DESIGN OF IT PLATFORMS FOR MONITORING DATA OF TRANSPORT INFRASTRUCTURE

IM-SAFE.EU

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Document Information

Name	Remark
Document Name	Design of IT platforms for monitoring data of transport infrastructure
Version No.	V1.0
Due date Annex I	15/07/2020
Report date	30/04/2022 (last updated in 29/04/2022)
Number of pages	73
Lead Author	Mattia Rigotti (IBM)
Other Authors	Ahmed Mossad Ibrahim Elseicy (UVIGO), Ana Sánchez Rodriguez (UVIGO), Andres Justo Dominguez (UVIGO), Cristiano Malossi (IBM), Ioana Giurgiu (IBM), Isabella Alovisi (SACERTIS), Patrycja Sanecka (Mostostal), Rizal Sebastian (TNO)
Dissemination level	Public

Document history

Ver.	Date	Remark	Author	Checked by
Rev. 0.1	30/10/2021	Creation of the document	M. Rigotti	
Rev 0.2	08/03/2022	Overview of software platform chapter	I. Alovisi	M. Rigotti
Rev 0.3	07/04/2022	Final addition of remaining platform descriptions	AAVV	M. Rigotti
Rev 0.4	11/04/2022	Finalization of chapter 3	I. Alovisi	M. Rigotti
Rev 0.5	14/04/2022	Finalization of chapter 4	M. Rigotti	
Rev 0.6	19/04/2022	Additional revision	I. Alovisi	M. Rigotti
Rev 0.7	28/04/2022	Document for reviewers	A. Elseicy, A. Sánchez Rodriguez, A. Justo Dominguez, C. Malossi, I. Giurgiu, I. Alovisi, M. Rigotti, P. Sanecka, R. Sebastian	M. Bohms, P. Darò, J. Żach
Rev. 0.8	29/04/2022	Document for approval	All	F. Cinquini and A. Bigaj-van Vliet
Rev. 1.0	30/04/2022	Document Uploaded	All	A. Bigaj-van Vliet

Document approval

Ver.	Date	Approval	Project's Role	Beneficiary
Rev. 1.0	30/04/2022	A. Bigaj-van Vliet	Coordinator	TNO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Preface

This report is part of the H2020 CSA IM-SAFE project results and is the outcome of WP4 ("Digitalization as enabling technology") Task 4.3 ("Cloud platform for monitoring data") activities, listed as delivery D4.3. This deliverable addresses the platform high-level specifications and design, including user experience, data collection, key domain-relevant functionalities and their backend implementation. It concludes with proposals and recommendations for common design of IT platforms for managing and monitoring data of transport infrastructure, and supporting tasks like asset management, structural health monitoring, and visual inspection for health assessment.

The goal of the overarching T4.3 is to support the generic design for a common data environment (CDE), with particular focus on underlying key consideration and providing recommendation regarding user experience, data handling and high-level functionalities to support the activities of engineers and other stakeholders in the transport infrastructure technical and management sectors.

Task Leader: IBM RESEARCH GMBH (IBM)

Contributors: SAFECERTIFIEDSTRUCTURE INGEGNERIA S.R.L (SAC)

UNIVERSIDAD DE VIGO (UVIGO)

MOSTOSTAL WARSZAWA S.A. (MOS)

TNO – NETHERLANDS ORGANISATION FOR APPLIED SCIENTIFIC RESEARCH

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Table of Content

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Table of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

1 Introduction

1.1 Objectives

This deliverable addresses the design of IT platforms for management of transport infrastructures inspection and monitoring data, and in particular the specification for data management, the user experience, the screen flows and the backend.

The goals of this document are to

- give an overview of use cases of software platforms in use in the transport infrastructure industry
- provide some general considerations for the design of a software platforms based on the provided use cases
- give general recommendations for the design of future platforms frontend/backend.

The stated goal of providing guidance on the design of a IT platforms is motivated by the obvious role that digitalization has been taking for the storage, sharing and analysis of information about the condition and maintenance of transport infrastructures. The current importance of digitalization of data is only overshadowed by its potential and projected growth in importance as a means to cope with increasing traffic loads and resilience threats looming over transportation assets.
Therefore, digitalization has become an important aspect in updating the existing European standards and developing new standards for inspection, monitoring, maintenance and safety of transport infrastructures.

1.2 Challenges in specifying design recommendations

One of the results of the use case overview that was conducted during the preparation of this deliverable is the realization of the wide diversity of approaches and functional technical needs that are being supported by the variety of data and software platforms currently in use in the industry. This heterogeneity makes it very challenging to give specific recommendations about standardization that will meaningfully encompass all use cases and don't limit the range of operation of domain practitioners like engineers, data scientists and other stakeholders.
The challenge in being specific about standardization if further compounded by the relative novelty of the use of IT infrastructures like cloud environments and the applications and use cases for digital data like machine learning and predictive analytics. The rapidity with which these technologies are being adopted by the civil engineering and the transport infrastructure industry is contributing to a rapidly evolving landscape of growing innovations, and the risk of defining possibly overly strict landmarks with premature standardization requirements might risk stifling this progress.
The final recommendations will try to take into account this diversity and effervescence in the field by sidestepping specific solutions and focusing on the other hand on general design considerations aimed at guiding the design and evaluation an IT platform for the industry.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

2 Overview of software platforms for data management in use in the construction sector

2.1 Introduction

The main goals of digitalization for inspections, monitoring and maintenance data are to endow practitioners like civil and structural engineers with the tools to carry out their data-related tasks in a way that is problem-focused, robust and effective. To achieve these goals it is crucial to consider all the different factors involved in the overall workflow of the practitioners that, from the site inspections, leads to obtain processed information and Performance Indicators useful to support the structural assessment. These factors become relevant at different phases of the asset management flow and affect, at various levels, the subjects involved in the process.

As the technology underlying inspection, monitoring and maintenance matures, a serious challenge to be faced during the design of a digital platform is that it must be versatile, so as to be adaptable and targeted to respond to the evolving needs of the broadest possible diversity of users.

In the overall definition of purposes for a digital platform, the ideal approach would be to primarily identify the stakeholders involved and their actual daily requirements and expectations, to fit the functionalities to their needs. This is both a practical consideration as well as the result of sociotechnical analyses that identified the direct involvement of relevant stakeholders as a central tenet of a comprehensive strategy to quantifying and resolving questions around safety and risks involved in the deployment of new technologies (see e.g. (1)).

The following overview of existing data platforms allows us to showcase the current practices, offering a general idea of customized solutions currently available for technical realities that have already faced similar challenges in their working life.

2.2 Case Studies selection

The following overview of existing data platforms in the transport infrastructure industry and current practices offers a general idea of customized solutions currently available. The present chapter aims to provide a rational selection of case studies, that clearly represent a wide set of scenarios including local vs national realities, few vs thousands of assets managed, management vs technical scenarios of needs and different adopted solutions.

Obtained by means of direct interviews with the users of such digital platforms in their own daily work, the following table gives a synthetic description some of the primary European stakeholders in the infrastructure monitoring process, together with the digital platforms they use, with particular focus on the data-related functionalities and requirements characterizing them.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

Stakeholder	Local/National environment	Platform Name	Inspection Data	Structural Health Monitoring (SHM) Data
Sina S.P.A.	Engineering company	SIOS	Ingestion, Storage	*Not handling*
Comune di Torino	Local Municipality	WeBridge	Ingestion, Storage	*Not handling*
Ministero delle Infrastrutture e dei Trasporti (MIT)	Ministry	AINOP	Ingestion, Storage	*Not handling*
Rete Ferroviaria Italiana	Operator	DOMUS	Ingestion, Storage	*Not handling*
Sacertis Ingegneria srl	Engineering company	SACERTIS PLATFORM	*Not handling*	Ingestion, Storage, Processing
Spanish Road Directorate (MITMA)	Ministry	CELOSIA	*Not handling*	Ingestion, Storage
IBM	Sofware Company	OCL PLATFORM	Ingestion, Storage	Ingestion, Storage, Processing
RELATICS	Software Company	Relatics	Storage	*Not handling*
SHM Systems Sp. z o.o., Sp. kom.	Engineering Company	SHM MONITOR	Storage	Ingestion, Storage, Processing

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.3 SIOS – SINA S.p.A. (Società Iniziative Nazionali Autostradali, Italy)

2.3.1 Platform overview

Input Data	Text Documents (PDF) Images Videos
Functionalities	Data ingestion and storage Data annotation Geolocation Analysis and ranking of intervention priorities
Input Users (data feeding user)	Site inspectors Maintenance dep.
Output Users	Structural design dep. Maintenance dep.

SIOS (Sistema Ispezione Opere Sina) is a digital platform, developed by SINA S.p.A. (Società Iniziative Nazionali Autostradali) to be used as intercompany database, in which different types of data are stored; thanks to a web interface, automatically populated in almost real time, it can always be consulted by the users.

The platform has been developed to meet both the requirement to store big quantity of different type of data, accessible on demand in near real time, and to standardise the outcomes of site inspections in order to fit current standards regarding structural defects classification (2); the data stored, belong to more than 900 bridges located on 6 highways in the North of Italy.

The platform has also been intended to allow the communication between inspectors and those in charge of maintenance programs or engineers to perform structural assessment, by using common protocols and tools running on the SIOS.

The users of this platform are mainly:

- Data Feeding Users
  - Inspectors
  - Technical consultants
- End User
  - Technical department of Road operator companies
  - Engineering companies

Data Feeding Users are site inspectors and/or technical experts who have the role to create a digital twin of the infrastructure, in the very first phase, filling a pre-set form to fully characterize the real structure.

In SIOS it is possible to register different kinds of asset, from bridges or pedestrian overpasses to structural steelwork for signs or noise barriers. The analysis of each structure begins with the search for historical data with the evaluation of:

- terrain morphological and environmental data;
- structural data (drawings and reports);
- data relating to the previous operational life (i.e. previous interventions);
- data relating to the current status with evaluation of the structural safety factors.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

The following digitalization of the collected information is made in the following ways:

1) Decomposition into structural units:
the individual structures are divided into units (set of spans) according to the similarities in terms of material, construction methodology, age etc.

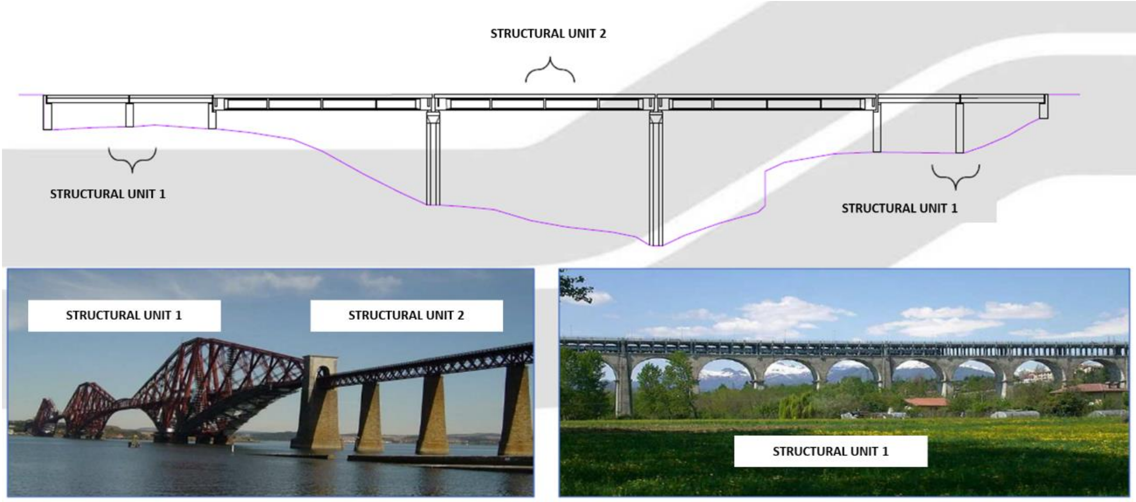


Figure 2.1 - Decomposition into structural units

2) Decomposition of structural units into typological elements:
after highlighting the structural units, for each of them the recurring typological elements are identified; these are grouped by the structural engineer, for a first breakdown of the structure according to the following parameters: material, geometry, construction type, usage class (operating loads).

3) **Identification of single component:**
each single elementary component must be associated with its typological form which it will inherit all previously seen characteristics: in this phase the list of elements is created. Each elementary component may vary slightly from the typological i.e. in terms of shape (Area), while it maintains all the other characteristics (material, list of anomalies, manufacturer, etc.).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

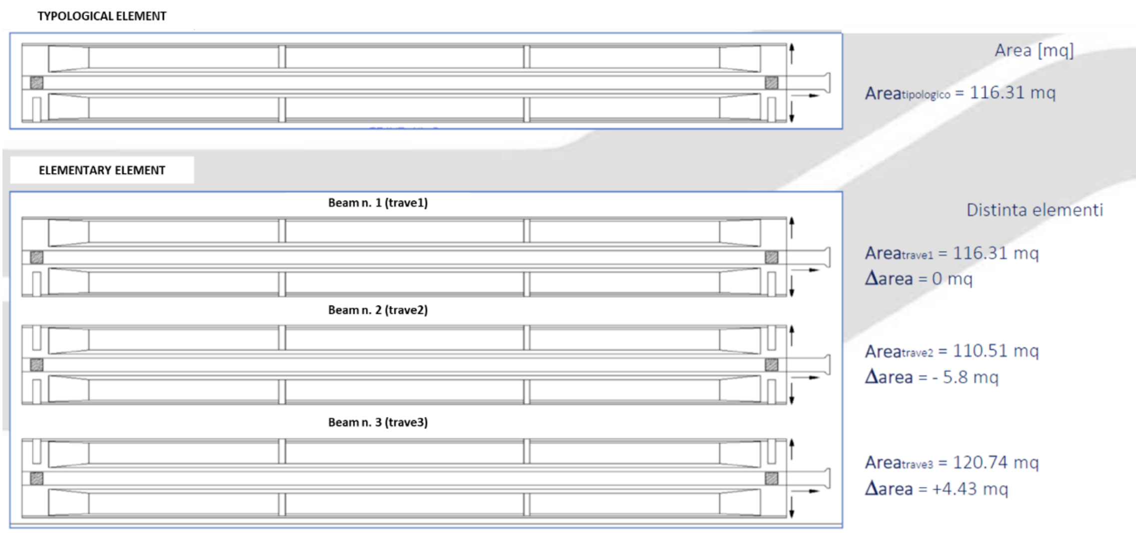


Figure 2.2 - Elementary Element

4) Link between each elementary component with:

- Structural defects (percentage of extension and location by site inspection)
- Monitoring parameters to be referred to

At the end, the SIOS software calculates the actual area of the anomalies per each element and correlates the results with a value index that takes into account the intensity and extent of the anomaly.

Once all the documentation concerning the structure has been collected and the digitalization has been performed, it is possible to carry out the visual inspection; inspection methodologies are mainly two: with graphic rendering of defects/ anomalies or with a percentage estimate of them. The principal aim of this first visual inspection is the detailed mapping of the degradation status of each elementary element.

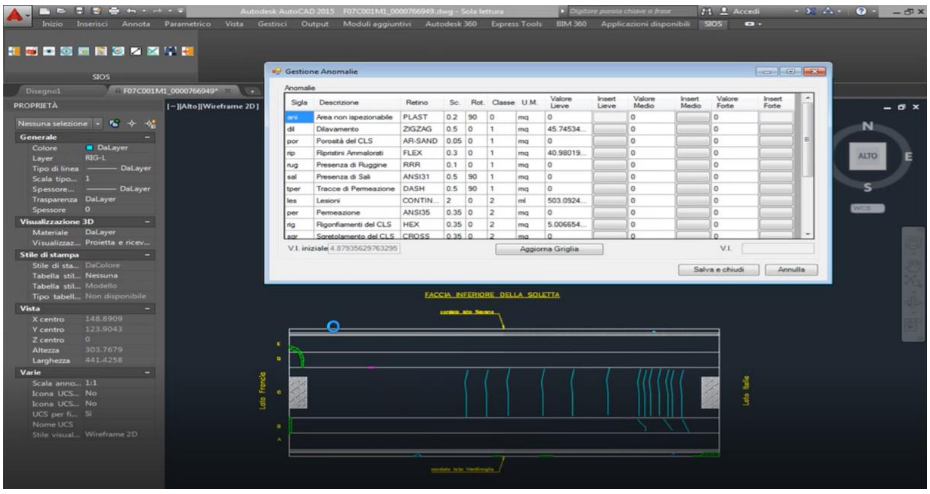


Figure 2.3 - CAD representation of anomalies

2.3.2 Case study: surveillance visual inspection for structural assessment

The main inspections described in the previous section are the basis for follow-up SURVEILLANCE visual inspections which are scheduled periodically (every 3 months). These are aimed at an assessment of the evolutionary status of the main anomalies identified in the previous visual inspections, in addition to checking the previously recorded reports. When the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

inspector triggers an alert for a particular anomaly, SIOS highlights it to the responsible engineer, who must approve it. Once approved, the anomaly is also highlighted on the web system and can be immediately consulted by final users.

The anomalies are the same as in the MIT (Ministero delle Infrastrutture e Trasporti - Italian Ministry of Transport) defect catalog associated with the guidelines. In SIOS there are cards that allow the inspector to understand what are the anomalies that can occur on a specific typological element.

Figure 2.4 - Defects cards SINA/MIT

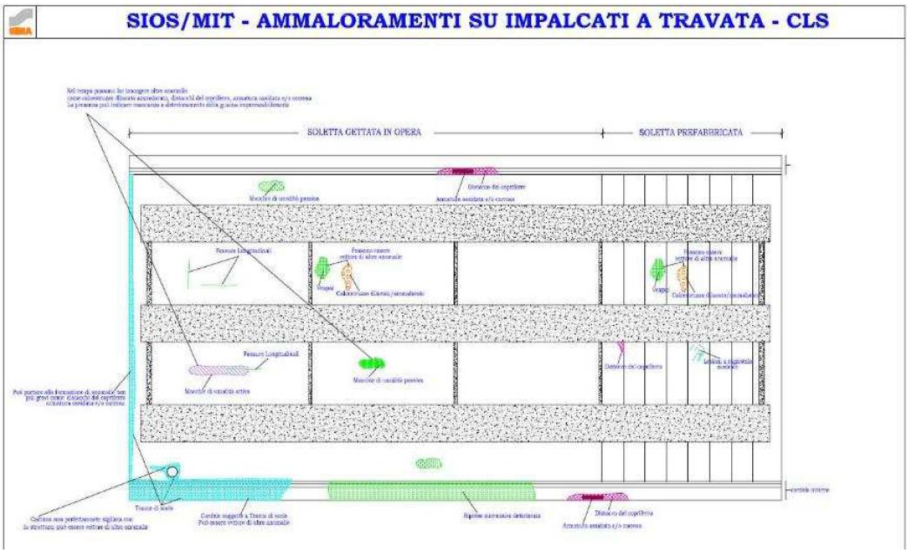


Figure 2.5 - Graphic scheme of a deck: representation of possible anomalies

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

2.3.3 User interface and user experience

The site inspection is supported by a dedicated android application: in the SIOS web interface (LAISAP) the inspector can access all previous collected data and alerts highlighted for the specific structure and can upload new anomalies he eventually finds.

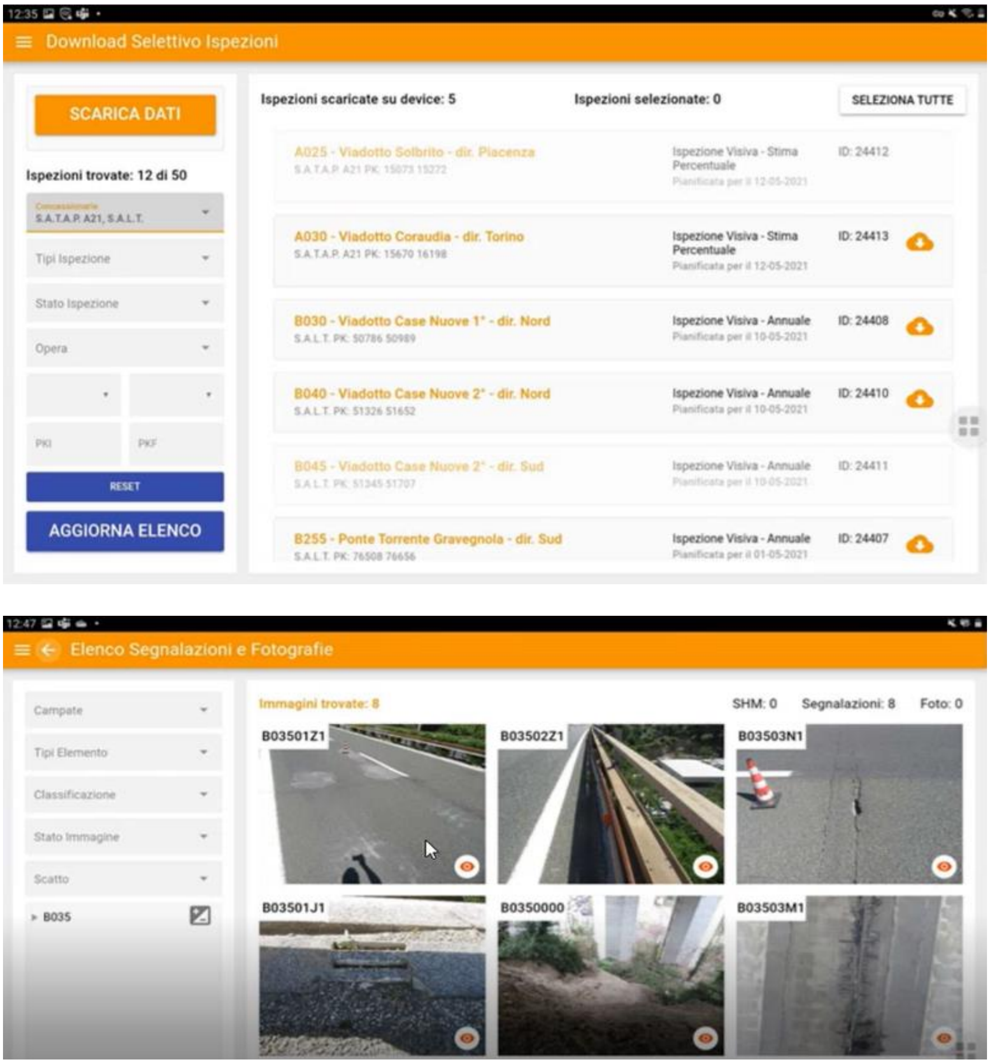


Figure 2.6 - Web interface of SIOS - List of viaducts/bridges and their pictures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

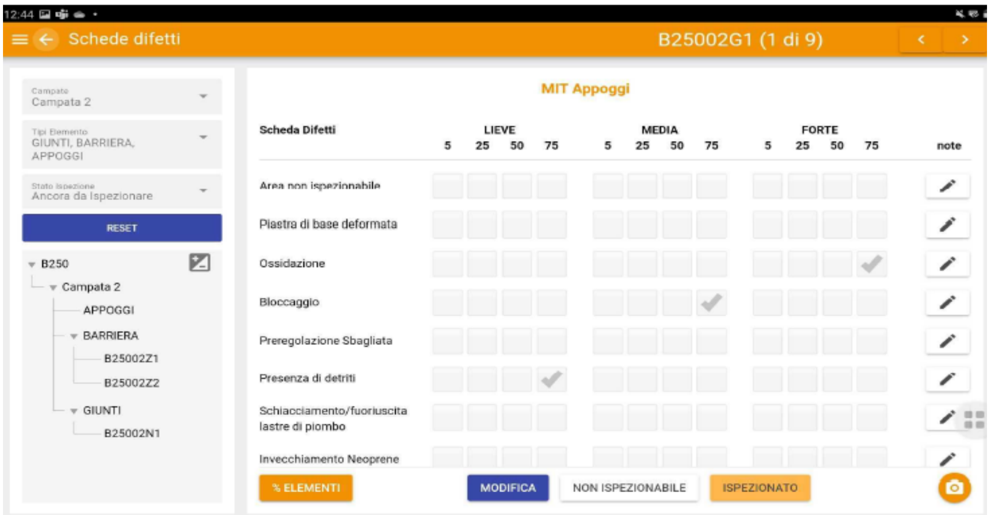


Figure 2.7 - Defect cards for an elementary element with extensions

SIOS gives users the possibility to download synthetic documents, such us: evolution of the average index value by type of element; photographic documentation; reports; data concerning tests conducted, test certificates.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

2.4 WeBridge – Comune di Torino (Italy)

2.4.1 Platform overview

Input Data	Text Documents (PDF) Images Videos
Functionalities	Data ingestion and storage Data annotation Geolocation Analysis and ranking of intervention priorities
Input Users (data feeding user)	Site inspectors
Output Users	Asset owners Asset operators

WeBridge is a software tool for the automated management of bridges and viaducts, developed by 4 EMME Service SpA. Italian Administrations use WeBridge to store their data and to be able to consult them at any time.

The software provides to Administrations a valuable support in the decision-making process that is the base for bridges and the infrastructure network management, and it helps to build up effective ordinary and extraordinary maintenance programs. Furthermore, the WeBridge platform can be integrated with the public works database, AINOP (see paragraph 2.7).

Through the WeBridge platform it is possible to carry out a census of the infrastructures to digitalize inspections directly on site by filling in the appropriate forms. The inspectors are qualified experts who have the task of inspecting the structure on site and completing, based on the results of their observations, the pre-set evaluation forms. The inspector, then, deals with the digitization of the structures and the assignment of scores to represent in a quantitative way the state of each component of the structure (with special attention to the defects features: typologies, extension, location). The responsible structural engineer, after the inspector has filled the forms, must validate the data provided, homogenizing the various assessments. Based on these scores, the software returns a global index that represent the status of the overall structure.

2.4.2 Case study: public infrastructures management

The Municipality of Turin (Italy) uses this software to support the management of about 250 infrastructure objects, divided into: bridges, overpasses, underpasses and cycle-pedestrian walkways.

At the end of each year, based on the indices returned by the software, a ranking of priorities in terms of interventions and maintenance strategies is drawn up, according to the available funds.

2.4.3 User interface and user experienced

The frontend of WeBridge consists of an initial panel that offers users an overview of the listed facilities and their inspection status.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Figure 2.8 - Home page.

It is also possible to view the listed structures on a map and, for each single structure, the results of the inspections carried out during its life.

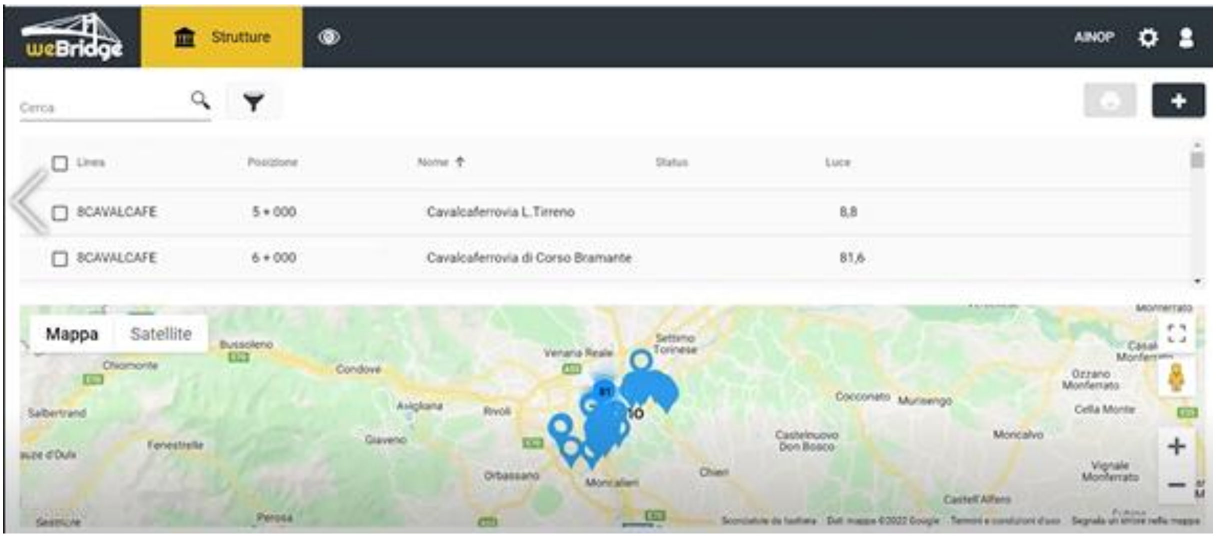


Figure 2.9 - Map of listed structures.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

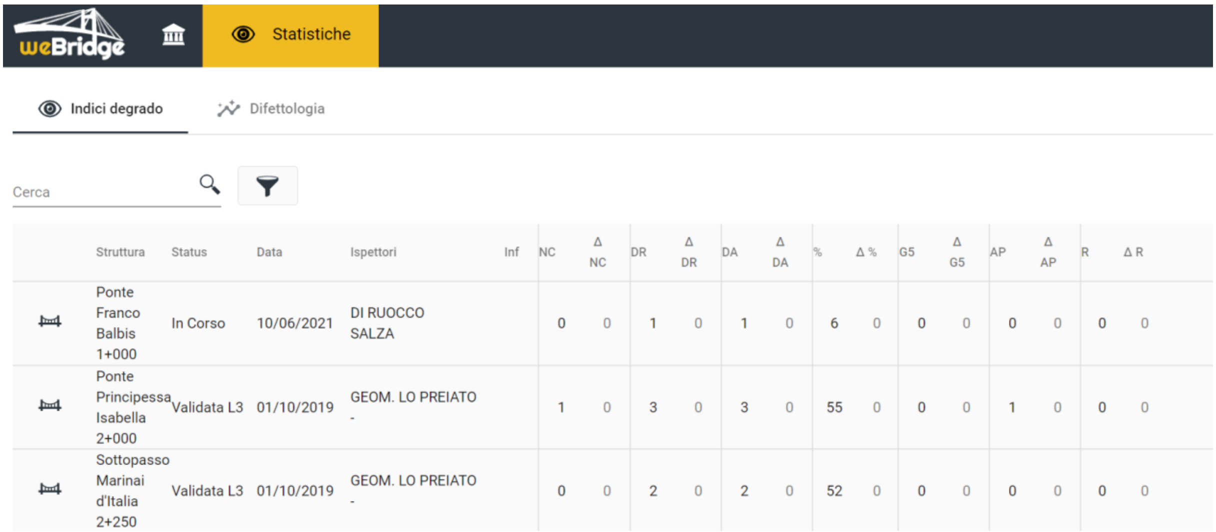


Figure 2.10 - Records of the inspections carried out

The data entry and the filling out of the evaluation forms take place through different levels:

- Level 0: *Geolocation / census.* The Census is carried out by the inspector using an Acquisition Form:
  - Identification: road / railway line, progressive kilometre, name, etc.
  - Administrative data: owner, road operator, locality, etc.
  - Geometric data: total length, deck width, type of spans and piles, etc.
  - Structural data: foundations, piers, beams, joints, supports, etc.
  - Service data: road class, load limitations, etc.
    Level 0 also offers the possibility to insert photos of the structure.

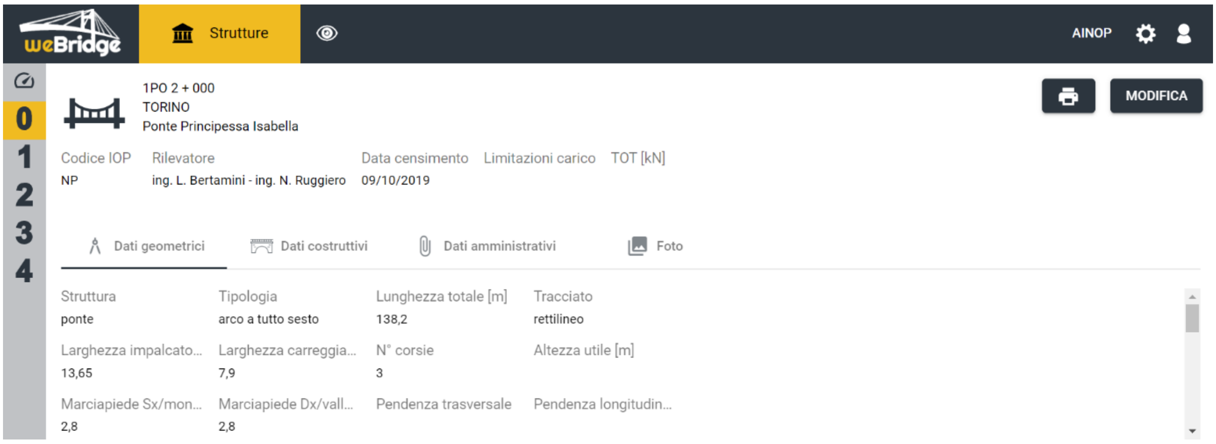


Figure 2.11 - Geolocation and census.

- Level 1: *Inspections and defect sheets*. This level includes all the information obtained during the visual inspections through the compilation of defect cards for the various structural elements. Visual Inspections must be carried out at regular temporal intervals (i.e. 6 months) and cover all components of each infrastructure. The inspector, during the site survey, can also upload the photographs of the specific defect, to store the images in the common database for future comparison.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

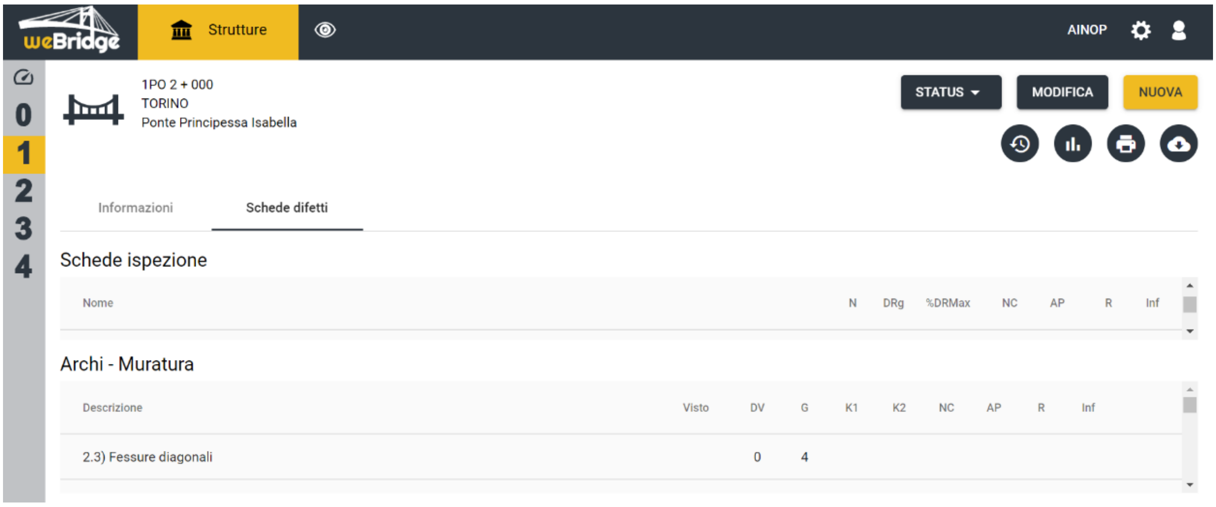
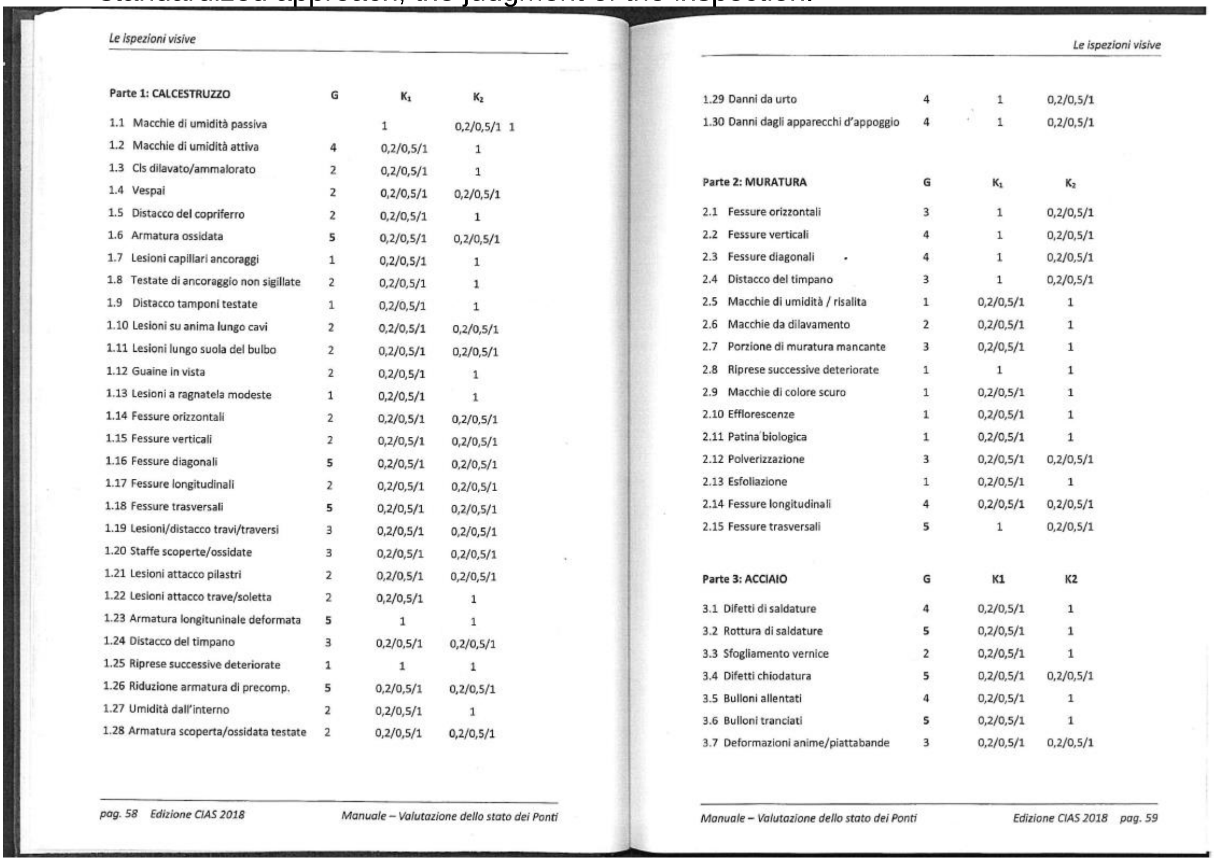


Figure 2.12 - Defect cards.

Inspectors during the site survey, use technical defects sheets, to support, with a standardized approach, the judgment of the inspection.



Le ispezioni visive

Parte 1: CALCESTRUZZO	G	K1	K2
1.1 Macchie di umidità passiva		1	0,2/0,5/1 1
1.2 Macchie di umidità attiva	4	0,2/0,5/1	1
1.3 Cls dilavato/ammalorato	2	0,2/0,5/1	1
1.4 Vespai	2	0,2/0,5/1	0,2/0,5/1
1.5 Distacco del copriferro	2	0,2/0,5/1	1
1.6 Armatura ossidata	5	0,2/0,5/1	0,2/0,5/1
1.7 Lesioni capillari ancoraggi	1	0,2/0,5/1	1
1.8 Testate di ancoraggio non sigillate	2	0,2/0,5/1	1
1.9 Distacco tamponi testate	1	0,2/0,5/1	1
1.10 Lesioni su anima lungo cavi	2	0,2/0,5/1	0,2/0,5/1
1.11 Lesioni lungo suola del bulbo	2	0,2/0,5/1	0,2/0,5/1
1.12 Guaine in vista	2	0,2/0,5/1	1
1.13 Lesioni a ragnatela modeste	1	0,2/0,5/1	1
1.14 Fessure orizzontali	2	0,2/0,5/1	0,2/0,5/1
1.15 Fessure verticali	2	0,2/0,5/1	0,2/0,5/1
1.16 Fessure diagonali	5	0,2/0,5/1	0,2/0,5/1
1.17 Fessure longitudinali	2	0,2/0,5/1	0,2/0,5/1
1.18 Fessure trasversali	5	0,2/0,5/1	0,2/0,5/1
1.19 Lesioni/distacco travi/traversi	3	0,2/0,5/1	0,2/0,5/1
1.20 Staffe scoperte/ossidate	3	0,2/0,5/1	0,2/0,5/1
1.21 Lesioni attacco pilastri	2	0,2/0,5/1	0,2/0,5/1
1.22 Lesioni attacco trave/soletta	2	0,2/0,5/1	1
1.23 Armatura longituninale deformata	5	1	1
1.24 Distacco del timpano	3	0,2/0,5/1	0,2/0,5/1
1.25 Riprese successive deteriorate	1	1	1
1.26 Riduzione armatura di precomp.	5	0,2/0,5/1	0,2/0,5/1
1.27 Umidità dall'interno	2	0,2/0,5/1	1
1.28 Armatura scoperta/ossidata testate	2	0,2/0,5/1	0,2/0,5/1

pag. 58 Edizione CIAS 2018 Manuale – Valutazione dello stato dei Ponti

Le ispezioni visive

1.29 Danni da urto	4	1	0,2/0,5/1
1.30 Danni dagli apparecchi d'appoggio	4	1	0,2/0,5/1

Parte 2: MURATURA	G	K1	K2
2.1 Fessure orizzontali	3	1	0,2/0,5/1
2.2 Fessure verticali	4	1	0,2/0,5/1
2.3 Fessure diagonali	4	1	0,2/0,5/1
2.4 Distacco del timpano	3	1	0,2/0,5/1
2.5 Macchie di umidità / risalita	1	0,2/0,5/1	1
2.6 Macchie da dilavamento	2	0,2/0,5/1	1
2.7 Porzione di muratura mancante	3	0,2/0,5/1	1
2.8 Riprese successive deteriorate	1	1	1
2.9 Macchie di colore scuro	1	0,2/0,5/1	1
2.10 Efflorescenze	1	0,2/0,5/1	1
2.11 Patina biologica	1	0,2/0,5/1	1
2.12 Polverizzazione	3	0,2/0,5/1	0,2/0,5/1
2.13 Esfoliazione	1	0,2/0,5/1	1
2.14 Fessure longitudinali	4	0,2/0,5/1	0,2/0,5/1
2.15 Fessure trasversali	5	1	0,2/0,5/1

Parte 3: ACCIAIO	G	K1	K2
3.1 Difetti di saldature	4	0,2/0,5/1	1
3.2 Rottura di saldature	5	0,2/0,5/1	1
3.3 Sfogliamento vernice	2	0,2/0,5/1	1
3.4 Difetti chiodatura	5	0,2/0,5/1	0,2/0,5/1
3.5 Bulloni allentati	4	0,2/0,5/1	1
3.6 Bulloni tranciati	5	0,2/0,5/1	1
3.7 Deformazioni anime/piattabande	3	0,2/0,5/1	0,2/0,5/1

Manuale – Valutazione dello stato dei Ponti Edizione CIAS 2018 pag. 59

Figure 2.13 - Inspection types

- Level 2: *Attention classes.* Level 2 allows inspectors to enter data relating to environmental conditions, landslides and /or hydraulic data. The WeBridge software, based on all information recorded in the previously panels, automatically computes the Attention Classes and common ordinary maintenance interventions are suggested for each asset, based on expertise on similar issues.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

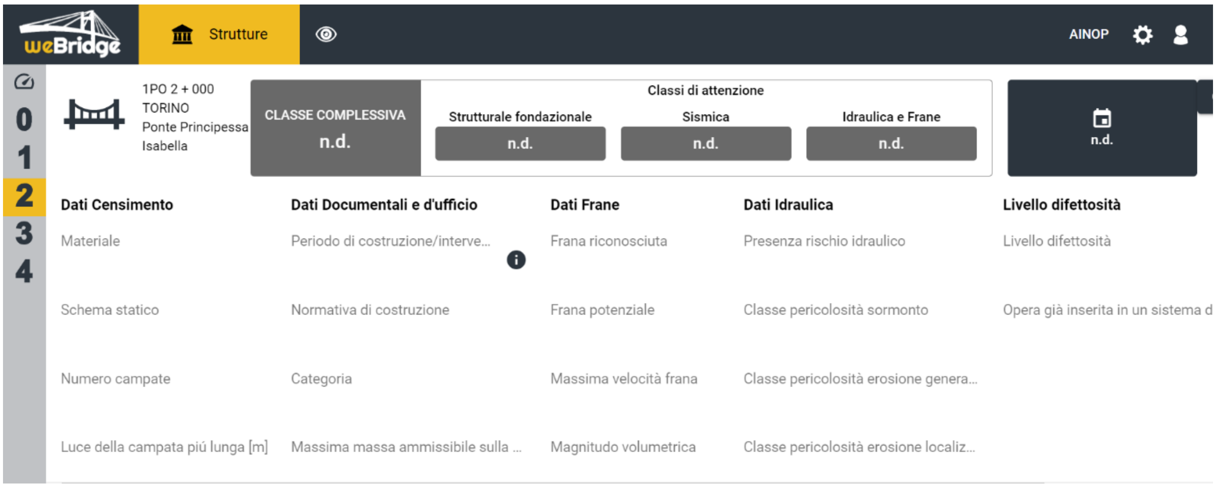


Figure 2.14 - Environmental condition form.

- Level 3: *Preliminary analyses*. Medium or Medium-High attention classes require preliminary analysis to estimate the structural performance; often a numerical comparison between the residual capacity calculated with technical standards used at the time of the design of the structure and the actual standards is carried out. For the structures belonging to these classes, extraordinary inspections are planned.
- Level 4: *Accurate analyses*. Infrastructures in High Attention Classes require individual control strategies through continuous monitoring systems supported by numerical finite element models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.5 DOMUS – RFI (Rete Ferroviaria Italiana) (Italy)

2.5.1 Platform overview

Input Data	Text Documents (PDF) Images Videos
Functionalities	Data ingestion and storage Data annotation 3D rendering and navigation Analysis and ranking of intervention priorities Management and planning of interventions
Input Users (data feeding user)	Site inspectors Maintenance department
Output Users	Structural design dep. Maintenance dep.

RFI (Rete Ferroviaria Italiana is the company of the Ferrovie dello Stato group, responsible for the overall management of the national railway network in Italy) has developed a proprietary BMS (Bridge Management System) called DOMUS which allows, through a 3D navigation of the various components of the infrastructure, the registration of any existing defect, classified according to a RFI catalog of defects. This digital platform represents a support to the inspectors in charge of the survey (3,4).

DOMUS is the acronym for “Diagnostica Opere d’arte Manutenzione Unificata Standard”; it is a system to digitalize infrastructures and support decision making process on ordinary and extraordinary maintenance. The objectives of the DOMUS system are:

- Ranking intervention priorities in an objective way (numerical judgment code), making sure that the judgment criteria are homogeneous throughout the national territory
- Optimizing the planning process of maintenance interventions (management), with minimization of reactive interventions (emergency)
- Sharing information, results and judgments.

The diagnostic process based on the use of visual inspection is supported, if necessary, by instrumental surveys and investigations and by the use of innovative technologies (drones, monitoring systems, etc.). At the end of the process, the analyses lead to an overall assessment of priorities with the development of intervention projects.

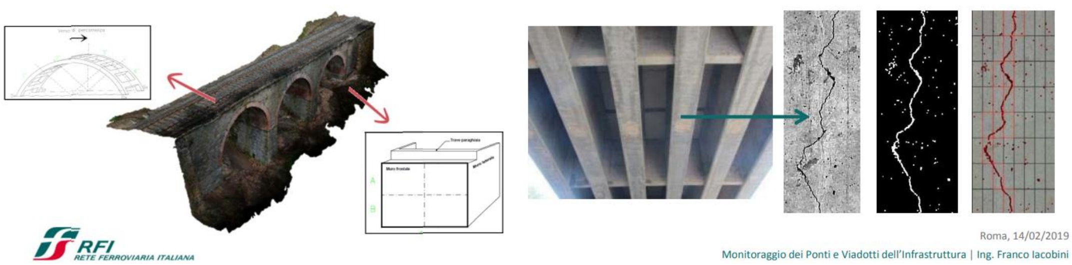


Figure 2.15 - Defects digitalization

The digitalization of each asset is made during the first site inspection, when the inspector, by using a combination of pre-selected groups of structural elements (pillars, girder deck, arch,

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171.

joints etc) and materials, fully describe the structure on DOMUS. Inspections, when needed, are supported with the help of special means (by- bridge and elevating platforms) to detail every component of the structure; for the inspection of spans with difficulties of access, an pilot activity is currently ongoing on the use of drones are used for the 3D photographic survey that are measurable and interfaceable with the DOMUS (5).

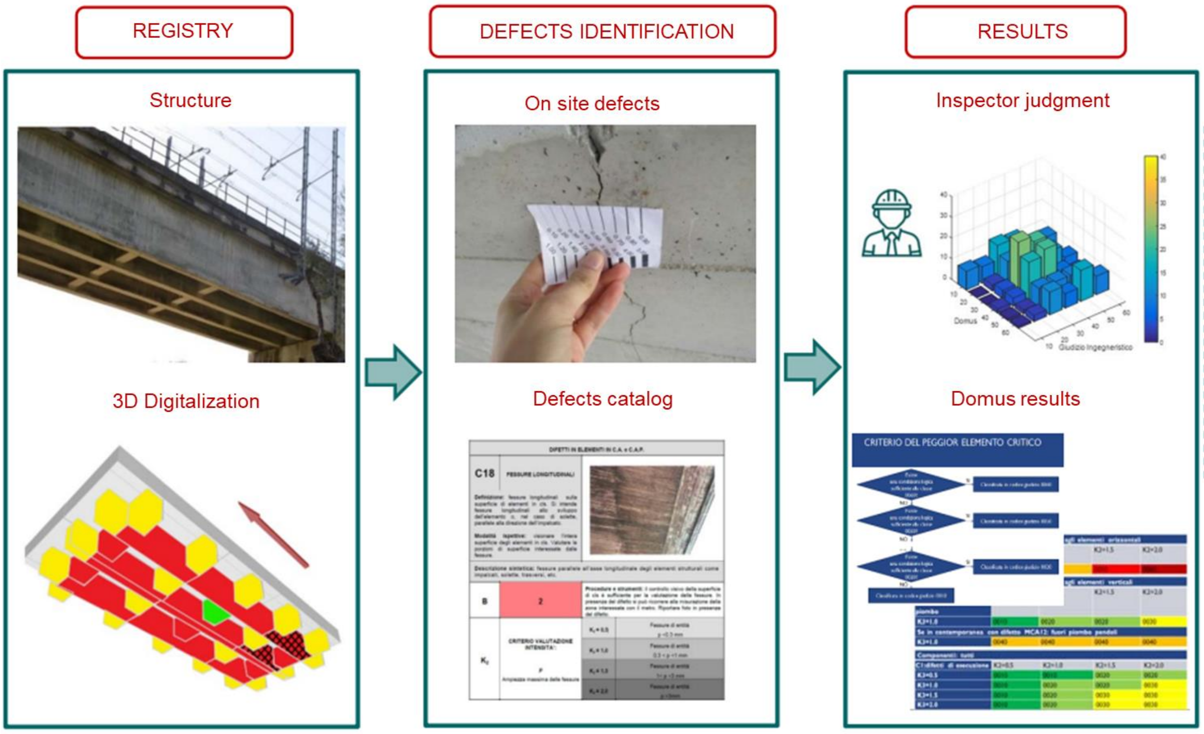


Figure 2.16 - DOMUS workflow

Domus consists in three main sub-systems (3):

1) inventory system
2) inspection system
3) evaluation system

1) **The Inventory system**. It contains the census of the overall bridges national network; each asset is described in the company IT Database (called INRETE 2000) with a Technical Sheet, where its position and function are uniquely determined. In each technical sheet it is possible to identify, in a hierarchical way, the components of each element of infrastructure. The inspector can:

- proceed to verify the data already existing in the platform INRETE 2000
- update information (it is possible to insert defects for each single component)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

Figure 2.17 - DOMUS panel for the Inventory section

2) **The Inspection System**. With these functionalities, defects are recorded via a tablet on which the DOMUS software is installed. With the aim of standardizing the procedure for recognizing defects, a Defects Catalog has been implemented, through which the inspector is able to describe the intensity, the extension and location of the defects in a unique way.

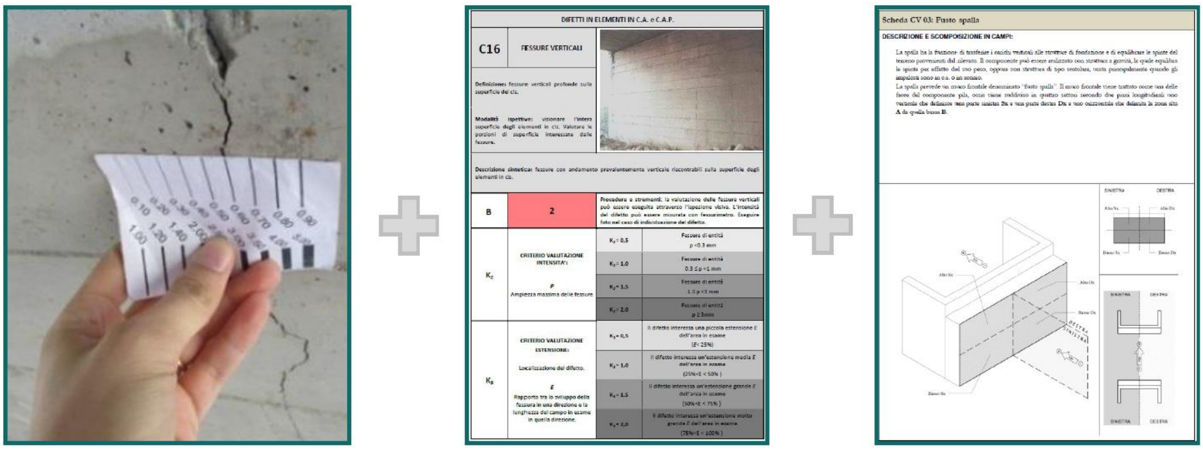


Figure 2.18 - DOMUS panel for the Inspection section

3) **The Evaluation System**. Starting from the data recorded during the visual inspection, through the DOMUS Mobile application, in which specific algorithms are implemented, it is possible to evaluate a series of indexes representative of the overall state of the structure. These indexes constitute a support information for expressing the final judgement on the efficiency of the structure and for managing their maintenance, facilitating the definition of intervention priorities. In particular, the algorithm generates the following three indices:

- **DOMUS Index**: it represents the state of deterioration and supports the handling of its maintenance
- **DOMUS Index 2.0**: it supports the inspector in defining the judgment code regarding the efficiency of the infrastructure with regard to the safety of railway traffic
- **Prevision index**: it represents the evolution of the state of the defects registered during the survey, and it allows a prediction concerning the judgment code of the next inspection.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

2.6 Sacertis Platform – Sacertis Ingegneria S.R.L. (Italy)

Input Data	Images Geolocation data
Functionalities	Data ingestion and storage Data visualization Data preprocessing
Users	Data annotators Data scientists Machine learning engineers Structural engineers

Sacertis Ingegneria is a highly innovative engineering company active in the field of Structural Health Monitoring, providing an affordable and reliable real time Monitoring and Diagnostics System, through a smart combination of civil engineering knowledge, advanced Cloud computing, easy-to-install sensors and SHM systems. Sacertis monitors in real time a network of more than 50 infrastructures and has developed internally its own hardware and software tools; its technological team designs customized applications/services/systems for monitoring infrastructures to fit structural engineering needs.

Sacertis digital platform uses the solution proposed by IBM Cloud, IBM's high-performance public cloud platform, which provides cloud computing services.

Figure 2.19 shows a simplified block diagram of the SW components that creates the Sacertis cloud architecture. The IBM proprietary services are represented in yellow and green, while the applications developed by Sacertis are symbolized in blue / light blue.

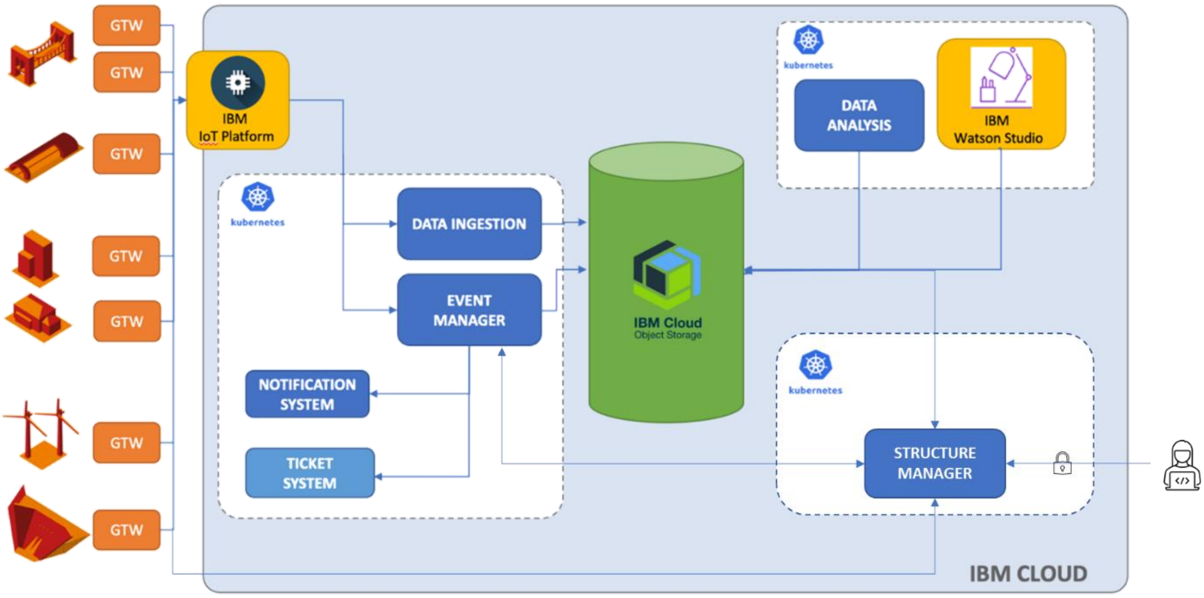


Figure 2.19 - Sacertis Cloud Architecture

The Sacertis cloud services architecture maintains the logical separation of each Customer data. Access to Customers data is allowed only to authorized personnel, strictly controlled by identity and access management policies, and monitored in accordance with Sacertis' internal privileged user control program.

A simplified description of the main components is presented as follows:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

- The IBM IoT Platform service is used to register and connect GTWs (Local Control Unit of the monitoring systems) to the cloud and receive or send data securely, using the open and lightweight MQTT messaging protocol with TLS.
- Data Ingestion application, developed by Sacertis, takes care of receiving the data from each monitored structures, through a secure API of the IBM IoT Platform, and storing them in a compressed format in the IBM Object Storage, without carrying out any modification or processing on the raw data, therefore without affecting the integrity of the data.
- IBM Object Storage provides durable storage with comprehensive security measures. The distribution capabilities in different geographical areas provide backup and data protection, which always becomes accessible, even in cases where disastrous events or multiple failures occur.
- Event Manager is an application developed by Sacertis able to perform a semi-real time continuous monitoring. It checks the status of the overall system, sensors and GTWs (control units), and handles anomalies due to data processing algorithms, both in the cloud (Data Analysis) and locally on the gateways. Based on the communication protocols, agreed between Sacertis and the Customer, the Event Manager activates the notification and ticket systems, depending on the type of event detected. This application also has interactions only within the Sacertis cloud, therefore protected by IBM security policies.
- Notification system - Ticket system. The notification and ticket systems are Sacertis proprietary applications capable of managing the communication protocols to the Customer, following structural or system anomalies detected and reported by the Event Manager. The notifications can be phone calls, PEC or e-mail and for each of them an external provider is used, with which the applications communicate via secure API.
- Data analysis is an application developed by Sacertis that periodically executes the data processing algorithms. Raw data from IBM Object Storage are processed, without being overwritten, and in the event of anomalies detected, an alert message is generated that the Event Manager will process. The data resulting from the processing is then also stored on the IBM Object Storage.
- IBM Watson Studio service enables data scientists, developers and analysts to build and manage machine learning models and optimize data processing algorithms. The access credentials are managed by the IBM IAM and the raw data of the structures are never overwritten, but only read. In doing so, the raw data, the processed data and the algorithms never leave the IBM platform, making the whole process safe and secure. When the tested algorithm is ready to be put into production it is loaded into the Data Analysis application seen above.
- Sacertis Ingegneria developed a centralized tool to handle all the installations named Structure Manager.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

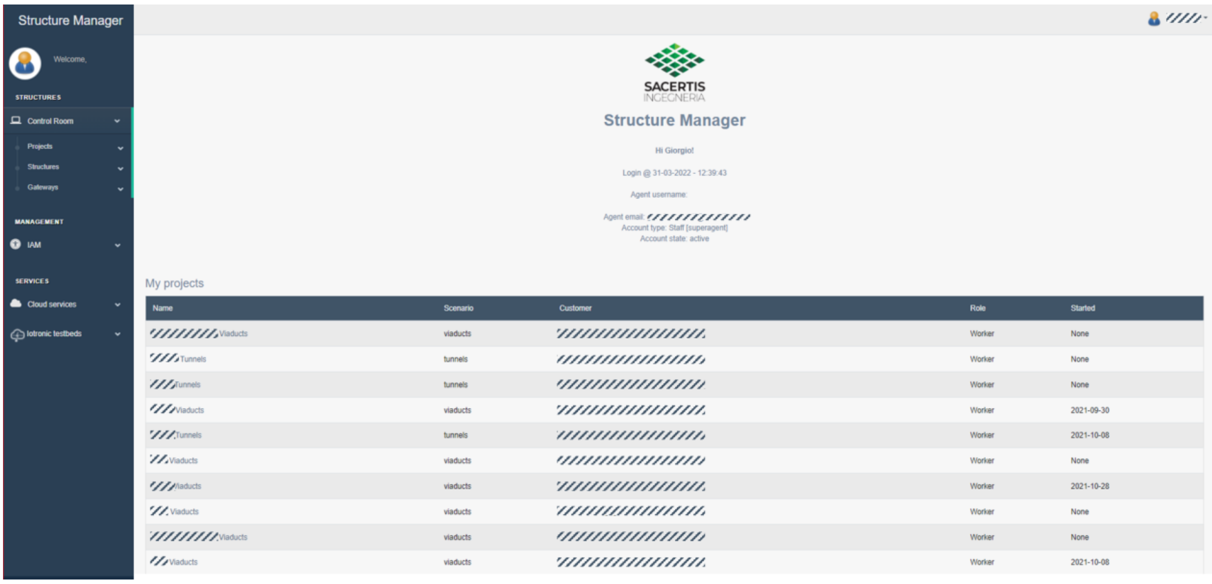


Figure 2.20 - Structure Manager-home page

Local control units (defined as “gateways”) are grouped into “structures” to which they belong, such that every structure can be assigned multiple gateways. The Structure Manager has been developed as a web-based tool to monitor health, life cycle of the asset and handle the functional testing of the monitoring system installed.

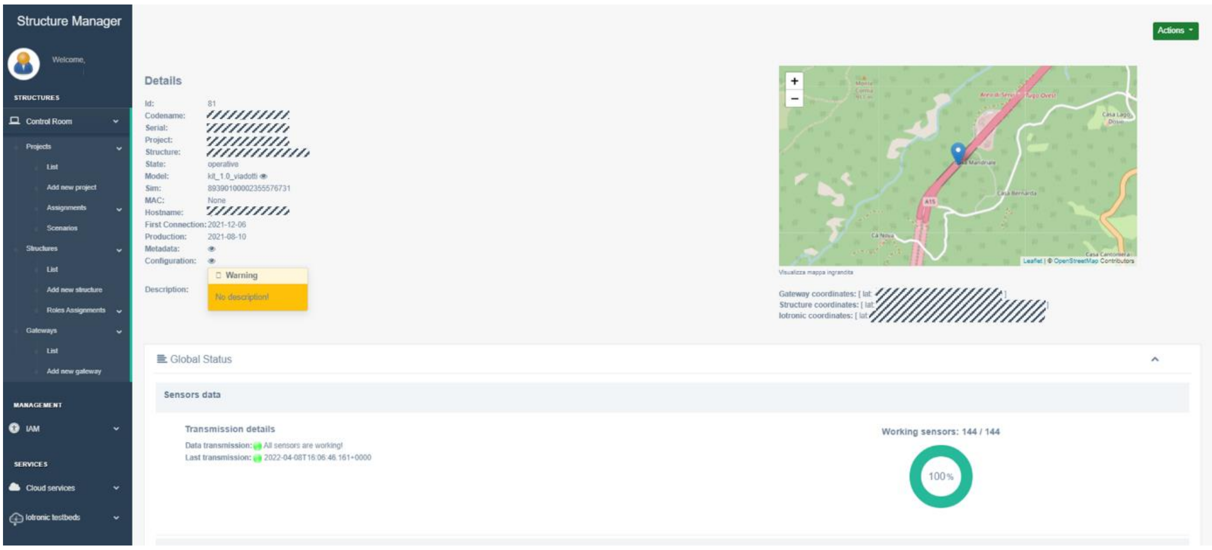


Figure 2.21 - Structure Manager- single structure panel

Geo-localization of structures is also displayed on maps and the layout of sensors on the monitored infrastructure is accessible by an integrated topology app which renders, not only the static disposition, but also the status of each sensor to easy investigate and support any maintenance action on the field.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171.

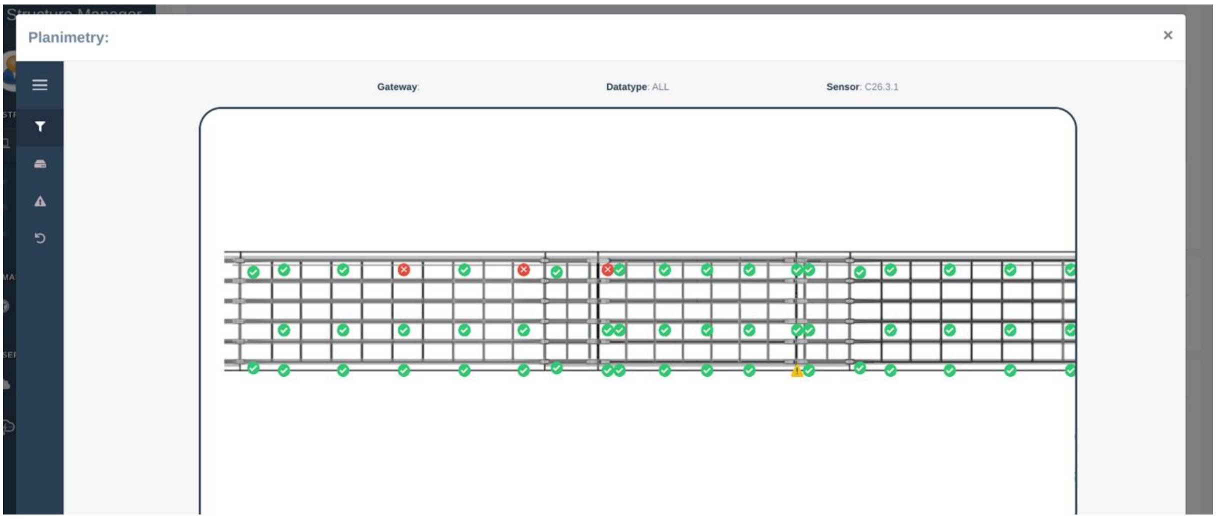


Figure 2.22 - Structure Manager- viaduct monitoring system status

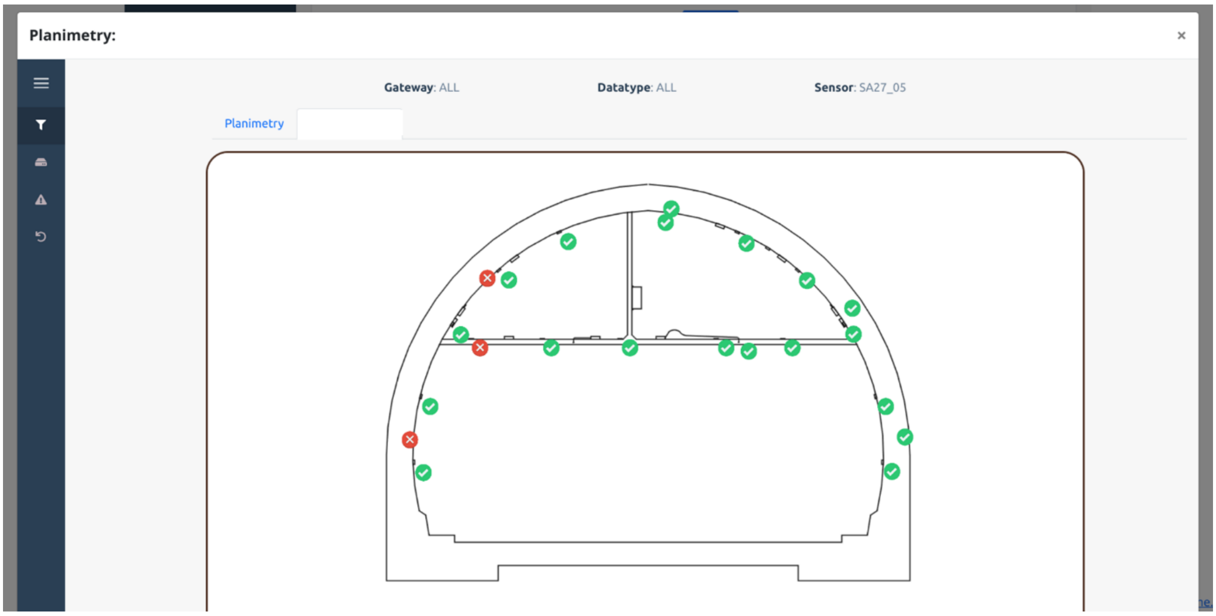


Figure 2.23 - Structure Manager- tunnel monitoring system status

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

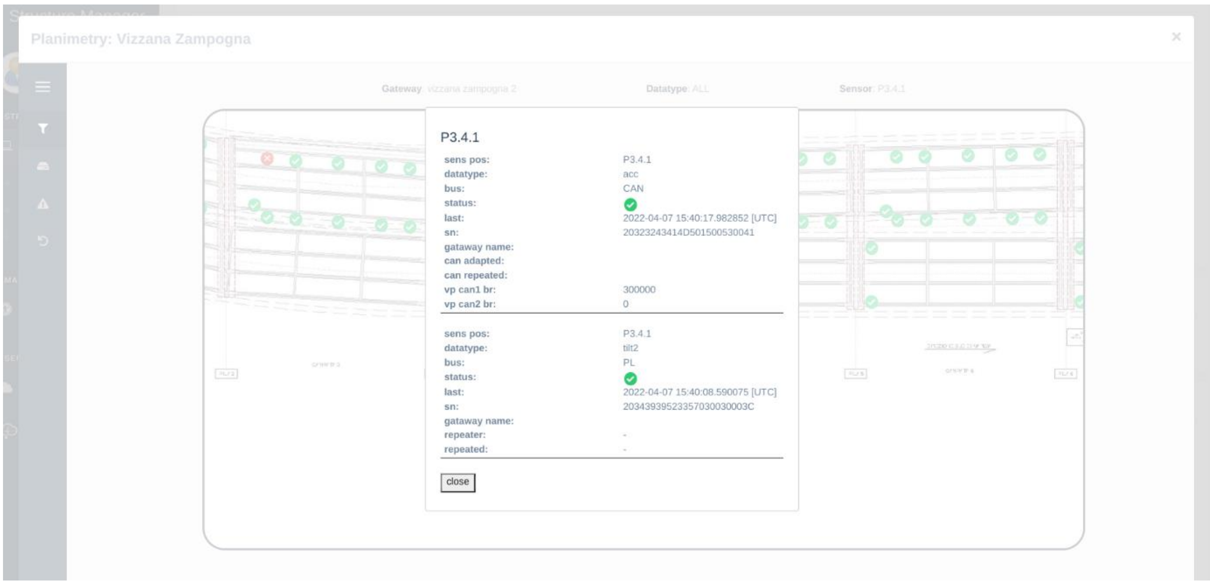


Figure 2.24 - Structure Manager- single sensor status

Furthermore, the Structure Manager can verify if a host (gateway) is online or offline.

The tool permits, to different type of actors, to interact with structures through role-based accounting system: a specific “role” must be assigned to every user, to provide him a suite of actions he can do on a subset of specific structures or gateways, ranging from register instances on DB, handle alarms, perform operations on plugins and services running on gateways’ control unit, up to access remotely and start to verify system health and basic functionalities. External customers can also access as simple viewers to monitor the status of their own monitoring systems.

Structure Manager answers to these requirements:

- Ease of use for non-expertise users
- Clarity to manage system crashes or critical situations
- Design to convey all the operations to minimize interactions with external tools.

To accomplish those requirements, all the elements managed by the Structure Manager are organized in separate web pages with all the operations exposed on specific sections.

Furthermore, real-time integration with external services will be granted using REST APIs to query both remote databases, perform operations on a subset of gateways or one, retrieve information to expose to the user, when a structure or a gateway page will be loaded.

Structure monitoring is also facilitated by an intuitive, easy-browsable UI, using icons of different shape and colour to give feedback about the current state of a service: running or failure.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.7 AINOP – MIT (Ministero delle Infrastrutture e Trasporti, Italy)

2.7.1 Platform overview

Input Data	Text Documents (PDF) Images Videos
Functionalities	Data ingestion and storage Data annotation
Input Users (data feeding user)	Asset owners Asset operators
Output Users	MIMS (Infrastructure and Sustainable Mobility Authority)

In Italy, in November 2018, the National IT Database of Public Infrastructures (AINOP, Archivio Informatico Nazionale delle Opere Pubbliche) (6), was set up by the Infrastructure and Sustainable Mobility Authority. Through the AINOP platform, it is intended to take a census of the patrimony of public assets belonging to the State, Regions, local autonomies (Autonomous Provinces, Provinces, Metropolitan Cities, etc.) and all Municipalities in the national territory. It provides a hierarchical safety list of the assets that need structural interventions, also classifying them according to urgency priorities.

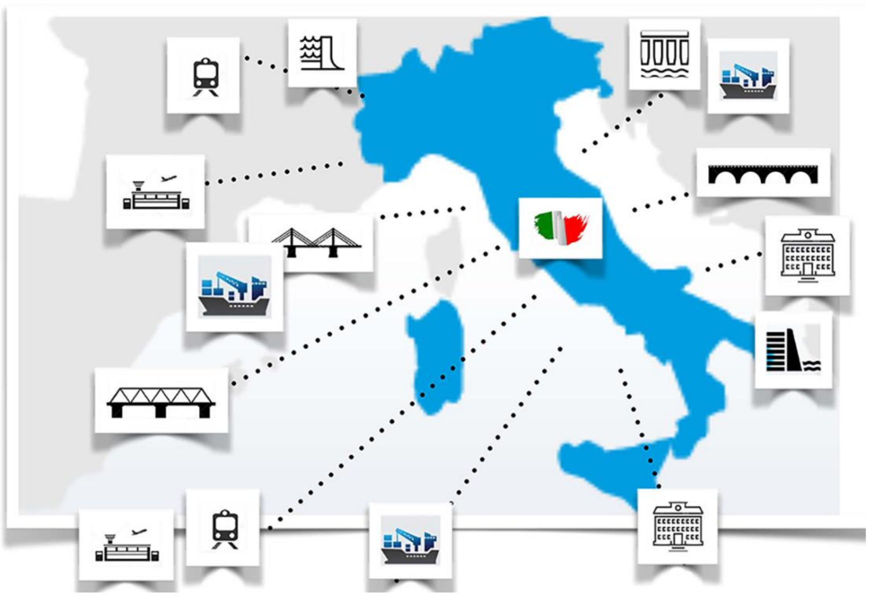


Figure 2.25 - AINOP take a census of the patrimony of public assets over the whole Italian territory.

The AINOP is structured in 9 sections:

- Road Bridges, viaducts and overpasses
- Railway Bridges, viaducts and overpasses
- Roads
- National and regional railways – metropolitan
- Airports
- Dams and aqueducts

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

- Railway tunnels and road tunnels
- Ports and infrastructures
- Public building.

Each section is divided into subsections to fully identify the asset: card registry, technical data, economic-financial data, technical monitoring of the work, maintenance interventions, work in progress, data relating to the territorial context, photographic documentation, environmental data and reports.

AINOP is a platform that enables to:

- identify an asset and its location
- consult data, information and documents of the structure
- receive information that allow the monitoring of the structure, with the aim to prevent critical issues, including smart alert systems on the state of the infrastructure
- identify potential workflows to make efficient maintenance and management phases.

To each asset is assigned a "Public Asset Identification" code called “IOP (Identificativo Opera Pubblica)”, which uniquely identifies the structure itself. The IOP is unique for the entire life of the asset, and it is automatically generated by an algorithm that processes the essential and distinctive features of the structure itself.

A metadata system is associated with the IOP which contains the qualifying information of the asset, such as type and name of the infrastructure, identification code of the operator, geographical and temporal characterization, data of the manager, of the supervisor etc..

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.8 IBM ONE CLICK LEARNING PLATFORM (OCL)

2.8.1 Platform overview

Input Data	Images COCO datasets (7)
Functionalities	Data ingestion and storage Data visualization Data preprocessing Training and deployment of machine learning models
Users	Data annotators Data scientists Machine learning engineers Structural engineers

The IBM One-Click Learning (OCL) platform is an ongoing effort internal at IBM Research aimed at providing end-to-end support for the full data science process, with particular focus on visual inspection of transport infrastructure with the goal of progressively incorporating a growing variety of application scenarios.

This section will focus on the functionalities of the IBM OCL platform that address some of the challenges arising in Big Data and Data Analytics in general, and within the use specifications of inspection and structural health monitoring (SHM) of transport infrastructures in particular.

In the final section of this platform overview the reasons behind the design choices are presented, translating them into recommendations to enable a unified standardized treatment of generic Machine Learning use cases and workflows with particular focus on the civil infrastructure domain. Many of the considerations in the present section will also be published in the paper (8).

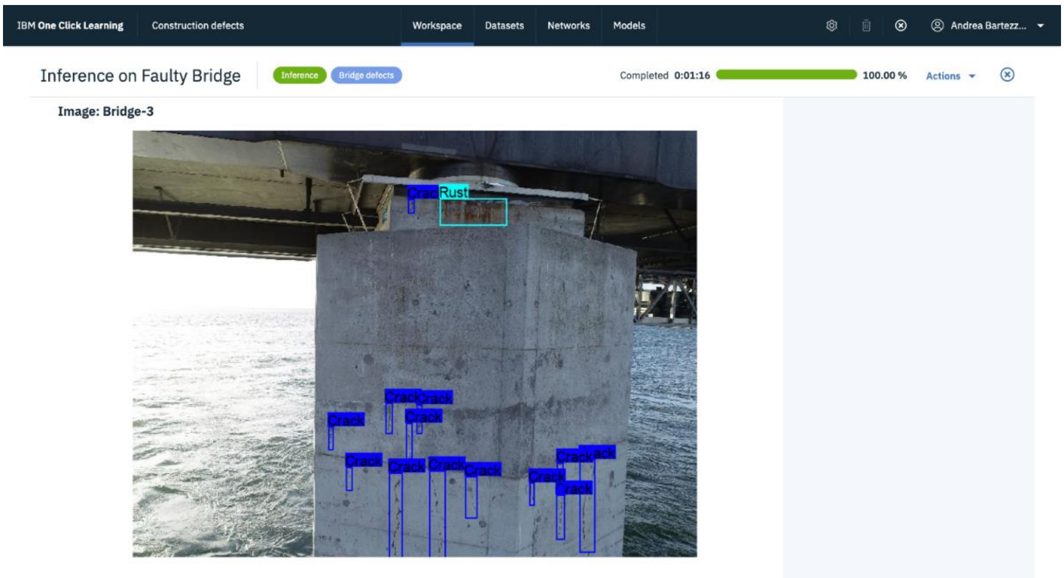


Figure 2.26 - Screenshot of OCL in use for defect detection on civil infrastructures

2.8.2 User interface and user experience

In order to guarantee broad accessibility irrespective of user expertise with coding, ML or hardware, the end user of OCL is provided with an intuitive UI that allows them to fully operate the platform through an interactive graphical interface, similarly to the majority of other commercial platforms.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

However, handling completely different tasks within a fixed user interface does not guarantee the best user experience: for example, exploring a civil infrastructure dataset composed of many annotated images organized in a hierarchical structure, is more efficient with the availability of specific navigation tools which exploit this particular structure, if compared to exploring a generic object detection dataset with standard tools.

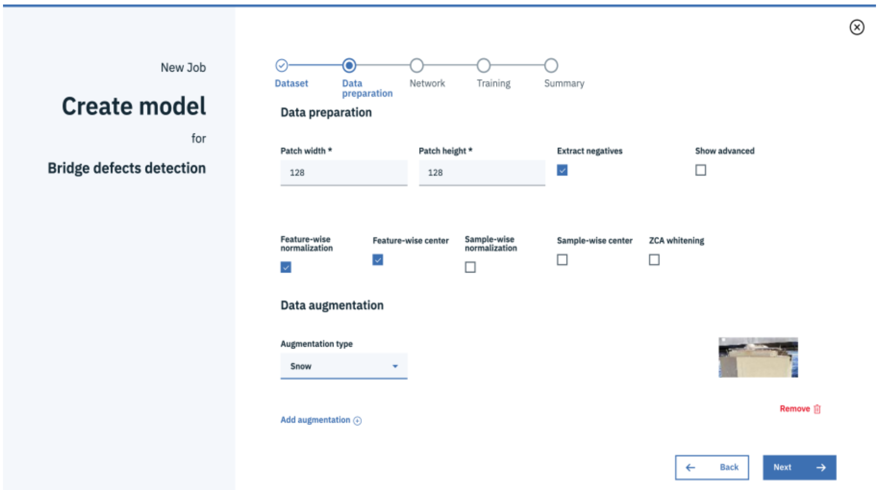


Figure 2.27 - Screenshot of the ML model creation UI

Therefore, the user interface is built with modularity and extensibility in mind in order to allow it to support and incorporate new requirements and user requests.

In addition, experienced users that require programmatic access to the platform are provided with a REST APIs that provides full control on all parameters and enables programmatic interaction with the platform, which could be necessary when clients need to integrate the system in existing data pipelines.

2.8.3 Backend implementations of the IBM OCL platform

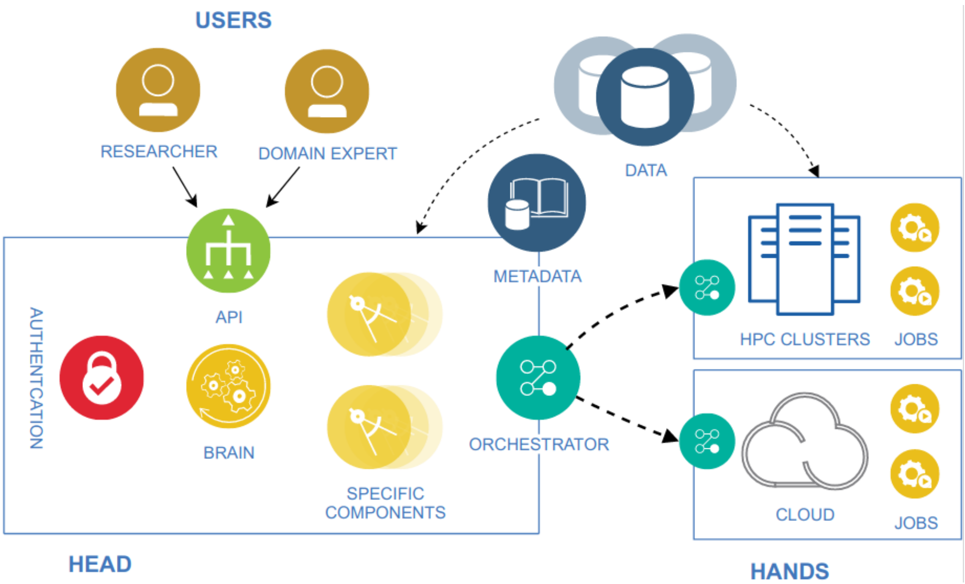


Figure 2.28 - Diagram sketching the OCL architecture

This section introduces the backend architecture at the core of the OCL platform sketched in Figure 2.28.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

The system is divided into two main logically separated parts: the *head* and the *hands*.

The **head** is the main centre of operations. It includes all components which are not computationally-intensive, which deal with:

- management of users and authentication providers
- high-level management of projects and data
- organization and access to the computing resources
- scheduling of jobs
- user interface and REST API.

The architecture is based on micro-services and is usually deployed in the cloud, for ensuring high availability.

Internally, the backend design is devised into a set of basic *entities* which are common to all AI workflows and are passed around the various services (among the others: *datasets*, *annotations*, *models*, *tasks*, *runs*, *users*, *projects*, etc.).

Use case-specific assets can define specialized entities on top of the basic ones: examples in the civil engineering domain are civil engineering asset, defect, and others.

All components talk to each other by means of a centralized document-based NoSQL database, where metadata regarding all *entities* is stored. The database contains no data by itself, but just information on how to access it as pointers to external sources. A convenient API layer provides the means to then access the data by UUID (Universally unique identifier**)** or name, transparently from its actual physical location.

All micro-services are stateless, to lower their complexity while ensuring scalability.

Specifically, OCL has two sets of services:

- *Server*: the main entry point of the system, providing the REST API layer offering access to all functionalities and hosting the user interface
- *Orchestrators*: multiple services handling the organization and scheduling of jobs.

Hands refer instead to the logical part running computationally-intensive jobs. This component can have special requirements in terms of access to the hardware: it may need one or multiple GPUs, exclusive access to multiple cores, extensive amounts of RAM, or big local storage as cache.

Each hand can be linked to different computational resources (being on cloud, or on other clusters).

In OCL multiple *hands* can be registered simultaneously and it is up to the *head* to provide transparent access to them, to handle data transfers in and out and to manage the execution of jobs.

In particular, jobs are scheduled in two stages:

- A high-level orchestrator checks the presence of new jobs to run from a global queue; these are usually pipelines divided into stages (or *runs*). Based on the pipeline needs in terms of hardware and the accessibility of the user to the available computational resources, the orchestrator decides to which hand to submit the runs.
- A hand-specific orchestrator takes the runs from its own specific queue, prepares the container or environment in which to execute them (based on the type of deployment to use) and submits them for execution.

This last step depends on the actual computing environment the job is destined to run on.

In case of cloud-based deployment, jobs are handled as *kubeflow pipelines (9)*. When dealing with HPC clusters, a specialized orchestrator connects to the login node, prepares the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

environment and data, and submits the job using the cluster-specific job queuing system (e.g. IBM Spectrum LSF).

It is the duty of the orchestrator services to keep track of all hands and to report real-time information about running jobs to the head.

2.8.4 Case study: Visual Inspection of transport infrastructure with OCL

This section, presents the application of OCL to the visual inspection of transport infrastructures, taking the particular example of bridge inspection.

This specific use case emphasizes the user- and data-centric aspects of OCL, in particular its modular design and extensibility that allowed it to integrate in a traditional ML workflow the management of customized application-specific assets and analysis tools in terms of specialized entities that are of interest to civil engineering users and domain expert.

Until recently, bridge inspections were exclusively a manual process conducted by reliability engineers. This practice can sometimes be difficult due to the complexity of the structure and difficult accessibility of some structural elements, but it is also estimated to cost around $50B corresponding to $2B man-hours annually.

The main objective of the inspection is to assess the condition of an asset and determine whether repair or further maintenance operations are needed; engineers observe the structural status, taking information by detecting defects, such as cracks, spalling, rust or algae, and assessing their severity, relative to the defect size and location in the structure.

The advances in drone technology and its falling costs have recently pushed this laborious process of manual inspection progressively towards automation.

Flying drones around a structure and using embedded high-resolution cameras to collect visual data from all angles not only speeds up the inspection process, but it also removes the human from potentially dangerous situations.

In addition, thanks to the power of artificial intelligence capabilities, defects can be detected and localized with high precision automatically and presented to the engineer for further analysis.

The OCL platform provides experts with such advanced capabilities.

By taking a data-centric approach, it focuses the user experience around the management of users' assets, providing specific features for:

- powerful data exploration of large datasets and high-resolution images with corresponding detected damage
- detection, characterization and measurement of damages
- reconstruction of infrastructure elements and localization of damages with automated image stitching
- quick extraction of actionable results
- empowering engineers and infrastructure managers to use pre-trained AI models for everyday inspections.

A compelling visual inspection use case is that of the Storebælt East suspension bridge, owned and operated by Sund & Bælt. The bridge is part of the Great Belt Fixed Link in Denmark. At 254 metres above sea level, the East Bridge has a length of 6790 metres and a free span of 1624 metres. A visual inspection of all 22 pillars was conducted in June 2021 with Matrix DJI 300 RTK drones. More than 23k high-resolution images were collected, where each image consists of ~6k x 4k pixels (24M pixels). The OCL platform is used to store, analyze and visualize these images.

First, as seen in Figure 2.29, general statistics are presented to the engineer. The left side presentes a view that provides a progressive decomposition of the bridge into its components by diving into the pillars, their orientations and corresponding images, and allowing the user

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

to understand the hierarchy of the structure to quickly locate data of interest. On the right side, a summarized view of the images collected during drone inspection and the defects detected and classified during the inference of the AI model is provided.

The defects are classified into categories specific to the use case (i.e., 6 here: crack, rust, spalling, algae, net-crack and crack with precipitation). Below the summary section and more detailed statistics are presented, such as average area (in m^2) per defect type and distribution of defects across the dataset.

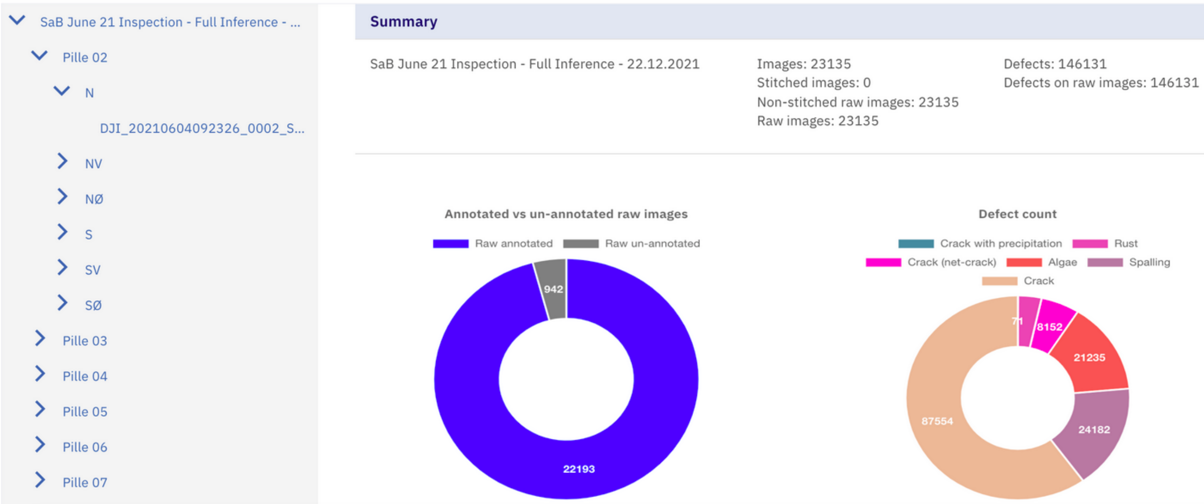


Figure 2.29 - Hierarchical view of assets (left); summary of dataset and associated defects, as detected and classified into 6 categories during AI-model inference.

Additionally, the platform provides filtering options to enable an intuitive and fast navigation through the defects. Specifically, the user can select one or more defect categories and provide ranges of interest for attributes like area (in pixels or m^2), confidence score as resulting from the model inference or severity rating. This will return the set of images satisfying these criteria. The user can navigate through the selected images and visualize all defects detected in a specific image, as seen in Figure 2.30.

By hovering with the mouse over a defect, more details of interest are provided, such as type of defect, area in pixels and m^2, prediction score and severity level.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

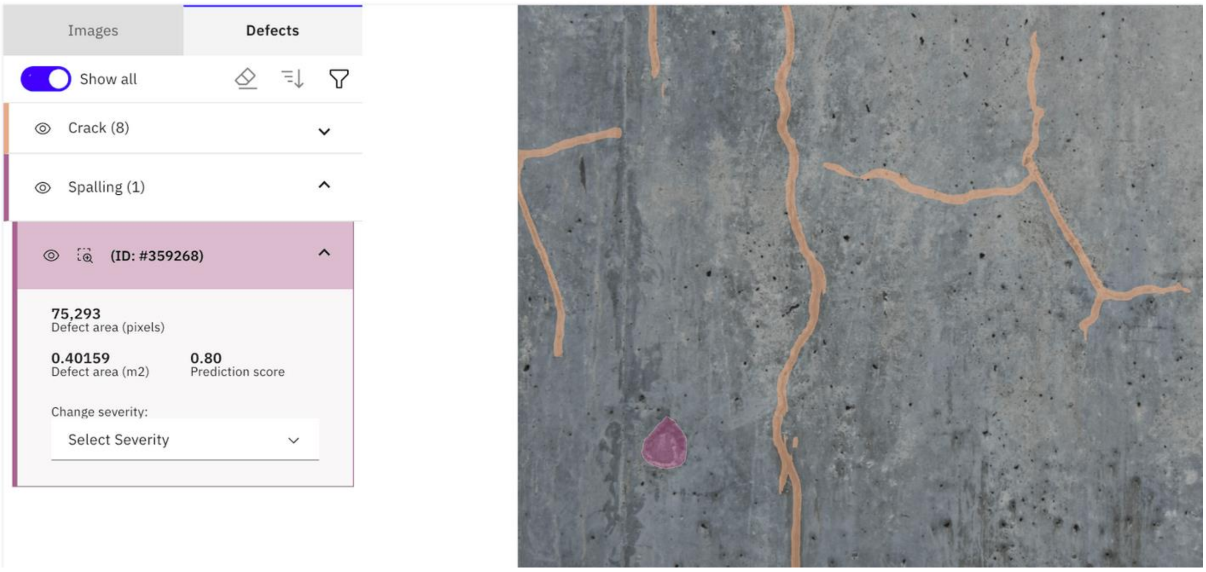


Figure 2.30 - Visualization of defects in specific image with associated confidence score as computed by the AI-model.

While the functionality described so far applies per individual image, it already provides great benefits to engineers. However, decisions around repair and maintenance take into consideration the locations of defects in a bridge structure as well. Powered by image stitching algorithms, OCL provides an overall view of each element, as bridge pillar.

These stitched images are reconstructed automatically from the raw high-resolution photos taken by the drone, by using image rectification and location reconstruction algorithms, and combine all defects detected in the individual images that were attached together, as shown in Figure 2.31. It is always possible for the engineer to go from the overview image to the underlying raw photos for further detailed inspection (e.g., by clicking on the red highlighted area in Figure 2.31. Intuitive navigation through the raw photos is possible by means of a minimap (shown in Figure 2.32), which allows also to have an overview of where in the pillar each defect is located.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

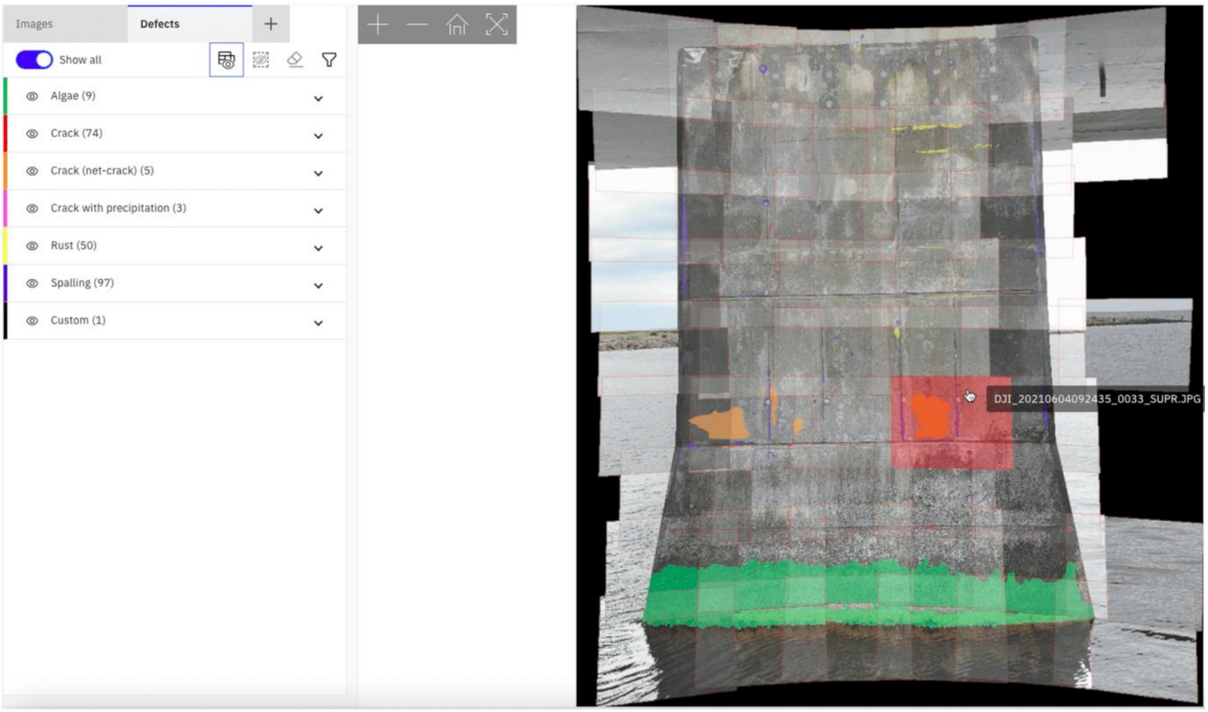


Figure 2.31 - Overall view of a bridge pillar, after image stitching algorithm was applied. Summary of defects is provided on the left and different categories are distinguished by color in the stitched image.

These stitched images are extremely large in size, as they preserve the full resolution of the already detailed raw photos. In order to deliver a smooth navigation, in the civil infrastructure component in OCL, images are divided in multi-resolution tiles and streamed to the browser from cloud object storage on demand. This is just one of the challenges that is necessary to overcome when dealing with such a big amount of data. Nevertheless, thanks to the flexible architectural design of the platform and the scalability of cloud-based solutions, OCL is able to deliver a smooth user experience in these complex use cases also.

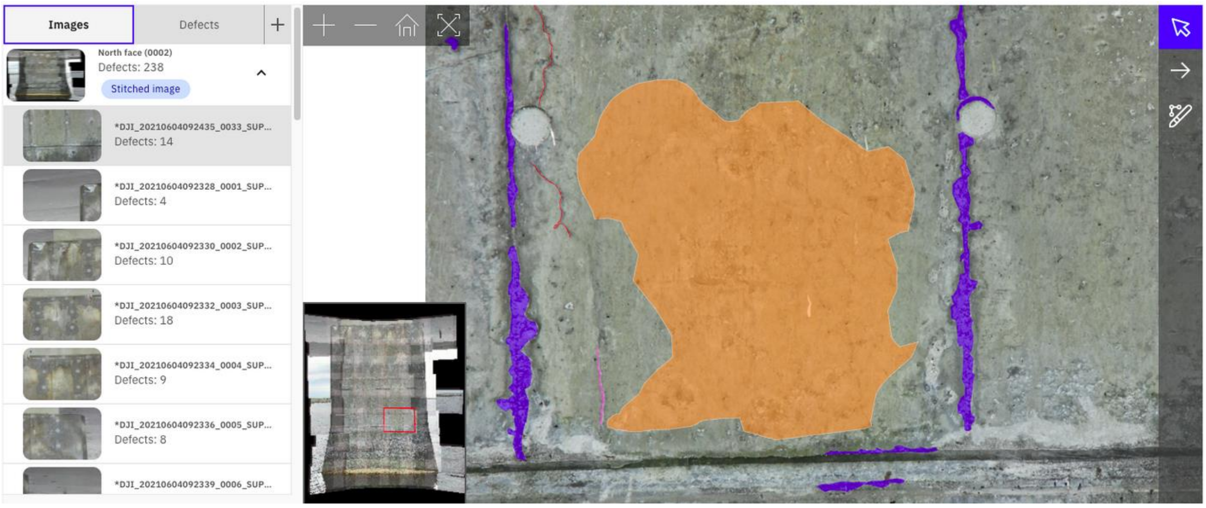


Figure 2.32: Highlight on a defect; a minimap with the overview stitched image is showing the corresponding location of the defect on the full pillar.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.9 SHM MONITOR Platform

2.9.1 Platform overview

Input Data	Images
Functionalities	Data ingestion and storage Data processing Creating and analysis of static analysis models Real-time data sharing Warning via e-mail and/or SMS
Users	Structural engineers Asset managers

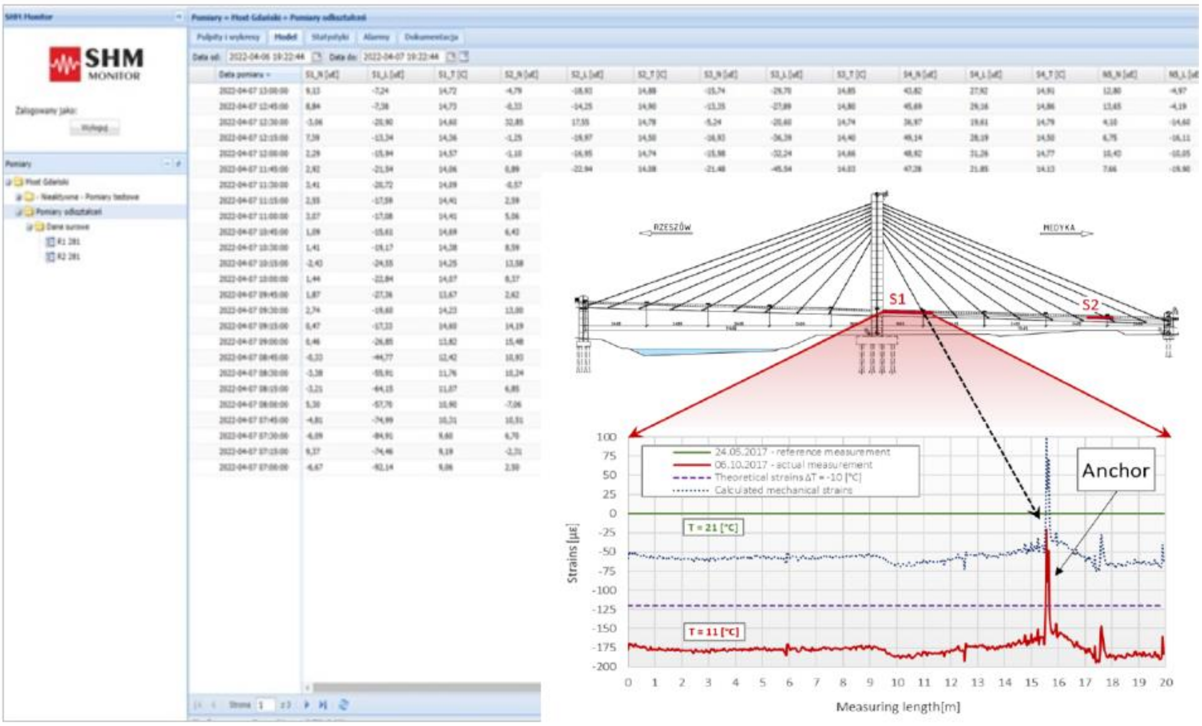


Figure 2.33 - Screenshots of the GUI of the SHM platform

One the main instruments for effective long-term structural infrastructures monitoring is the control of key performance indicators such as: strains, forces, displacements as well as temperature and humidity for the assessment of the technical condition of the structures. The SHM platform provides a fully-automated on-line system for the real time monitoring of these parameters on bridges and remote management of collected data. SHM Monitoring systems are based on sensing technologies for full-scale monitoring of civil engineering structures. The software is designed for collecting, processing and measurement data sharing.

The creators of the software, the company SHM Systems from Poland, is specialized in monitoring systems for safety in civil engineering and geotechnical structures. The company cooperates with R&D departments and universities in the European Union, but also outside of the EU. SHM Systems produce composite Distributed Fibre Optic Sensors systems for continuous geometrical measurements in many proven applications. The company also provides assistance and IT support for the customers.

The functionalities of the platform include: real-time data sharing in a form of charts, tables, statistics; warning about the exceedance of the defined limit values via e-mail and/or SMS; creating and analysing any computational models using the collected data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

SHM monitoring system can be adjusted for a specific application and used in various fields such as:

- non-destructive diagnostic of structures conditions (roads and bridges, tunnels, railway lines, pipelines)
- measurements of the ground vertical displacements
- reduction of the failure risk by early detection of deflections
- process optimization for better selection of construction technology
- control of vibrations generated during construction works
- detection of local damages (e.g. cracks) and assessment of the deformation state of monitored object.

2.9.2 User interface and user experience

The access to the platform is available from the website and after opening the first interface there is visible is login panel as below:

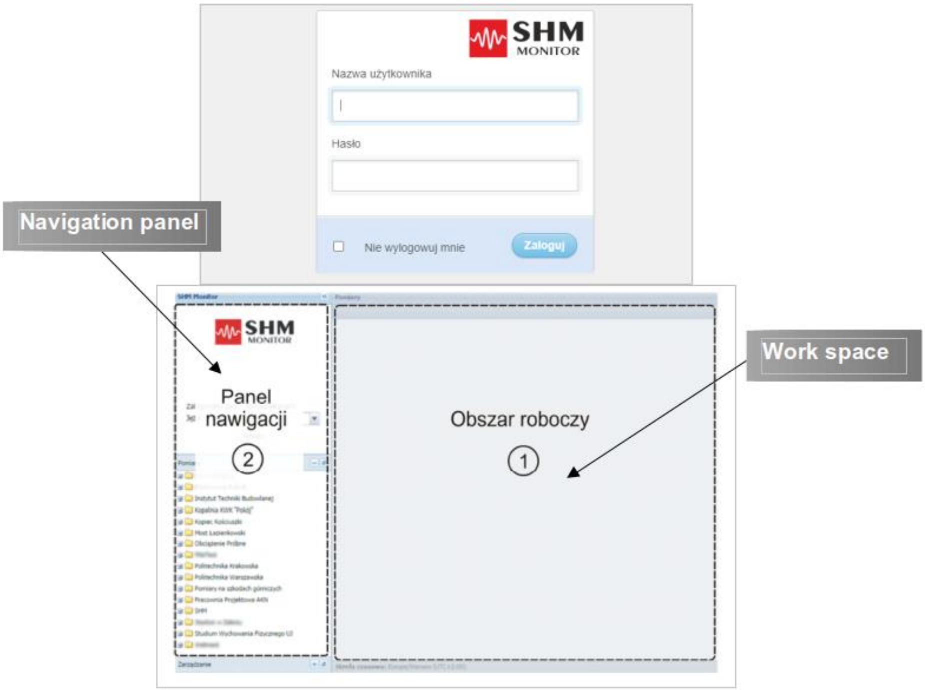


Figure 2.34 - Login, navigation and workspace panels

After entering the credentials, the user has access to the navigation panel which enables to open any registered project for different infrastructure objects with monitoring system:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

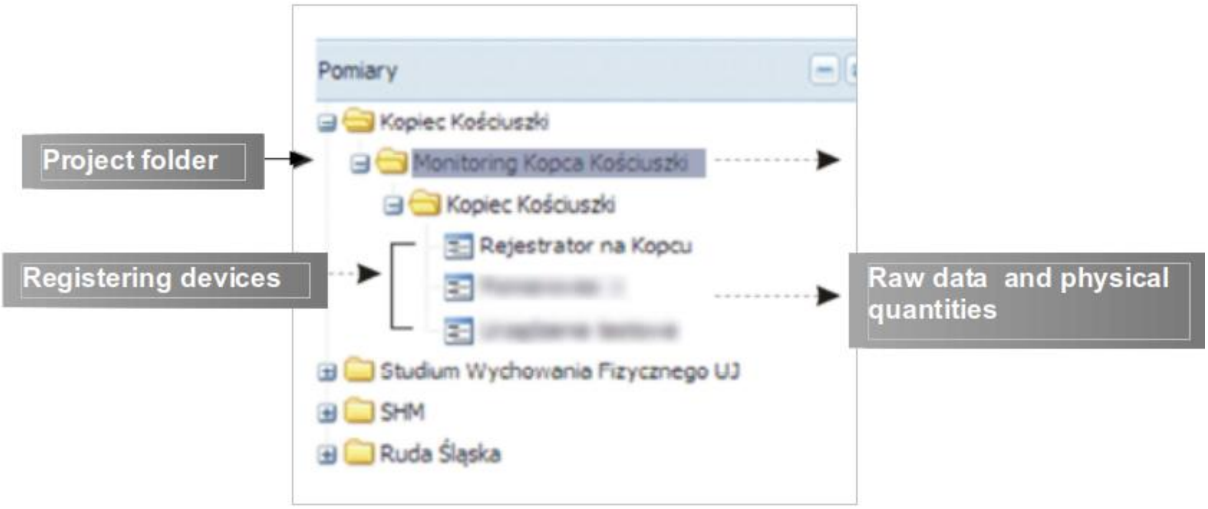


Figure 2.35 - Project management window

Unfolding the panels will open the placeholders where the user can find all the calculations and mathematical model for the monitored object and raw data table as below:

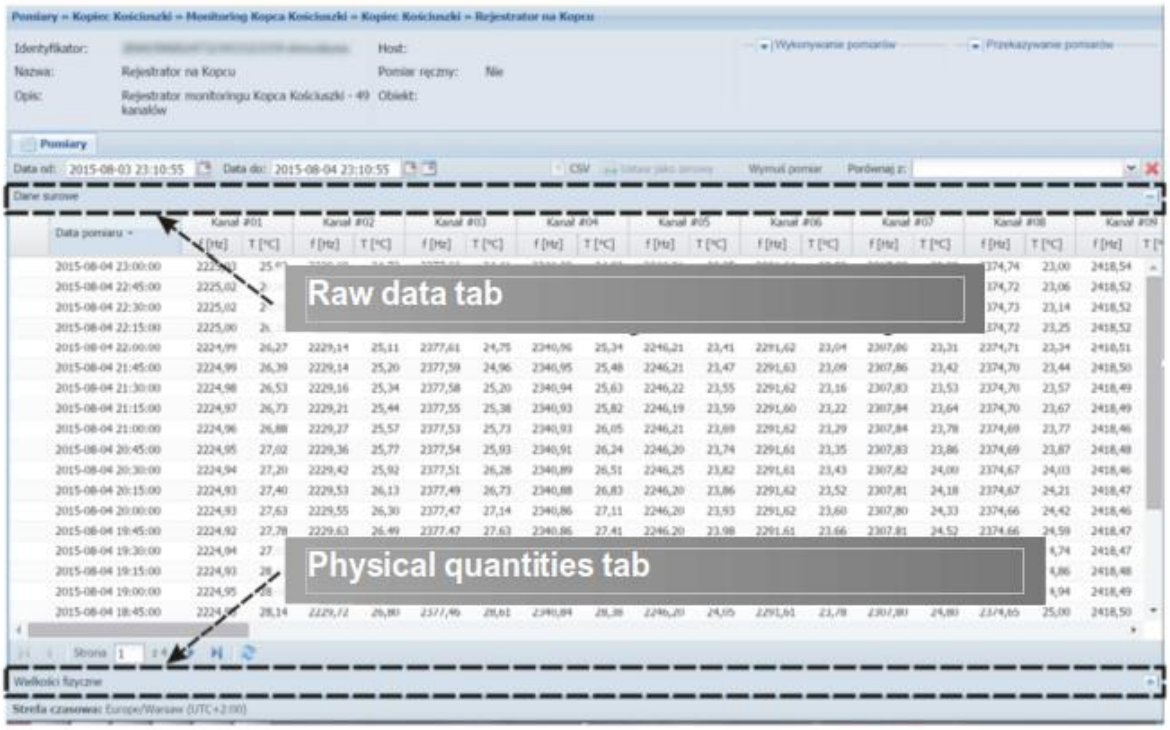


Figure 2.36 - Main view of the outputs of mathematical models

The main interface is simplified for intuitive navigation. Available navigation elements for the user are:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

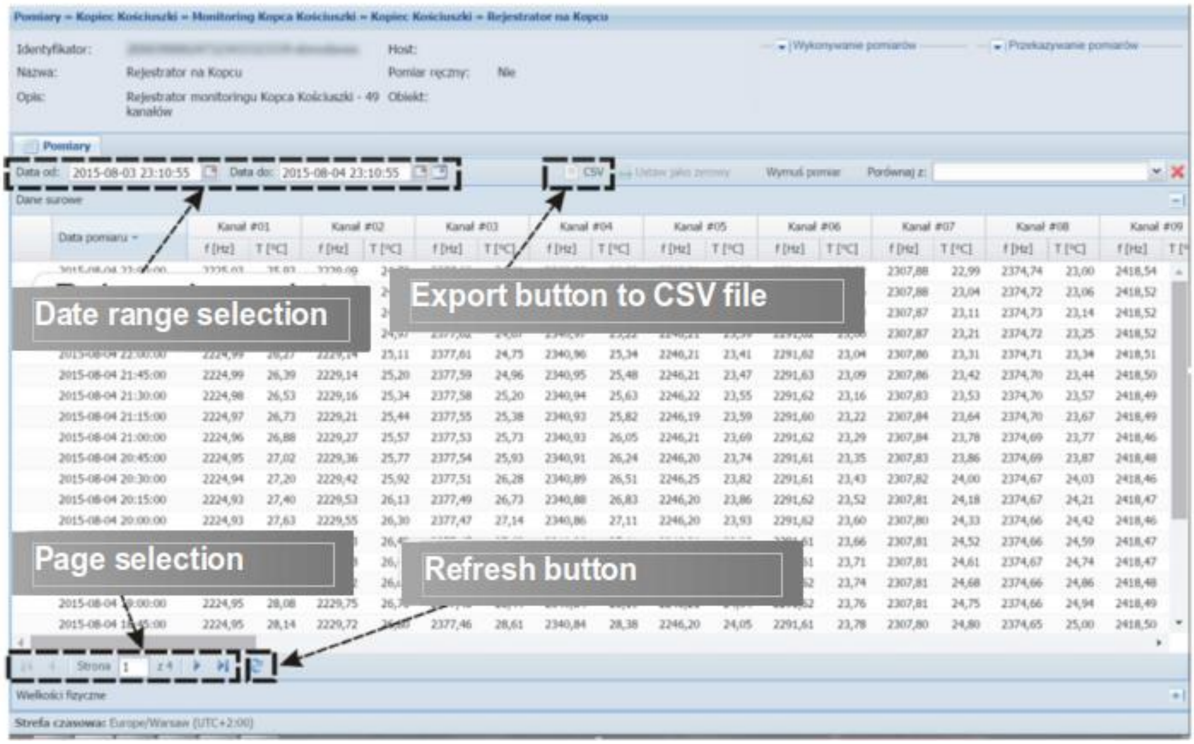


Figure 2.37 - Navigation of the main window

Project selection from the navigation panel gives access to the workspace presenting the calculated model sizes. The data are presented in the form of tables, charts, information dashboards and maps.

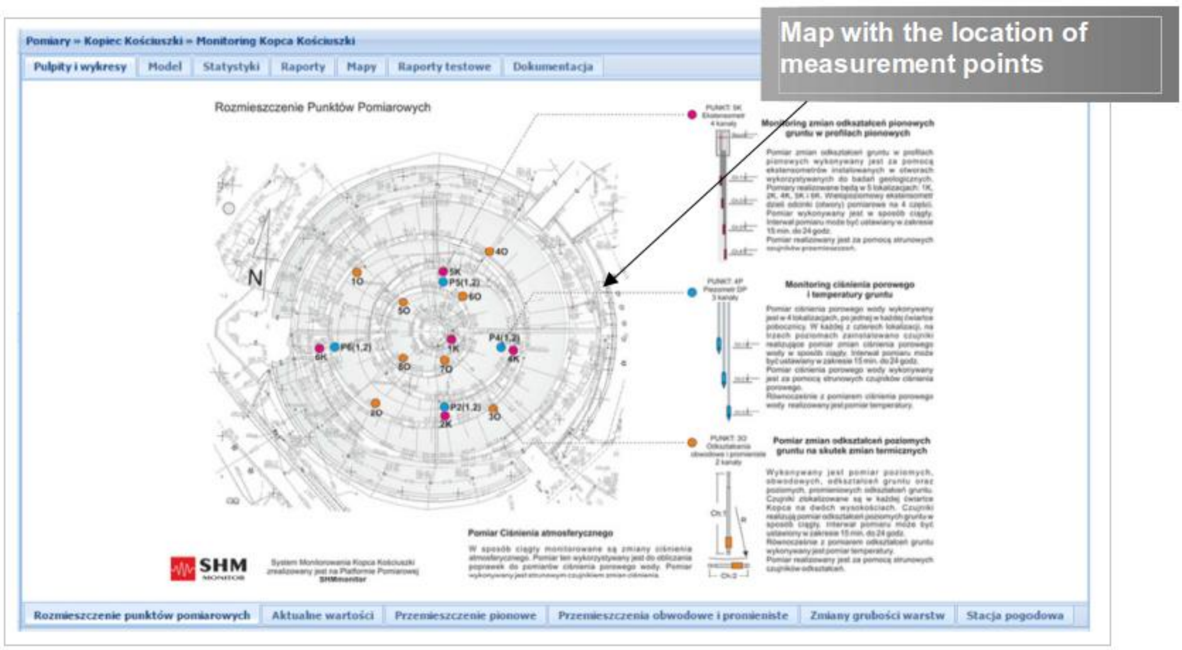


Figure 2.38 - Map of the sensors localization

In another tab called Desktop and Charts the user has access to the graphs with the registered data during selected period of time:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

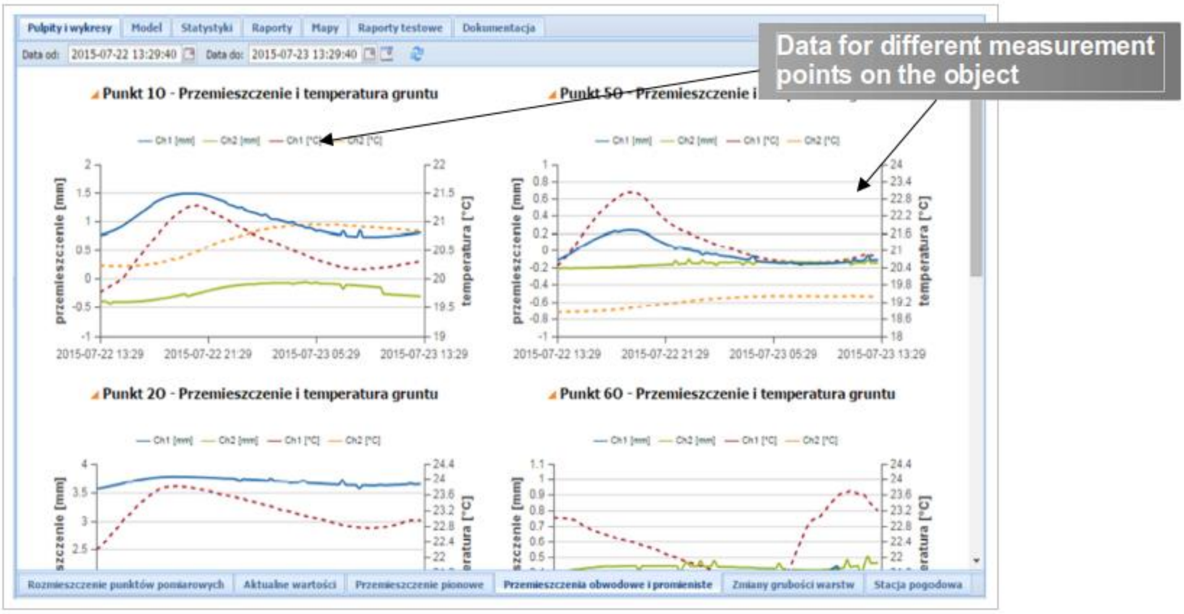
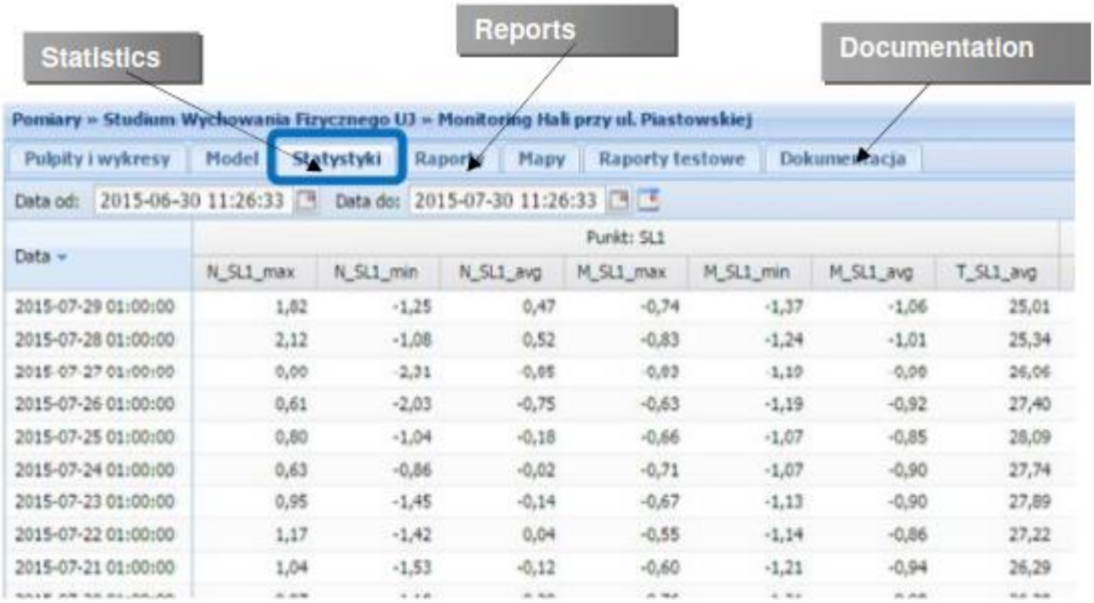


Figure 2.39 - Graphs with the recorder data from different locations

The user can easily enter the tabs with mathematical model as well as with the statistics over the data recorded in the specific period of time.



Data	Punkt: SL1						
	N_SL1_max	N_SL1_min	N_SL1_avg	M_SL1_max	M_SL1_min	M_SL1_avg	T_SL1_avg
2015-07-29 01:00:00	1,82	-1,25	0,47	-0,74	-1,37	-1,06	25,01
2015-07-28 01:00:00	2,12	-1,08	0,52	-0,83	-1,24	-1,01	25,34
2015-07-27 01:00:00	0,00	-2,31	-0,05	-0,83	-1,10	-0,90	26,06
2015-07-26 01:00:00	0,61	-2,03	-0,75	-0,63	-1,19	-0,92	27,40
2015-07-25 01:00:00	0,80	-1,04	-0,18	-0,66	-1,07	-0,85	28,09
2015-07-24 01:00:00	0,63	-0,86	-0,02	-0,71	-1,07	-0,90	27,74
2015-07-23 01:00:00	0,95	-1,45	-0,14	-0,67	-1,13	-0,90	27,89
2015-07-22 01:00:00	1,17	-1,42	0,04	-0,55	-1,14	-0,86	27,22
2015-07-21 01:00:00	1,04	-1,53	-0,12	-0,60	-1,21	-0,94	26,29

Figure 2.40 - Other available tabs

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

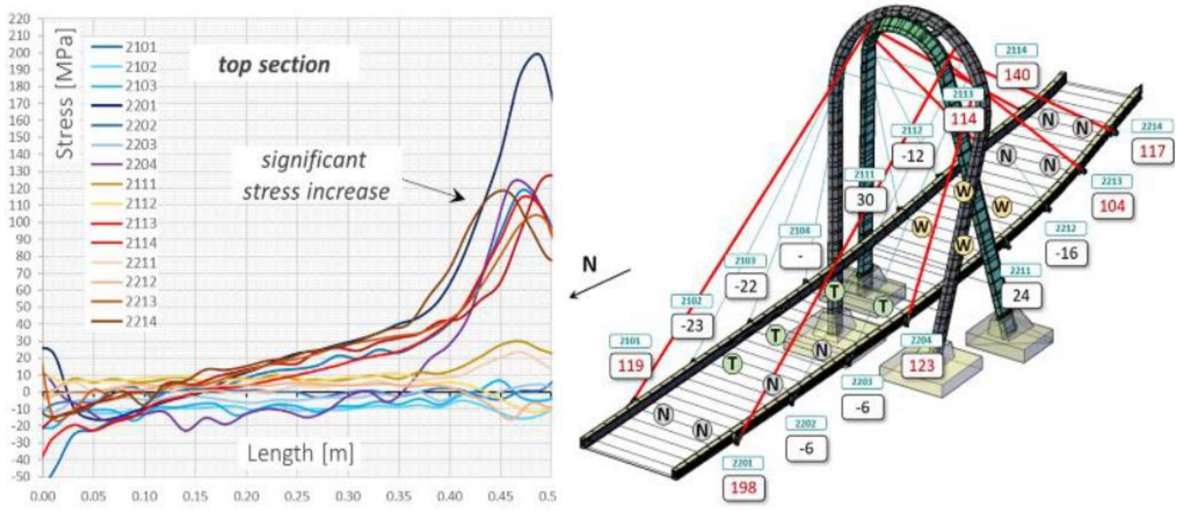


Figure 2.41 - Example of the data – stress measurement on various lengths of anchorages

2.9.3 Backend implementations of the platform

Data management from in the SHM Monitor from the backend view can be represented as follows:

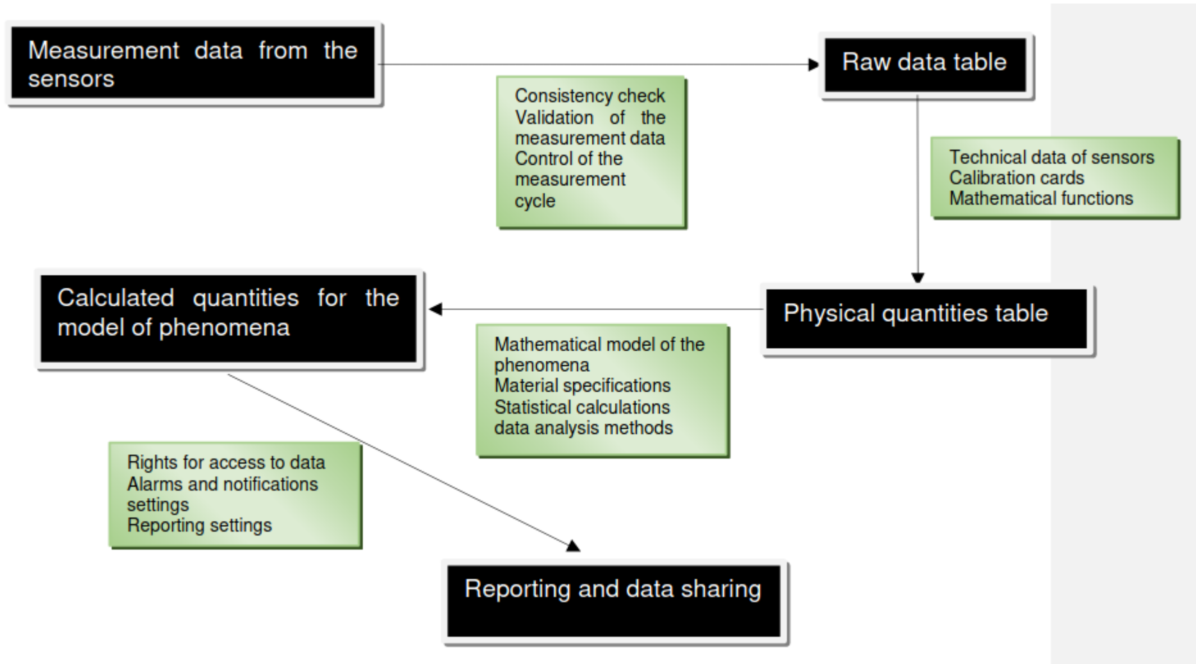


Figure 2.42 - Data management flow

Measurement data and all the information used for their processing are stored and available in the software. Data available for the user without administrator rights are as follows:

- raw data: data sent directly by measuring devices – fiber optic sensors. These data are not subjected to any mathematical operations

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

- physical quantities: data obtained as a result of calculating raw data according to formulas and calibration cards characteristic for individual sensors. As a result of this operation, the data represent specific physical quantities, i.e. deformation, pressure, temperature, forces
- calculated model values: these data are created as a result of any mathematical calculations performed on raw data and physical quantities. The user can introduce any mathematical model according to which the data will be converted
- results of statistical calculations.

The alarm station and navigation point on site is organized in a way that the end user is notified by e-mail or mobile after the critical values of the measured parameters will be exceeded, in some cases such as monitoring of vibrations there is also a sound alarm system on site:

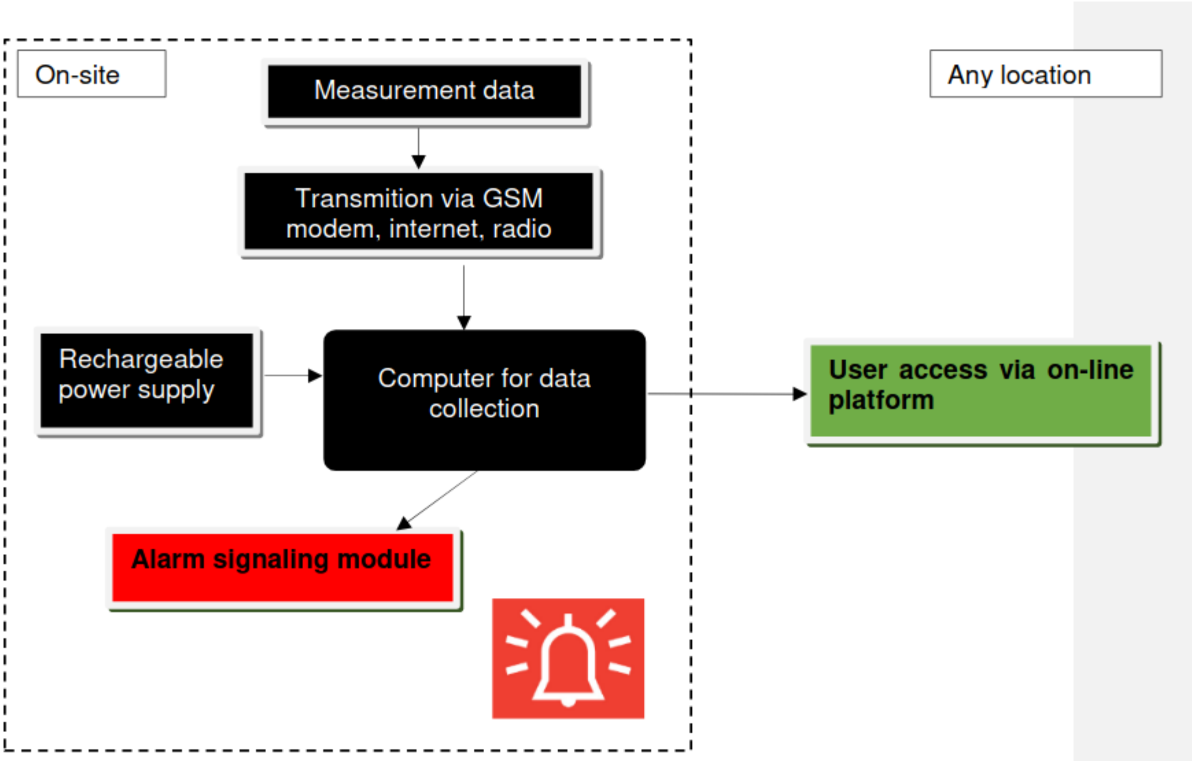


Figure 2.43 - Data transfer flow and on-site alarm station

2.9.4 Case study

The SHM platform for structural health monitoring based on the Distributed Fibre Optic Sensors (DFOS) has been implemented on a steel bridge over the Vistula River in Warsaw which has been constructed over 65 years ago in order to monitor deformations and temperature. Data transmission has been running since August 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Figure 2.44 - Gdański bridge in Warsaw

Monitoring system includes eight string sensors, eight fibre optic sensors, data registers and control station on-site. Underneath there is presented the sketch of the measurement points localization on the bridge structure:

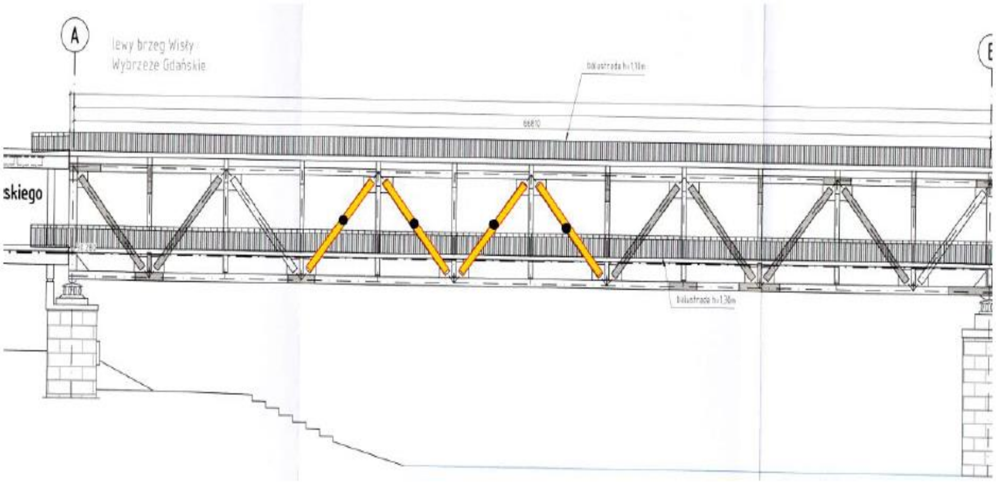


Figure 2.45 - Sketch of the bridge structural elements monitored`

The SHM monitoring platform enables remote view of the current state of the temperature and deformations changes on the object. After logging into the platform the user can open selected folder, in this case with Gdański bridge data and see the complete documentation of the implemented system as below:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

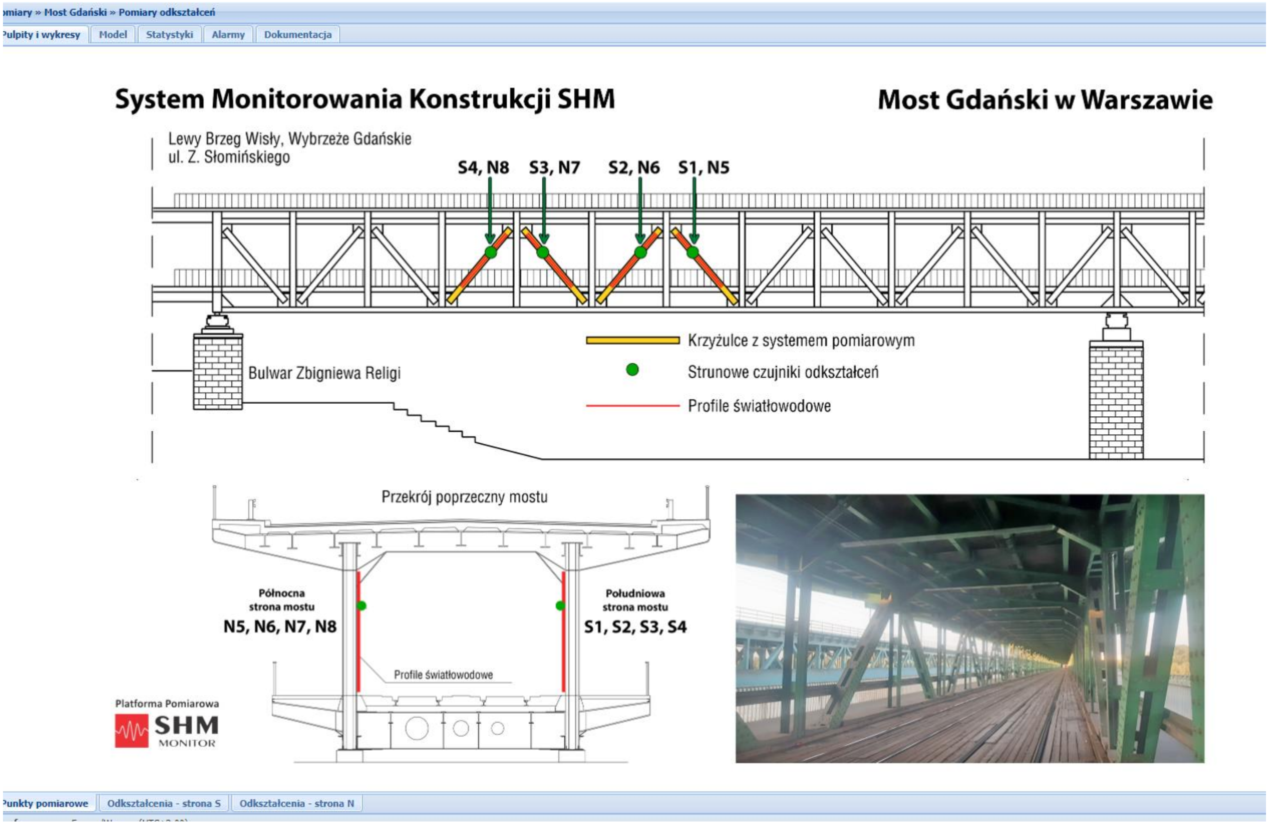


Figure 2.46 - Documentation of the monitored bridge

As shown above, there are sketches of the bridge parts of interest as well as a photography of the bridge. In the folder of the deformation measurements on the bridge there are data for each measurement point on the north and south side of the bridge, which are synchronized in real-time:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

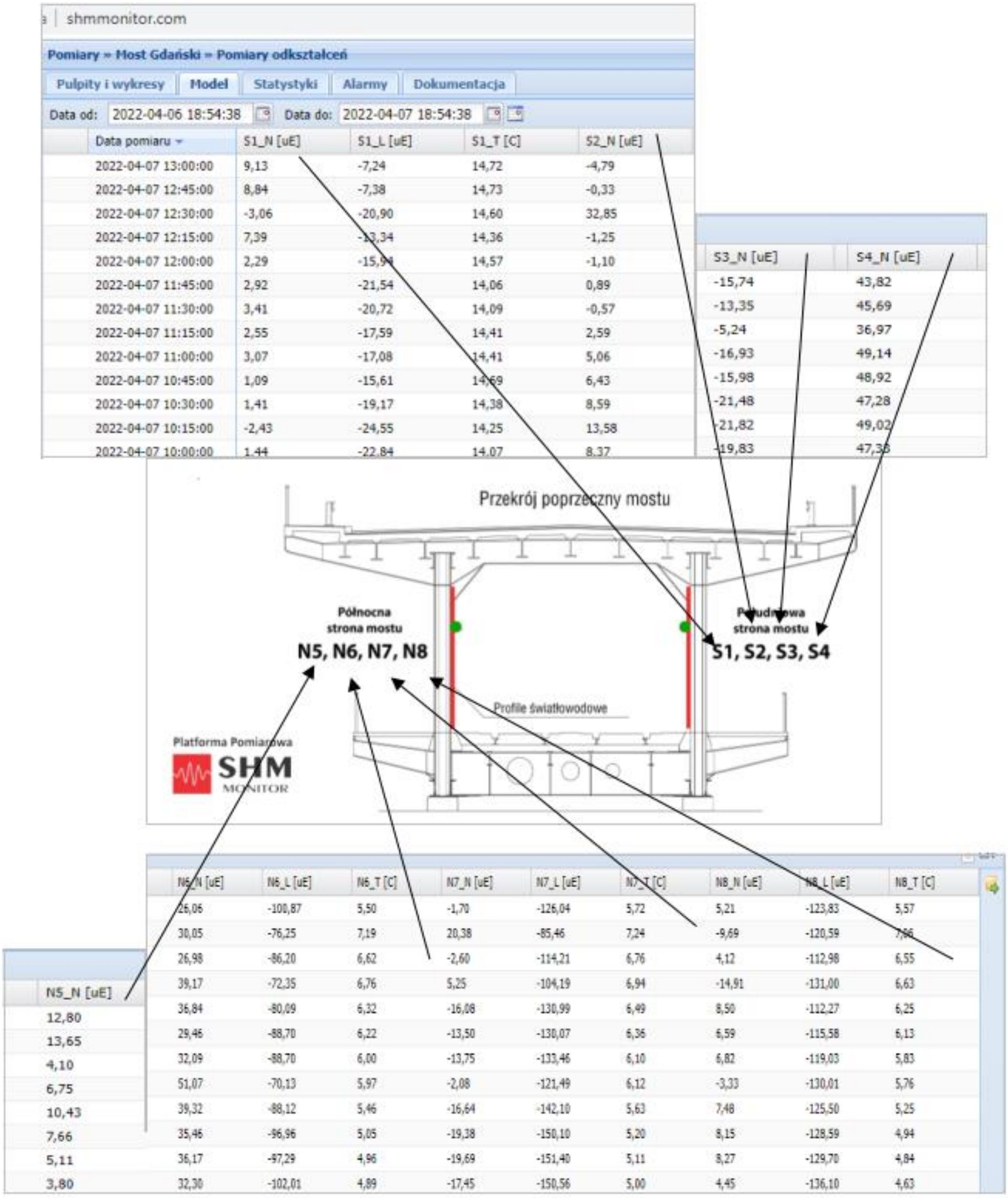


Figure 2.47 - Data assignment in the platform for specific case

The data can be the exported further to CSV file:

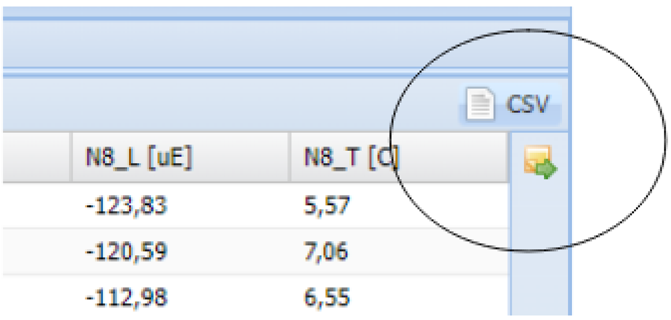


Figure 2.48 - Export of data in SHM platform

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

2.10 Relatics

2.10.1 Platform overview

Input Data	Text (documents, inspection data, georeferenced data) Documents (documents, articles, reports, minutes, technical notes) Images External data via web services
Functionalities	Project information management Progress monitoring Data integration
Input Users (data feeding user)	Project engineers Contractors
Output Users	Contractors Stockholders

Relatics is a Dutch company established in 2003. Its homonymous product is one of the leading software applications that supports Systems Engineering for construction projects. It is used in thousands of projects around the world and has helped more than 50,000 users manage their projects. The Relatics software is a model-based systems engineering (MBSE) application that offers a collaborative platform for both stakeholders and contractors within construction projects. It enables the users to manage project requirements, design, analysis, verification, and validation activities. Moreover, it is used to manage the infrastructure assets for inspection and maintenance projects.

It is a cloud-based multi-user platform that replaces the traditional document-based workflows with modern web interfaces and acts as a source for all information during the project lifecycle. Information is managed in a structured way within one information management system and the relations between data elements are explicitly defined in the application. Thus, data quality, consistency, and integrity are ensured without redundancy for all project workflows.

2.10.2 User interface and user experienced

The application is template-based where the agency or company has a source template for similar projects, and it can be extended to fit the new project requirements. The project owner can define the project models, and each model has its own attributes, properties, and relations to other models.

Also, each project has several customized dashboards to monitor the requirements, work progress, and risks.

In addition, the platform provides a configurable reporting system that can standardize the outcomes of the projects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

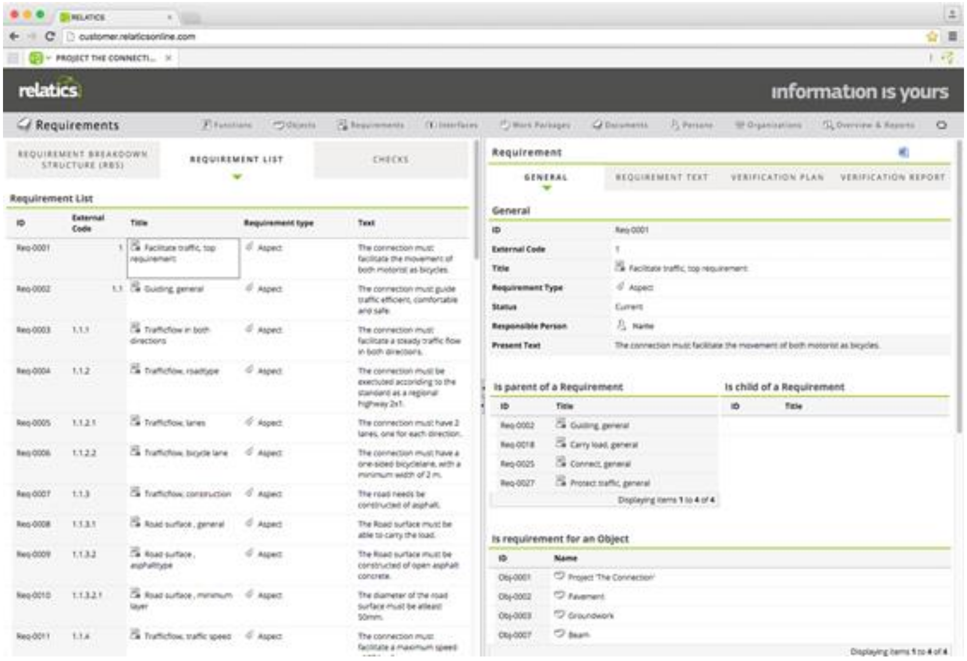


Figure 2.49 - Example of document control interface in Relatics (Source: www.relatics.com).

2.10.3 Backend implementations of the platform

Relatics is a cloud-based multi-user software as a service (SaaS) platform which enables the users to migrate from a document to a model-driven approach. It offers a configurable user interface that can be extended through a JavaScript API to add custom user controls and enhance the overall user experience. For example, the user can use JavaScript to add GIS web viewers or read external data interactively within the UI.

The platform can be integrated with external applications via web interfaces. All the object instances within the platform have a unique identifier called *RelaticsId*. The project owner can define web services to import data to the system via the Relatics web APIs module and then link, append, or update it with existing objects. The created web services work with REST and SOAP protocols. Also, the platform offers web integration services to share data with external platforms (i.e., CAD, GIS, BIM, DMS such as SharePoint). All the web services support OAuth2 authentication.

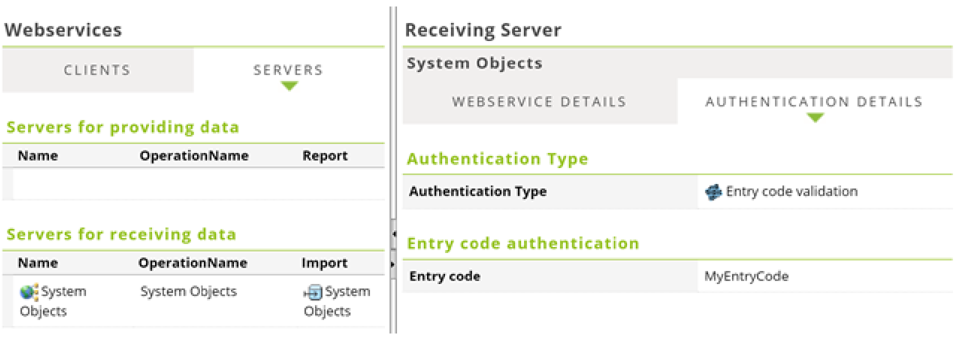


Figure 2.50 - Example of web services definition interface (Source (10)).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

2.11 CELOSÍA

2.11.1 Platform overview

Input Data	TXT (Experimental data, e.g.: movements, loads, accelerations, etc.) JPG/JPEG and PNG (schema and illustrative figures) PDF (documents, articles, reports, minutes, technical notes)
Functionalities	Data ingestion and storage Data processing Visualization
Input Users (data feeding user)	Installation responsible Maintenance responsible Monitoring manager
Output Users	Dep. Involved in monitoring and analysis of structures

CELOSÍA is a platform allowing the centred and homogeneous management of all the information generated by the structures of the Spanish Road Network with a monitoring system installed. Besides that, the platform is has the aim to unify and universalize protocols, formats and calculation processes, assessing the continuous monitoring of the structural behaviour of the structures.The platform was developed to monitor and analyse in real time infrastructures of the Spanish Road Network. The main objectives of the monitoring are:

- Tool to assist construction
- Control the evolution of an existing damage processes
- Monitoring the correct structural performance of the structure during service life
- Assistance in the operation of the structure
- Database for research projects
- Data lake for standardization

These should ensure an improvement of road safety and serviceability, as well as an early detection of potential infrastructure problems.

The users of this platform are mainly:

- Data Feeding User:
  - Installation responsible
  - Maintenance responsible
  - Monitoring managers
- End user:
  - Departments involved in the monitoring and analysis of structures.

CELOSÍA allows to register bridges and geotechnical works that have or have had during some period of their life cycle of a monitoring system. In this platform, the data transmission is automatic and in real time. Moreover, the information is presented in a universal way regardless the technology used for monitoring and the consultant who implemented it. The structures can be managed from the same platform, and the information is allocated in the same and unique database.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

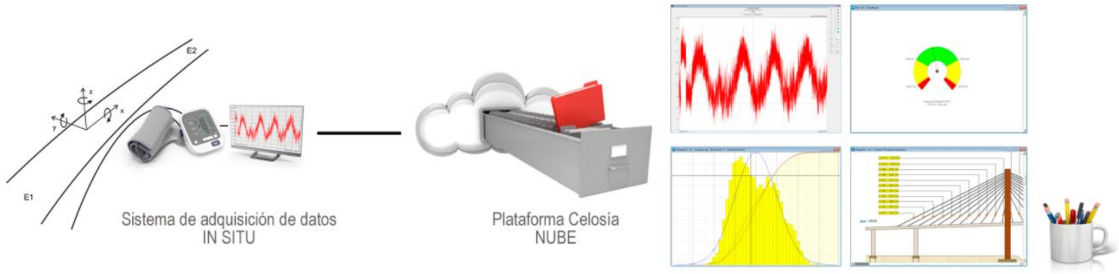


Figure 2.51 - Operation scheme of CELOSÍA (source (11))

In order to register a structure in CELOSÍA, the monitoring manager in each case should deliver the following information to the Platform Director:

(a) Name of the work: title and subtitle identifying the structure and the purpose of the monitoring. For example: viaduct over the river Mayor and historical construction record

(b) Worksheet: in plain text format (ASCII), with the following information:

Road

Section

Demarcation

Province

Project manager

Project designer

Technical assistance to site management

Technical assistance to the site management

Construction company

Site manager

Technical office of the construction company

Instrumentation company

(c) Geographical coordinates of the work (near its central point): latitude and longitude

(d) Photograph (or, failing that, a virtual image) representative of the work, in jpg/jpeg format. It must be of good quality, preferably landscape, with a minimum size of 3000 pixels in its largest dimension

(e) Number of static channels: approximate number of measurement points, real (physical) and virtual (calculated). It shall be an estimate rounded up, adjusted to the foreseen needs of the instrumentation. If it is necessary to increase this number in the future, it shall be communicated to the Platform Director to proceed with the resizing of the database.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

With this information, the Platform Manager will proceed to open a new monitoring item in the platform, which will be immediately reflected in the CELOSÍA's web portal. This item will contain a specific folder for the reception of data from the Monitoring Manager.

The person in charge of monitoring shall send to the Platform Manager the instrumentation sketches showing the position and direction of measurement of all channels. These channels shall be annotated as c1, c2, c3, etc. Together with these annotations, the type and number of the sensors shall be indicated (also in a standardised way).

In addition, the person responsible for monitoring may send the Platform Manager additional information such as technical articles, photographs, plans, diagrams, and figures of interest, which will also be hosted on the portal.

Nowadays, the platform stores 13 structures, with 913 information channels in total. In Figure 2.52, a map containing all the structures being monitored using CELOSÍA can be found.

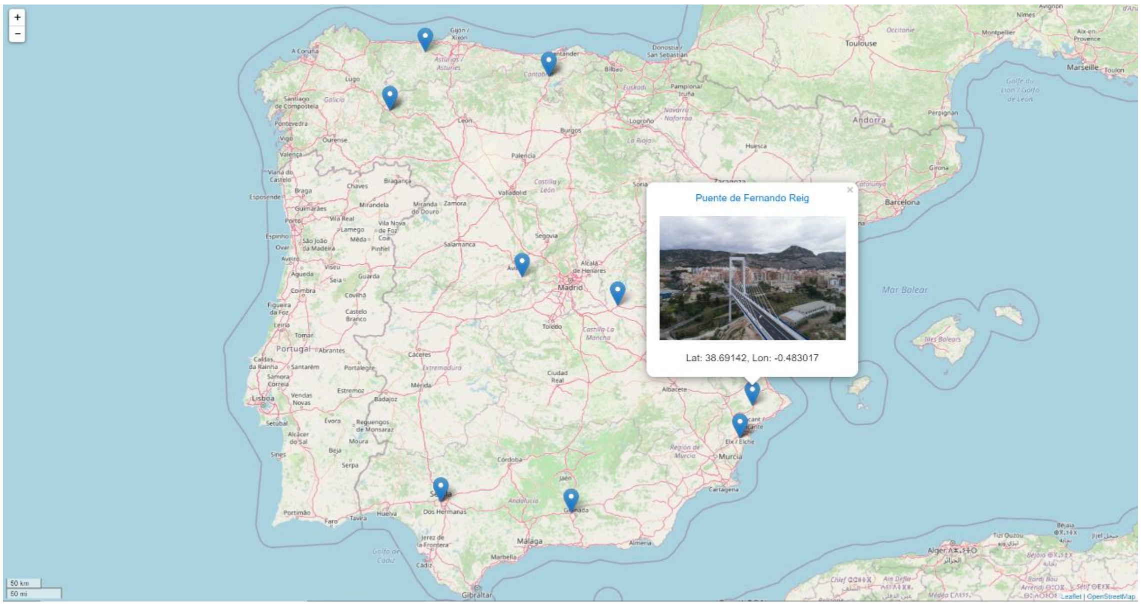


Figure 2.52 - Map containing the structures being monitored through CELOSÍA nowadays (source (11))

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

3 General considerations for the design of a software platforms for data management in the construction sector

3.1 Introduction

This chapter aims to provide a rational framework as a guidance to go through the features and main aspects covered by the selected digital platforms presented in chapter 2, focusing on how these solutions help identifying key technical input for standardization.
A rationalization of the main driving factors is presented below; it aims to analyze the problem from a tripolar point of view, considering the following related aspects: **data**, **function**, **usability**.

Figure 3.1 - Three factors that will be central for the analysis of digital platforms

- **Data**: Based on source, the platform has to ingest, store and process different type of data (i.e. reports, images, sensors data). Digital information can be input or output, it can be qualitative or quantitative, discrete or continuous and, once collected, has to be handled to extract significant information.
- **Function**: to represent the main functionalities of digital platforms it is useful to identify a temporal timeline, following the evolution of the needs, starting from the site inspection activities and ending with the structural assessment of an infrastructure.
- **Usability**: customization for needs and outcomes, requested by users, are main drivers to design complex system. The user, as main actor in the infrastructures management process, can be seen as an input feeder and/or an output checker. Users' interactions with the data can be requested with different levels of detail based on their final usage.

This chapter analyses the general design of software platforms for data management in the construction sector from the perspective of these three poles in turn by exposing their interrelatedness, starting from data, finally analysing the relation between data and functions, then discussing the implication for the downstream pole, the user.

3.2 Data

3.2.1 Overview of data formats and data preparation techniques to be managed in the construction sector

The first step to plan the design of a digital platform or analyse and compare available ones is to understand the type of data produced by each surveying technology and the file format in which such data is stored.

Because data types depend on the surveying technology that they support, they necessarily vary according to the surveying technology supplier. On the other hand, specific data formats depend on the type and processing level of the data. Raw data structures differ from processed software-ready data in ways that are determined by the sensor itself and standardization or downstream data access criteria set by the manufacturer.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

The following list of data and file formats (summarized from D4.1) is a cursory overview of the main file formats that are relevant specifically in the construction sector.

Technique	Expected raw output data format
Optical and visual inspection	• Photos and videos collected on a memory card. • Digital reports
Water penetration test/ Permeability test	The data collected by this technique is numeric, an adequate file format has to be used
Weight in Motion Systems (WIM-Systems)	There are no uniformity and standards with respect to the data management of these systems. The expected outputs include: • Images • Files (MS word, PDF) • Database
Fibre optic sensors	There are no specific guidelines regarding the data storage and management. Raw data from the acquisition system can be exported to a file with a .csv or .xlsx extension
Magnetic and Electrical Methods	There are no standards with respect to the data management of these methods. The information managed from these methods is shown on a digital display and also can be stored in an adequate format
Radiological and nuclear methods	Data can be numeric and graphic and it can be stored in a file or in a database. In both cases, it has to be used an adequate file format. • The appropiate file extension for images is SER • Numeric data can be stored in files with FID format
Surface Measurement	The information managed from surface measurement test are files with the numeric values relative to the measurements. The formats to use can be: • *.txt • *.xlsx
Guided Waves Propagation (GW)	There are some standards that regulate guide wave tests that automate the collection of the data generated in these tests. These standards are ISO 18211:2016 (*"Non-destructive testing – Long-range inspection of aboveground pipelines using guided wave testing with axial propagation"*), BS 9690-1:2011 and BS9690-2:2011 ("*Non destructive testing. Guide wave testing"*). The information managed by guided wawe propagation test have no defined formats and it include the following: • Sensors output in a graphical form • Text files • Database

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Technique	Expected raw output data format
Acoustic Emission	The software for the acoustic emission technique have to comply with EN13477-1 and EN13477-2. The expected outputs for this technique is a database file based on SQLite with one of the following extensions: • *.pridb for the parametric data, status data and user input data • *.tradb for the transient data
Water resistance / Absorption Test	The information generated by water resistance/absorption test is numeric, so it could be managed by a file consequent to this data type
Qualitative chemical methods	The data managed by this technique can be presented in the following data types: • Files • Graphical • Images The formats used for managed this information are: • *.txt • *.pdf • *.csv/*.xlsx
Quantitative chemical methods	The data managed by this technique can be presented in the following data types: • Numeric • Graphical • IR images To manage this information are used the following file formats: • *.txt • *.xlsx
Mechanical tests on cored samples	Software available on the market enables to present measurement data graphically and prepare customized reports according to the test method
Micro Electro-Mechanical Systems (MEMS) accelerometer	There are no standards defined for the management of the data provided by the accelerometers, but it is possible to define some formats to be used for the numeric data: • Files with extension *.csv • Database files • PARQUET column-oriented data files
Micro Electro-Mechanical Systems (MEMS) clinometer	There are not standards defined for the management of the data provided by the clinometers, but it is possible to define some formats to be used for the numeric data: • Files with extension *.csv • Database files • PARQUET column-oriented data files

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Technique	Expected raw output data format
Remote sensing	Satellite data comes in different formats depending on the satellite, sensor type (active or passive), image type, and processing level. Passive sensors usually acquire optical images. The images can be in binary data format for each band accompanied by a header file. The header file contains the necessary information to interpret the data, such as pixel size, projection, number of rows, and columns. Image formats can be common image formats such as GeoTIFF which has its geo-referenced data embedded, or other file formats such as PND, JPEG, and TIFF, and have an extra world file that contains the geo-referencing information. Also, it can be in other standard formats, such as Hierarchical Data Format (HDF4, HDF5, etc.), the Network Common Data Form (NetCDF), or as a set of compressed files. On the other hand, active sensors generate radar satellites images, usually with a high resolution. COSAR files can be a standard for this technique. COSAR files serve as a container for hosting complex sensor data and are formed by the following: • Annotation files for the schema definitions in this document: *.xml. • Images, with types of TIFF or GeoTIFF, with different depths and representations: *.tif • COSAR image format: *.cos • Binary raster file: *.bin • Standard PNG: *.png, *.txt • Text files: *.txt
LiDAR	The expected outputs of LiDAR systems will be different depending on the nature of the system (TLS or MLS). Point cloud data comes in different formats, depending on the sensor, and several standards such as .las, .laz or .bin. Moreover, the images captured by the RGB sensors are saved in the chosen image format, such as PNG or JPEG, or a raw format convertible by software offered by the manufacturer. In the case of RGB images taken by a 360° panoramic imaging camera, the output files will generally be produced in one file containing several images. Using the software or API given by the manufacturer it is possible to obtain both the panoramic images and an individual image for each camera. Depending on the RGB sensor type, information from IMU and GPS sensors can be used to calculate the position of every image.
GPR	The data from GPR is provided in a binary file format, that varies depending on the manufacturer of the system such as: • Sensors & Software: *.dt1, *.hd, *.gps • MALA geoscience: *.rd3, *.rd7, *.rad, *.cor • Impulse radar: *.iprb, *.iprh, *.cor, *.time, *.mrk • GSSI: *.dzt, *.dzx • DS: *.dt, *.gec • SEG-Y: *.sgy • Radar systems, Inc.: *.sgy • 3dradar: *.3dra, *.vol • ASCII: *.txt (3 column format; X, t, amplitude)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

Technique	Expected raw output data format
UAVs	UAVs can carry different sensors as their payload, such as cameras, lidar or radar. Therefore, the expected outputs can be of different types of data, including: • Numeric data • Alphanumeric data • Images The above data could be stored in formats as: • Text files • Binary files • Database • Images in formats as RAW, TIFF or JPG There are no any particular standard described for UAVs.

3.2.2 Data security

Clearly data security is a topic of paramount importance related to data. This section summarizes the main concepts and practices that have been treated in more depth in IM-SAFE document D4.1. For more details, the reader is therefore referred to Chapter 5 “Data security” of the D4.1 IM-SAFE dodument.

The fundamental principles in data security are known as the “CIA triad” (12) *confidentiality*, *integrity*, and *availability*.

Figure 3.2 - The CIA triad of information security (image from (13))

- **Confidentiality** is concerned with preventing unauthorized access to sensitive data. The two main ways to guarantee confidentiality are *data encryption* (which prevents unauthorized access by rendering the data unusable by anyone other than an authorized user), and *data access control* (which, analogously to *physical access control,* selectively restricts access to data to a set of authorized users who can exhibit appropriate *credentials*).
- **Integrity** is concerned with assuring the accuracy and completeness of data over its entire life-cycle (14), and in particular protecting data from improper modification.

More generally, integrity from a broader information security perspective touches upon aspects that have to do with guaranteeing credibility, completeness, accuracy and trustworthiness of the data (15).

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171.

- **Availability** has the goal of ensuring that authorized users have uninterrupted access to the data that they needed. Ensuring availability also involves preventing service disruptions due to hardware failures and power outages, as well as cyber attacks such as denial-of-service attacks (16).

Data Access Control is a key component to implement the protect function of risk management strategies that aim to address the confidentiality and integrity principles of the CIA triad. The way Data Access Control achieves that is by guaranteeing that only authorized users with appropriate credentials have read and write access on the data.

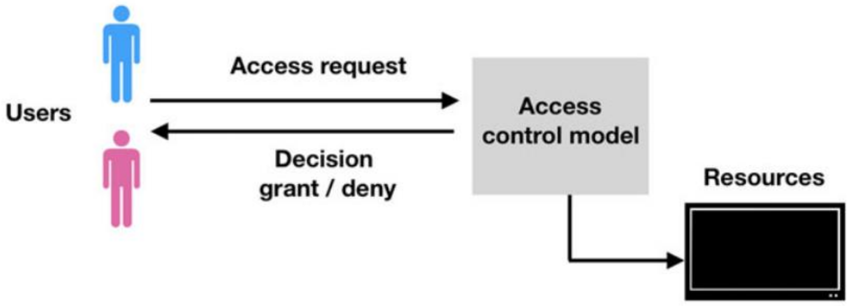


Figure 3.3 - Access control (from (17))

Data Access Control is based on three key mechanisms: *identification*, *authentication*, and *authorization* (17).

Figure 3.4 - Basic mechanisms of Data Access Control (from (17))

Please refer to IM-SAFE document D4.1, Chapter 5 “Data security” for more details on how Data Access Control mechanisms are implemented.

3.2.3 Data security issues in cloud computing

Cloud computing has many advantages compared to more traditional of provisioning IT infrastructures. One of these advantages is the data security aspect, since the majority of companies reported experiencing better security in the cloud than on-premises (18).

Still, there are security issues that can emerge or become of particular interest specifically in a cloud environment. IM-SAFE document D4.1, Chapter 5 “Data security” also goes into detail in these considerations, and the reader is referred to the corresponding section in that chapter for more information. Here we summarize the main considerations with regard to the cloud data life cycle.

In cloud infrastructures *data storage* is typically distributed over multiple data centres and *data processing* can happen in virtual machines that are also being deployed across possibly multiple physical servers. This geographically distributed character of cloud computing comes with its sets of new regulatory and security considerations, since transfer of data calculation and storage to a third part involves the transfer of responsibility associated with their security and compliance to this third part (19,20).

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

The processes involved in distributing storage and data processing across servers implied in on-cloud data processing can be broken down into three distinct states relating to the data in a cloud infrastructure, each one with its specific security risks:

- **Data-at-rest**, where the data is stored in a low cost but high latency (i.e. slow) medium. Data-at-rest in cloud computing is subject to risks that fall into three categories:
    - *Risks associated with storage media sharing,* which stem from the fact that physical infrastructure is shared between different users, meaning that a security breach might simultaneously affect multiple users (see (21) for more details)
    - *Risks associated with data location,* which derive from the fact that data is potentially stored in different locations distributed around geographies. In this case, users have to be aware of the consequences of their data being stored in specific locations, like the fact that they are subject to the local laws that might confer to the local authorities to access the data under given circumstances
    - *Risks related to storage media reliability,* these risks come from the fact that users do not control physical access to data on a cloud storage platform. This means that users will depend on the cloud provider to secure their data, and guarantee availability.
- **Data-in-transit**, which denotes data being transferred for instance from storage to a processing virtual machine, and back. Data-in-transit are typically subject to higher risks than data-at-rest, because data relocation compounds the *risks associated with data location* across all the locations across which the data transits. Secondly, risks associated with data transmission have to be considered (see following section "Requirements for fast and secure data transmission").
- **Data-in-use**, which refers to data being accessed and processed. When data is being processed on cloud infrastructures, the risks of misuse increase, due to the large number of users being hosted on the same infrastructure. An example of such threats is a problem known as *data remanence*, which refers to the fact that upon deletion data residues might still persist on the cloud infrastructure. Consequently, subsequent users gaining access to the same physical storage resources could in principle be able to restore the data.

3.3 Data to functions

3.3.1 Data lifecycle layers

In order to support the analytics required by the use case and business case under consideration data needs to be stored and prepared. This data lifecycle is divided in 3 layers; these 3 layers can be combined to support different needs.

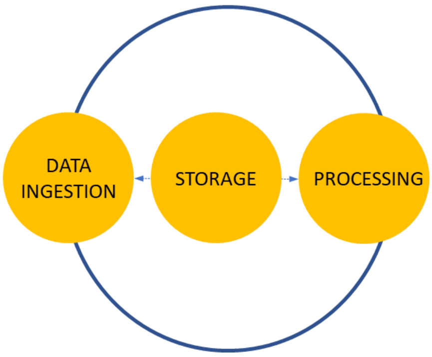


Figure 3.5 - The layers describing the data lifecycle

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

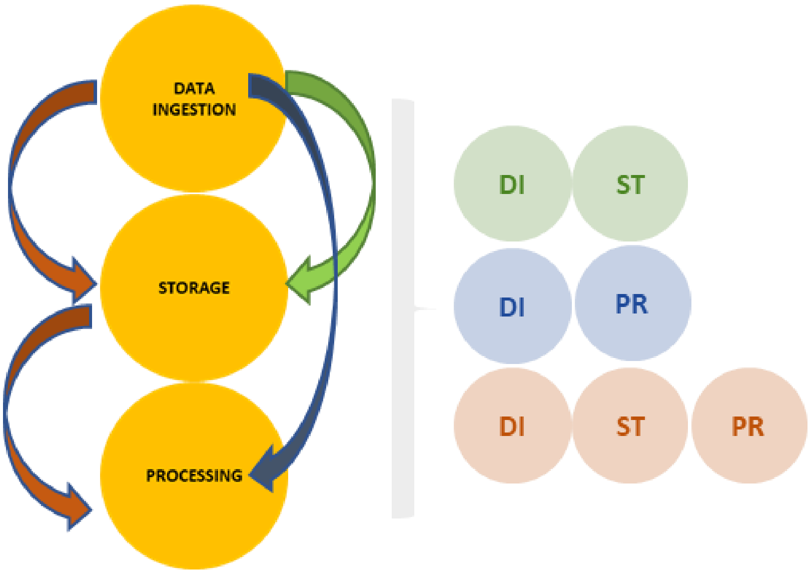


Figure 3.6 - Relation between the layers

- **Data ingestion** refers to the process of transporting data from various sources – sensors, IoT devices, databases, and SaaS applications — into storage. There are two basic ways in which this is done, depending on the type of data and the downstream processing to which the data will be subject: *batch* or *streaming (or real time ingestion) (22)*.
  - **Batch data ingestion** covers the cases where the data is moved from the data sources to the data target on scheduled intervals, like on a daily or weekly basis.
  - **Streaming data ingestion** is the data ingestion mode of choice when the data collected is time-sensitive. For example, in the case of a sensor monitoring critical infrastructure whose sudden change of state might trigger the need of an emergency response.

Below we provide a list of open source data ingestion frameworks (under the Apache license) that are commonly used, along with their main characteristics and surveyed in reference (23). Please refer to the reference for more details.

Tool	Prominent usage	Performance	Integration tools	Type of stream supported
Apache Kafka	For ingesting data, no computations	2 Million writes per second using 03 clusters	All prominent languages	Real time
Apache NIFI	For ETL purposes with vizualization	1000 events per second using 03 nodes cluster	Language, Apache Spark, Apache Kafka, Apache Atlas	Batch, near to real, real time
Amazon kinesis	For videos analytics	3000 writes per second using 03 node cluster	Java, Python, Ruby, Node.js,.NET, Amazon RedShift, Amazon Dynamo Database, Amazon S3	Real-time streaming
Apache flink	For ETL as well as business purposes	4 million records per second using 01 node cluster	Apache Kafka, Elastic Search, HDFS, RabbitMQ, Amazon Kinesis Streams, Twitter, Apache NIFI, Apache Cassandra, Apache Flume	Batch and real time
Apache spark	For ETL as well as business needs	approx 2.5 million records per second using 01 node cluster	All prominent languages, big data tools as well as frameworks	Batch and near to real time
Apache storm	Only for data computation purposes	aprrox 01 million tuples processing per second using 01 node cluster	With all queuing tools like Kafka, Kinesis, JMS and database systems like SQL, MongoDb, etc.	Real time
Apache gobblin	Data ingestion, cleansing, integration	approx 100 TB/day using three clusters	Majorly Java connectors, Apache Kafka	Batch and real time

Figure 3.7 - Access Major data ingestion frameworks, use cases and performance comparison chart (from (23))

- **Data storage** is the step in the data lifecycle that follows ingestion. It deals with the challenge of keeping the data in the right place based on usage so as to remain available

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

upon demand. In particular, it aims to prevent inactive or historical data from being lost and to be kept in regulatory compliance, for chronological reference, or because it can provide analytical value over time. Data storage needs scalable solutions as network-based storage, possibly supporting data sharing and collaboration. More in general, the data storage layer is tasked to store the data in data structure that are can accommodate the downstream processing of the data. For instance, if the downstream operations will require relational persistence, then the data has to be stored in a relational database.

- **Data processing** is the layer of the data operation pipeline which essentially defines the whole pipeline, in that this is where the data flow converges to analytics. Clearly, the data processing layer will be completely defined by the application and use case supported by the platform. The main characteristic of interest in this sense will be whether the processing of the data will have to happen in real-time stream fashion or can perform in batch. In the first case, obviously the prerequisite is that the ingestion and storage layers support the streaming mode. Below we list some features distinguishing batch processing from real-time processing that will be useful on deciding which modality is suited to support the downstream application and use cases.

Batch processing	Real-time stream processing
Discrete chunk of data at a certain interval	Import data when it is produced
Prominent in traditional analytics	Prominent in real-time analytics in current trend
Collection and loading of information are two different tasks	Collection, loading, and processing of data are under one umbrella
Output or analytics is based on history data	Output or analytics is based on real-time or current data
Store first before processing	Processing can be first even before storage
Data volume is generally specific	Data volume is dependent on tool being used
Decision making is much slower than real-time processing tools	Real-time decision-making capability

Figure 3.8 - Comparison of batch data processing versus real-time stream processing (from (23))

3.4 From functions to usability and users

3.4.1 Importance of a user-centric design

The performance of a data infrastructure can be measured based on functionalities that it provides to its end-users and its efficacy and usefulness is evident when it empowers them to be more productive and effective in their tasks.

This is the basic recognition of User-Centric Design (UCD) which is a set of methodologies aimed at gaining a deep understanding of who will be using a product in order to adapt the product development accordingly. The international standard 13407 (24) is the basis for many UCD methodologies. The key tenet of UCD is that users should be involved throughout the design and development lifecycle and such involvement should be iterative.

The following are the general phases of the UCD process:

- **Specify the context of use**: Identify the people who will use the product, what they will use it for, and under what conditions they will use it
- **Specify requirements**: Identify any business requirements or user goals that must be met for the product to be successful
- **Create design solutions**: This part of the process may be done in stages, building from a rough concept to a complete design

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

- **Evaluate designs**: Evaluation – ideally through usability testing with actual users – is as integral as quality testing is to good software development.

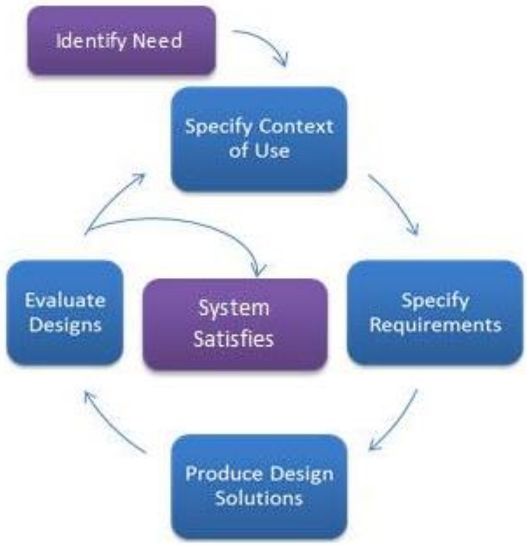


Figure 3.9 - The User-Centric Design process (from (25))

There are several common techniques supporting UCD at various stages of the design and development process of a product. Figure 3.10 lists some of the main ones, several of which are based on the idea of interviewing users.

This is akin to the idea of user stories, a common tool in software development and product management, which we are examining next.

Technique	Purpose	Stage of the Design Cycle
Background Interviews and questionnaires	Collecting data related to the needs and expectations of users; evaluation of design alternatives, prototypes and the final artifact	At the beginning of the design project
Sequence of work interviews and questionnaires	Collecting data related to the sequence of work to be performed with the artifact	Early in the design cycle
Focus groups	Include a wide range of stakeholders to discuss issues and requirements	Early in the design cycle
On-site observation	Collecting information concerning the environment in which the artifact will be used	Early in the design cycle
Role Playing, walkthroughs, and simulations	Evaluation of alternative designs and gaining additional information about user needs and expectations; prototype evaluation	Early and mid-point in the design cycle
Usability testing	Collecting quantities data related to measurable usability criteria	Final stage of the design cycle
Interviews and questionnaires	Collecting qualitative data related to user satisfaction with the artifact	Final stage of the design cycle

Figure 3.10 - Main technique of user-centric design (from (26))

3.4.2 User stories as a tool to relate functions and user personas

User stories are a commonly adopted tool in software development and product management to either define, document or create a common understanding around the relationship between

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

a system or product and its end users (27). In practice, a user story is an informal description of features of a software system from the perspective of a user of a system with the scope of exposing the motivation, needs and goals of the user in order to elucidate the systems role in relation to the user's workflow.

Below an example of user story is described in the contest of analysing or designing a typical data platform for Structural Health Monitoring of the transport infrastructure which captures many of the features and goals of the platforms covered in Chapter 2.

- **User persona:** *Inspector*
  - **Goals**:
    - Provide documentation of defects
    - Provide health scores aggregated across structural elements
  - **Tasks**:
    - Inspect structure to document defects position, shape and length on structural elements
      → need: retrieve data about structure
    - Tabulate defects from a predefined items catalogue relating materials to defects
      → need: retrieve database about material and defects
    - Assess evolution of defects in the context of retrieved historical data from previous inspections
      → need: retrieve data about historical inspections
    - Gather pictures to document longitudinal temporal evolution of defects
      → needs to store and share large quantities of picture data
    - Co-register pictures with structure plan and geo-localization of focal points
      → need: access to geo-location data and annotation tools on recorded data

- **User persona**: *Data analyst*
  - **Goals**:
    - Relate sensor readouts to structural status
    - Enable structural engineers to use data for structural analysis
    - Perform analytics on data
  - **Tasks**:
    - Preprocess sensor readouts to relate raw data to physics, structural and mechanistic quantities
      → need: access to sensor readouts, storage, and processing layers
    - Perform statistical analyses
      → need: access to processing layer
    - Identify outliers in the data to help discriminate noise from rare events
      → need: access to processing layer
    - Communicate analysis results to structural engineer
      → need: share data with other platform users

- **User persona**: *Structural engineer*
  - **Goals**:
    - Relate sensor readouts to structural status
    - Provide structural health assessment
  - **Tasks**:
    - Gather historic monitoring data and inspector outputs about the structure under analysis
      → need: retrieve data about historical inspection
    - Provide structural analysis of data by computing mechanical quantities
      → need: access to processing layer
    - Access to knowledge of sensors (sampling rate, device characteristics, etc) to validate the relation between acquired sensor data and mechanical quantities

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

→ *need: access to database of device characteristics and access to sensor readouts*

- **User persona**: *Asset manager*
  - **Goals**:
    - Evaluate structural performance and functional status of asset
  - **Tasks**:
    - Evaluate structural health assessment
      → *need: access to reports provided by structural engineer*
    - Evaluate and respond to alerts
      → *need: real-time streaming access to output of alarm system.*

A great feature of user stories, beside grounding the understanding of the product development team in a common user-centric framework, is that they help to expose the interdependencies among users, functions supporting the users' tasks, and the needed resources. This not only guides the design of the user interface (UI), but also indicates to which data layers each user persona needs access, and what requirements these data layers should satisfy. For instance, one would naively assume that, being at the end of the data lifecycle, an *asset manager* would only need access to highly elaborated data downstream from all processing layers and that it would therefore be enough to offer long-latency batch access. In fact, asset manager might also need access to streaming data from sensors feeding into real-time alarm systems. The data platform will therefore be designed so as to accommodate these types of technical needs.

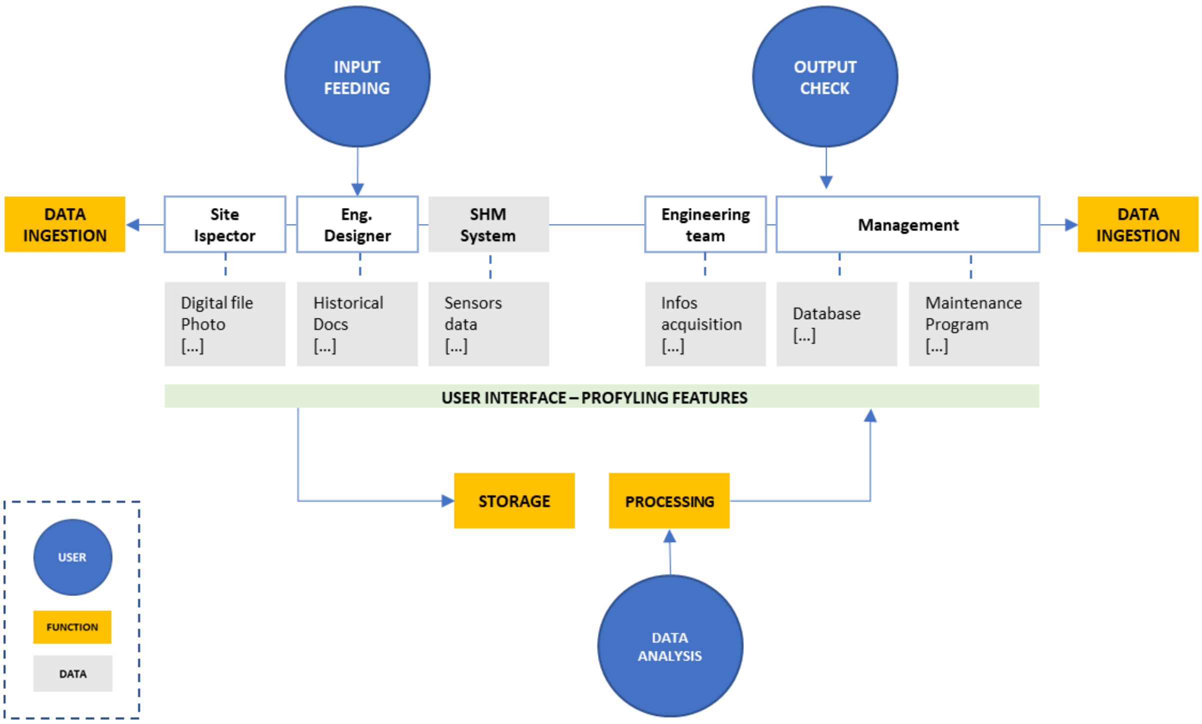


Figure 3.11 - Diagram illustrating the relations between user personas, data types, and data functions in a typical data platform for Structural Health Monitoring

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

4 Conclusions and recommendations

4.1 General considerations

This chapter (based on reference (6) of many of the authors of this chapter) will briefly elaborate on a set of functionalities and usability considerations that have been identified as essential in order to design a data platform that would best support the users on the platform and the stakeholders in the domain for the case of applications in the transport infrastructure inspections and monitoring sector.

The main considerations have to do with the needs and requirements of different user personas being very diverse but also highly interdependent from those of other user personas, and these interdependencies evolving themselves dynamically in response to changes in technology, regulation, business priorities, etc.

Being able to cater to the individual needs in such a diverse and evolving landscape requires designing a platform explicitly geared towards two main *usability desiderata:* **broad accessibility** on one hand, and **extensibility and specialization** on the other one.

- **Broad accessibility** means that the platform should be suitable for all user personas at their level of order to enable practitioners with no coding, machine learning analytics, hardware deployment or framework-specific expertise to explore their data, synthesize models, analyse results.
- **Extensibility and specialization** aims to make sure that the platform is useful for data science, engineers, inspectors, asset managers and machine learning researchers as well, both as working environment to conduct research and development of new uses of the data, but also as a means to quickly integrate research products in order to *extend the platform with the latest cutting-edge developments* and *specialize it to novel emerging needs,* be they driven by changes in technology, regulation, or business considerations.

This chapter will indicate some recommendations aimed at addressing these accessibility desiderata in mind during the design of the data platform.

4.2 Recommendation 1: User-centric design

Developing a platform with a User-Centric Design approach in mind prompts to focus on the user personas of the platform. On the other hand, the material and list of data platforms for the transport infrastructure sector gathered so far in this chapter suggest that this might be a challenging endeavour, since a data platform in the civil engineering and the transport infrastructure sectors typically has to support a broad diversity of user personas, with a diverse set of technical requirements and needs.

The first recommendation for choosing or designing a data platform for a domain like the transport infrastructure sector with such a diversity of the highly technical user personas is to have a user-centric approach and start by identifying the main stakeholders in the domain and the main type of users. This can be done by means of *interviews* and *user stories*, whose scope is to define the user personas that will be using the platforms, their tasks, the functional needs to support those tasks, the type of access to the data lifecycle that they need, and the interdependencies between them. These interdependencies will tend to look analogous to those between the steps of a typical machine learning lifecycle as identified by the CRISP-DM Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (41) (see Figure 4.1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

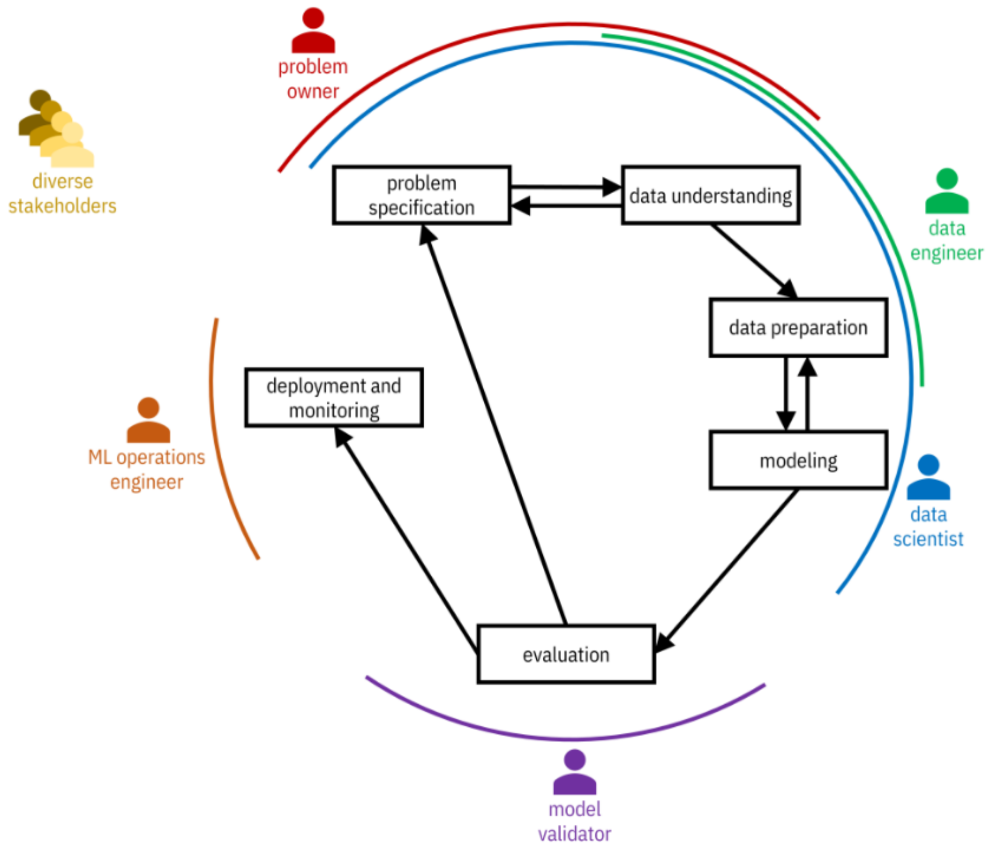


Figure 4.1 - High interdependency among the steps of the machine learning lifecycle codified in Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (28). Different personas participate in different phases of the lifecycle (from (29))

4.3 Recommendation 2: Automate resource allocation

The main goal of the platform is to enable users to carry out their tasks without being obstructed by the technical implementation details that have to do with the platform itself. In other words, all the backend infrastructure of the platform should be rendered invisible to the user, that should be able to work on top of an abstraction layer that directly addresses the technical needs to address their tasks.

An important way to achieve that is to free the user from having to worry about how to provision compute infrastructure. This will be particularly important considering that most users will not have the technical expertise to deal with hardware resource provisioning nor should they be expected to subtract time from their main tasks for that. This translates into the recommendation to design the data platform such that it implemented mechanisms for *automating resource allocation*.

Automated resource allocation is a challenge that has been tackled for instance by (30–32). This line of work was technically supported by studies that investigated the suitability specifically of cloud infrastructure for specific ML workloads (33–39).

Other important developments to lower the barrier for entry to large-scale machine learning were to provide generic ML as a Service (MLaaS) (40), and to allow users to access a prediction service through an API and a Graphical User Interface as demonstrated by (41).

Some of these considerations have been implemented in several existing commercial platforms among which some of the most popular are: IBM Cloud Pak for Data (42) (an end-to-end ML platform available on IBM Cloud which integrates IBM Watson Studio for model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

automation), Google Cloud AI (43), Azure Machine Learning (44) and Amazon SageMaker (45) (enterprise-grade platforms for the end-to-end ML lifecycle from Google, Microsoft and Amazon, respectively); H2O.ai (46), an open source end-to-end ML platform).

All mentioned commercial platforms provide Graphical User Interfaces that abstract away resource allocation as well as several aspects of the selection of the appropriate algorithms and model hyperparameters (i.e. fixed parameters of the model) for a particular task of interest.

That helps considerably lowering the barrier of entry into the utilization of ML methods to practitioners that are not expert in data science and managing its supporting hardware infrastructure.

4.4 Recommendation 3: Modular and extensible User Interface (UI)

In order to guarantee broad accessibility, end-users should have to be able to fully operate the platform from an intuitive UI, similarly to the majority of other commercial platforms. However, handling completely different tasks within a fixed user interface does not guarantee the best user experience: for example, exploring a civil infrastructure dataset composed of many annotated images organized in a hierarchical structure, is more efficient with the availability of specific navigation tools which exploit this particular structure, if compared to exploring a generic object detection dataset with standard tools.

This prompts another recommendation regarding the UI/UX design: the user interface should be built with *modularity* and *extensibility* in mind so as surface to the users the resources and data layers that they need at the abstraction level that is most appropriate whenever the needs arise.

As an example, some data platforms (like for example IBM OCL) provide a graphical user interface suited for high-level interaction with abstract data objects, and, in addition to that, experienced users that require programmatic access to the platform are provided with a REST API that gives control on all parameters and enables programmatic interaction with the platform. This REST API allows quick integration with new functionalities or other data platform, which might be essential when new integration needs arise.

4.5 Final Recommendation: Adherence to data security and data governance best practices

The final recommendation is the first in order of importance, and that is that the data platform should be built from the ground up to strictly adhere to the industry's best practice in terms of data *security* and *data governance.* For more details on security, please see the "Data Security" chapter in D4.1 (summarized in the previous chapter of this document).

The recommendation to adhere to *data governance best practices* reflects very much the recommendations identified by the D4.2 on data integration and interoperability. The main challenges arising from designing a data platform in a domain like civil engineering and the transport infrastructure industry have to do with on one hand having to accommodate a large variety of established workflows and legacy data formats and software, and on the other hand being ready to incorporate changes in the technology and evolution in use cases.

The recommendation of D4.2 elaborated in such a context where data is diverse, heterogeneous and distributed across many (legacy) systems were to first of all realize that "A single system or solution will not be able to deal with all [...] use cases. Accordingly, implementation must be based on open, robust and easily extensible solutions". Supporting **open standards** is in other words a recommended strategy to both ensure compatibility with current systems and data, as well as a way to guarantee the longevity of future Information and Communication Technoslogy (ICT) systems. As elaborated in D4.2, the principles of transparency and openness at the root open standards are in fact expected to help drive

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

innovation in a way that also fairly takes into account already established workflows and practices.

Moreover, as D4.2 pointed out, “There is not a single standard that covers all data being of interest for monitoring scenarios. It is thus expected that a combination of ICT standards will be needed to support semantic data integration”. In other words, standardization will require the adoption and or definition of multiple standards, in accordance to the multiplicity of scenarios, use cases, and technical needs that the technology will have to accommodate.

Other aspects of *data governance* related to the use of data in machine learning applications are going to be covered in the chapter on “Trustworthy Adoption of AI/ML models” in D4.4.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171.

5 References

1. Dobbe RIJ, Gilbert TK, Mintz Y. Hard Choices in Artificial Intelligence: Addressing Normative Uncertainty through Sociotechnical Commitments. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2022 Apr 11]. p. 242. Available from: https://doi.org/10.1145/3375627.3375861

2. Ministero delle Infrastrutture e dei Trasporti, Consiglio Superiore dei Lavori Pubblici. Linee guida per la classificazione e gestione del rischio, la valutazione della sicurezza ed il monitoraggio dei ponti esistenti. Ministero delle Infrastrutture e dei Trasporti, Consiglio Superiore dei Lavori Pubblici: Roma, Italy. 2020;

3. RFI. Ispezione e manutenzione per la durabilità dei ponti: criteri di gestione, contesto normativo, metodi di intervento. Milano, Italy: Rete Ferroviaria Italiana; 2019 Mar.

4. RFI. Dalla manutenzione ciclica alla manutenzione predittiva. Italy: Rete Ferroviaria Italiana; 2020 Nov.

5. RFI. Sperimentazione sull'impiego dei droni nell'ispezione dei ponti ferroviari. Milano, Italy: Rete Ferroviaria Italiana; 2015 Oct.

6. Ministero dell'Infrastruttura e della mobilità sostenibile. AINOP [Internet]. 2022 [cited 2022 Apr 19]. Available from: https://ainop.mit.gov.it/portale/#/

7. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer; 2014. p. 740–55.

8. Bartezzaghi A, Giurgiu I, Marchiori C, Rigotti M, Sebastian R, Malossi C. Design of a Cloud-Based Data Platform for Standardized Machine Learning Workflows with Applications to Transport Infrastructure. In: IEEE MELECON. 2022.

9. Bisong E. Kubeflow and kubeflow pipelines. In: Building machine learning and deep learning models on google cloud platform. Springer; 2019. p. 671–85.

10. Relatics. Relatics [Internet]. 2022 [cited 2022 Apr 19]. Available from: https://www.relatics.com/en/

11. Ministerio de Transportes, Movilidad y Agenda Urbana. Plataforma CELOSÍA [Internet]. 2022 [cited 2022 Apr 19]. Available from: http://www.celosia.es/

12. Dhillon G, Backhouse J. Current directions in IS security research: towards socio-organizational perspectives. Information systems journal. 2001;11(2):127–53.

13. Purcell A. 3 key ideas to help drive compliance in the cloud [Internet]. Cloud computing news. 2018 [cited 2022 Apr 13]. Available from: https://www.ibm.com/blogs/cloud-computing/2018/01/16/drive-compliance-cloud/

14. Boritz JE. IS practitioners' views on core concepts of information integrity. International Journal of Accounting Information Systems. 2005;6(4):260–79.

15. Sebastian-Coleman L. Measuring data quality for ongoing improvement: a data quality assessment framework. Newnes; 2012.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171.

16. Loukas G, Öke G. Protection against denial of service attacks: A survey. The Computer Journal. 2010;53(7):1020–37.

17. Andress J. The basics of information security: understanding the fundamentals of InfoSec in theory and practice. Syngress; 2014.

18. McAfee. New McAfee Report Finds Eighty-Seven Percent of Companies Experience Business Acceleration from Use of Cloud Services [Internet]. 2019 [cited 2022 Apr 13]. Available from: https://www.businesswire.com/news/home/20190617005945/en/New-McAfee-Report-Finds-Eighty-Seven-Percent-of-Companies-Experience-Business-Acceleration-from-Use-of-Cloud-Services

19. Singh VK, Singh T. Present data security issues and their resolving technique in cloud computing. Int J Sci Technol. 2016;1:1–6.

20. Kacha L, Zitouni A. An Overview on Data Security in Cloud Computing. In: Silhavy R, Silhavy P, Prokopova Z, editors. Cybernetics Approaches in Intelligent Systems [Internet]. Cham: Springer International Publishing; 2018 [cited 2022 Apr 13]. p. 250–61. (Advances in Intelligent Systems and Computing; vol. 661). Available from: http://link.springer.com/10.1007/978-3-319-67618-0_23

21. Singh S, Jeong YS, Park JH. A survey on cloud computing security: Issues, threats, and solutions. Journal of Network and Computer Applications. 2016;75:200–22.

22. Mătăcuţă A, Popa C. Big data analytics: Analysis of features and performance of big data ingestion tools. Informatica Economica. 2018;22(2):25–34.

23. Sharma G, Tripathi V, Srivastava A. Recent Trends in Big Data Ingestion Tools: A Study. In: Kumar R, Quang NH, Kumar Solanki V, Cardona M, Pattnaik PK, editors. Research in Intelligent and Computing in Engineering. Singapore: Springer; 2021. p. 873–81. (Advances in Intelligent Systems and Computing).

24. ISO. ISO 9241-210:2010 [Internet]. ISO. 2010 [cited 2022 Apr 15]. Available from: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/20/52075.html

25. USGSA. User-Centered Design Basics | Usability.gov [Internet]. Department of Health and Human Services; 2017 [cited 2022 Apr 15]. Available from: https://www.usability.gov/what-and-why/user-centered-design.html

26. Abras C, Maloney-Krichmar D, Preece J. User-centered design. Bainbridge, W Encyclopedia of Human-Computer Interaction Thousand Oaks: Sage Publications. 2004;37(4):445–56.

27. Dimitrijević S, Jovanović J, Devedžić V. A comparative study of software tools for user story management. Information and Software Technology. 2015;57:352–68.

28. Berwind K, Bornschlegl M, Hemmje M, Kaufmann M. Towards a cross industry standard process to support big data applications in virtual research environments. CERC2016. 2016;82.

29. Varshney KR. Trustworthy machine learning. Chappaqua, NY, USA: Independently Published; 2022.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

30. Messina F, Pappalardo G, Rosaci D, Santoro C, Sarné GM. A trust-aware, self-organizing system for large-scale federations of utility computing infrastructures. Future Generation Computer Systems. 2016;56:77–94.

31. Caballer M, Zala S, García ÁL, Moltó G, Fernández PO, Velten M. Orchestrating complex application architectures in heterogeneous clouds. Journal of Grid Computing. 2018;16(1):3–18.

32. Salomoni D, Campos I, Gaido L, de Lucas JM, Solagna P, Gomes J, et al. INDIGO-DataCloud: A platform to facilitate seamless access to e-infrastructures. Journal of Grid Computing. 2018;16(3):381–408.

33. Pallickara S, Ekanayake J, Fox G. Granules: A lightweight, streaming runtime for cloud computing with support, for map-reduce. In: 2009 IEEE International Conference on Cluster Computing and Workshops. IEEE; 2009. p. 1–10.

34. Catak FO, Balaban ME. CloudSVM: training an SVM classifier in cloud computing systems. In: Joint International Conference on Pervasive Computing and the Networked World. Springer; 2012. p. 57–68.

35. Li K, Gibson C, Ho D, Zhou Q, Kim J, Buhisi O, et al. Assessment of machine learning algorithms in cloud computing frameworks. In: 2013 IEEE Systems and Information Engineering Design Symposium. IEEE; 2013. p. 98–103.

36. Lin X, Wang P, Wu B. Log analysis in cloud computing environment with Hadoop and Spark. In: 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology. IEEE; 2013. p. 273–6.

37. Meng X, Bradley J, Yavuz B, Sparks E, Venkataraman S, Liu D, et al. Mllib: Machine learning in apache spark. The Journal of Machine Learning Research. 2016;17(1):1235–41.

38. Preethi G, Krishna PV, Obaidat MS, Saritha V, Yenduri S. Application of deep learning to sentiment analysis for recommender system on cloud. In: 2017 International conference on computer, information and telecommunication systems (CITS). IEEE; 2017. p. 93–7.

39. Ishakian V, Muthusamy V, Slominski A. Serving deep learning models in a serverless platform. In: 2018 IEEE International Conference on Cloud Engineering (IC2E). IEEE; 2018. p. 257–62.

40. Ribeiro M, Grolinger K, Capretz MA. Mlaas: Machine learning as a service. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE; 2015. p. 896–902.

41. Chan S, Stone T, Szeto KP, Chan KH. Predictionio: a distributed machine learning server for practical software development. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. 2013. p. 2493–6.

42. IBM. IBM Cloud Pak for Data as a Service [Internet]. Available from: https://www.ibm.com/products/cloud-pak-for-data/as-a-service

43. Google. Google Cloud AI [Internet]. Available from: https://cloud.google.com/products/ai

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. © 2021

44. Microsoft. Azure Machine Learning [Internet]. Available from: https://azure.microsoft.com/en-us/services/machine-learning/

45. Amazon. Amazon SageMaker [Internet]. Available from: https://aws.amazon.com/sagemaker/

46. H2O.ai. H2O.ai [Internet]. Available from: https://www.h2o.ai

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958171.