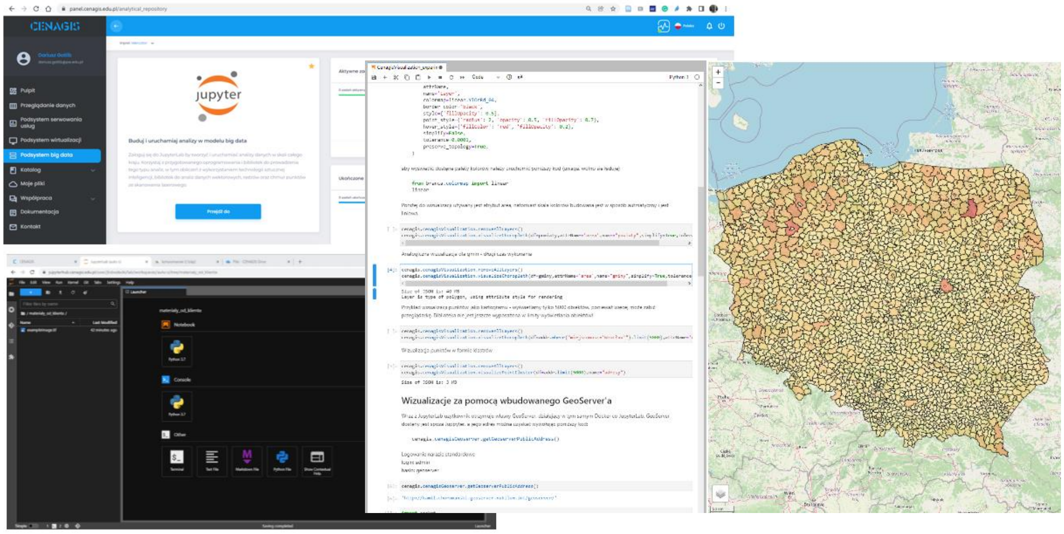


Figure 6 Access to a big data environment through the JupyterLab interface

It is a pre-configured environment for geo-data scientist allowing to process and analyse data in Python language (which is a most popular programming language regarding both PYPL (PYPL Index Website, 2022) and TIOBE (TIOBE Index Website, 2022) indexes). Users have the possibility to use preconfigured libraries, e.g. to facilitate the use of artificial intelligence methods.

3.5. Team collaboration

Apart from the typical data processing tools, the CENAGIS Platform provides additional utilities supporting user collaboration (shared disk space, chat, forum, teleconference platform, bulletin board). The purpose of this type of solution is primarily:

- Finding partners for research projects
- Validating ideas
- Promoting and sharing the results of research
- Promoting and sharing data
- Promoting new technology solutions
- Collaborating to create more complex projects

4. SAMPLE PROJECTS DEVELOPED ON THE CENAGIS PLATFORM

Although CENAGIS Platform is only in the first phase of deployment, it is already used by many users from different research units and companies. In addition to conducting test and experimental work, geo-cyberinfrastructure supports specific R&D projects carried out in various consortia. Below are some examples of how Surveyor 4.0 works.

1) "AI tool for detecting building components and site changes"
The aim of the project is to develop a service for inventory and modelling of key objects of technical and transport infrastructure in BIM technology with the use of artificial intelligence tools. The project involves research on implementation of deep machine learning solutions based on three-dimensional photogrammetric data acquired by drones equipped with cameras and laser scanners. The project is conducted in a consortium of Warsaw University of Technology and SkySnap company.

2) "Improvement of methods of acquisition and processing of remote sensing data from unmanned UAV flying platforms"
The aim of the project is to study the possibility of using laser scanners and multi and hyperspectral cameras in the analysis of forested areas. The project is carried out in cooperation with Dragonfly Vision company. In the framework of the project, air missions with the use of UAV measurement platforms are performed over the forested areas, both in the leafed and leafless season. Systematic air raids will allow to examine the repeatability of the obtained results and to conduct a long-term analysis of the stand condition and support of remote sensing in its inventory. The result of the project will be a methodology of acquisition and processing of the data from unmanned aerial vehicles in forested areas. Within the framework of the

project, a software as a service (SaaS) is being developed, which will enable automatic data processing and making the results available to end users in an interactive form.

3) „Methods for analysing, forecasting, and recommending COVID-19 containment with a particular focus on geospatial analysis".
The goal of the project is to develop a methodology for analysing the development of the COVID-19 pandemic in time and space to build a decision support system for social distancing. Implementation of the project requires collection and processing of multi-source epidemiological, geospatial, demographic, economic, climatic and social data. The collected data will enable the preparation of three simulation models using: multi-agent modelling, deep learning, and Monte Carlo simulations.

4) „MARS - Development of a concept and prototype system for simulation of Mars terratransformation process"
The aim of the project is to develop a prototype simulation system for planning the terratransformation of Mars. The system created within the project will be used for variant analyses of terratransformation of the planet. The proposed methods will use multiscale relief models of planetary fragments in order to select optimal terrains for different management modes, e.g., delineation of communication routes and support of localization and navigation systems. The system uses deep learning machine learning methods, multi-source data and terrain relief models, as well as planetary measurements realized by Opportunity and Curiosity Mars rovers.

5) „Implementation of the Smart Villages concept in the Mazowieckie Voivodeship "
The project aims to support the community and rural development of Mazovia, strengthen traditional and create new networks of relationships between stakeholders through modern means of communication, and above all to raise public awareness of rural development. The basis for achieving the objectives will be conducting research and scientific analysis focusing on identifying elements slowing down the development of agriculture, assessing farming conditions and determining factors affecting the slowdown in agricultural development.

6) „Maloutena and Agora in the urban plan of Paphos"
The aim of the project is to reconstruct the layout of the ancient city of Nea Paphos, located in Cyprus. The result of the integrated research will be the reconstruction of the cityscape and buildings in the form of 3D models. For this purpose photogrammetric UAV flights are used among other measurement tools. The 3D reconstructions will then be spatially analysed to determine the relationship between the layout of buildings and streets and the functioning of the city in terms of visibility, population flow, potential population, etc.

5. SUMMARY

As mentioned in the introduction, in the era of Industry 4.0, the role of the Surveyor is also changing. The scope of work performed and the tools used are altering dynamically. More and more digital data resources are created. It provides a number of benefits and extraordinary opportunities, but also poses new challenges for surveyors. The surveyor 4.0 needs an access to powerful computing resources, large volumes of geospatial data from a variety of sources, and IT solutions to meet new customer requirements. However, not every surveyor has advanced education and skills in IT or geoinformatics.

The dynamics of change generates the need for access to testing environments with functionality adapted to different levels of surveyors' experience. Environment in which, on the one hand, one can learn new technologies and, on the other hand, to experiment and look for a new technological solutions with commercialization potential. Such considerations lay at the core of the CENAGIS geo-cyberinfrastructure project.

The platform is in the first phase of operation. During the Second Phase (II Phase) it is planned to develop the unique cyberinfrastructure of spatial information (an IT/geoinformation system as a computational cloud, basing on spatial big data) towards integration with selected university IT centres, which are conducting research in the field of geoinformation/geomatics, in order to achieve the "university geo-GRID" and to successively modernise the repository and the Platform for Scientific Geospatial Analyses, developed during the first stage of the project.

REFERENCES

Deloitte. (2019). *Raport: Trendy technologiczne 2019, Jak przełamać bariery technologiczne?*

Gotlib, D. (2020). Nowe wyzwania polskiej społeczności geodezyjno-kartograficznej w świetle ery Przemysłu 4.0. *Przegląd Geodezyjny*, pp. 11-15.

Hughes, J., Annex, A., Eichelberger, C., Fox, A., Andrew, H., & Ronquest, M. (2015). GeoMesa: a distributed architecture for spatio-temporal fusion. *Geospatial Informatics, Fusion, and Motion Video Analytics*, pp. 128 - 140.

Kepner, J., Arcand, W., Bestor, D., Bergeron, B., Byun, C., Gadepally, V., . . . Yee, C. (2014). Achieving 100,000,000 database inserts per second using Accumulo and D4M. *IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1-6.

Kluyver, T., Ragan-Kelley, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J., . . . Willing, C. (2016). Jupyter Notebooks -- a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pp. 87-90.

Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. *Linux journal*, p. 2.

PYPL Index Website. (2022). Retrieved from https://pypl.github.io/PYPL.html

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop Distributed File System. *IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, pp. 1-10.

TIOBE Index Website. (2022). Retrieved from https://www.tiobe.com/tiobe-index/

Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, pp. 264-277.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., . . . Shenker, S. (2016). Apache Spark: A Unified Engine for Big Data Processing. *Communications of the ACM*, pp. 56–65.

BIOGRAPHICAL NOTES

Dr Dariusz Gotlib is a cartographer-geoinformatics researcher, assoc. professor of Warsaw University of Technology employed at the Faculty of Geodesy and Cartography since 2001. Associate Dean for Development and Outreach. Originator and manager of the project of creating a Center for Scientific Geospatial and Satellite Analyses at the Warsaw University of Technology. His research and teaching interests include issues in geoinformation, in particular: geoinformatics, mobile and navigation cartography, LBS (Location-Based Services), spatial data infrastructure, and GIS.

Dr Bogusław Kaczałek is Computer Systems Architect and Deputy Head of IT department in OPEGIEKA Sp. Z.o.o. He has over ten years of experience in processing geospatial data and designing and building geospatial information systems from small scale to national scale systems.

MSc Kamil Choromański is PhD candidate on Warsaw University of Technology.and machine learning specialist in SkySnap Sp. Z.o.o. His main areas of expertise are spatial big data and machine learning. He is a member of CENAGIS development team.

CONTACTS

Dariusz Gotlib
Warsaw University of Technology, Faculty of Geodesy and Cartography
Plac Politechniki 1
Warszawa
POLAND
Tel. + 48 501048277
Email:dariusz.gotlib@pw.edu.pl
Web site: www.cenagis.edu.pl, www.gik.pw.edu.pl

Kamil Choromański
Warsaw University of Technology, Faculty of Geodesy and Cartography
Plac Politechniki 1
Warszawa
POLAND
Tel. + 48 602102760
Email:kamil.choromanski@pw.edu.pl
Web site: www.cenagis.edu.pl, www.gik.pw.edu.pl

Bogusław Kaczałek
OPEGIEKA Sp. Z. o.o.
Ul. Aleja Tysiąclecia 11 82-300 Elbląg
Tel. +48 604561272

Email: boguslaw.kaczalek@opegieka.pl
Web site: www.opegieka.pl