LODE PROJECT

Conditions of replicability and interoperability with existing databases

–

Deliverable n° 3.3
Delivery date 15.07.2021

Reference N° 826567

LODE Loss Data Enhancement for DRR & CCA management

Project acronym	LODE PROJECT
Contract Number	826567

Authors: Anna Faiella	
Partner	Polimi

Versioning			
Version	Date	Name	Organization
Version 01	10/06/2021	Anna Faiella	Politecnico di Milano
Version 02	30/07/2021	Scira Menoni	Politecnico di Milano
Review	30/08/2021	Pierluigi Plebani	Politecnico di Milano
Version 03	08/09/2021	Scira Menoni	Politecnico di Milano

Index

1. Process replicability

Developing an information system is both a science and an art. It is a science as it grounds on the logical steps and splitting of any procedure, operation, structuring of data and information into simpler and unambiguous components that can be translated into algorithms and being processed by a software. Also the coding that is necessary to shape its features and development grounds on computer science and expertise that has evolved tremendously in the last decades, moving up towards increased complexity and corresponding performance. Software nowadays are not written from scratch each time a new program has to be developed, they are building on pieces of software that already exist and combine them in new ways adding and integrating new coding which makes the entire work and result much more complex and challenging. It is an art in the sense that the abstraction and the grouping of elements and components and the identification and characterization of each element to become part of the system is left to the understanding and the capacity to the analyst to fragment a problem or a task without losing the cognition of its overall meaning, its pertaining to an entity.

This understanding has been at the core of the mutual exchange of knowledge and expertise between the LODE partners, most of whom are not computer scientists, and the developers of the system at the Politecnico di Milano with the key implementing role of the external consultant. We may well state that the experience has been somehow unique and original. In most cases we are aware of, software developers ask the domain experts to provide them with adequate requirements for the system to be produced and for its functionalities. In many cases, and we can say this on the basis of our own past experience, users, domain experts do not have a full understanding neither of what a requirement is exactly nor of how they should express such requirements in a way that is usable at best by the developers. On the other hand, computer scientists are not particularly keen to enter too much into the domain, they wish to get the minimal understanding they deem to need in order to provide the information system. In the LODE project a different approach has been taken. It was decided to train partners to represent their knowledge domain in a form that is already schematic and structured enough to be more “comprehensible” to the computer scientist and in the meantime to ask them to make the maximal effort of abstraction and analysis of each sector that has been tackled by the project in detail. Basic questions were “what does constitute damage for each sector and sub-sector?”; “how damage and what damage (physical, indirect) should be considered?”; “what attributes of each component of each sector should be represented as elements of the database?”; “what relations exist between the data that we would like subsequently to query?”. Clearly some partners had more difficulties than others in understanding and uptaking this approach, yet a number of partners actively contributed by replying to the questions above as explained in the previous deliverables 3.1 and 32.

On the other hand, the computer scientists and the developer who actually produced the tool, entered into the knowledge domain in an active and propositional way. The GIS expert and the software engineer already had a solid background on disaster studies and had already both professional and research experience in the domain of risk mitigation, prevention, damage assessment using advanced technologies. The developer had already extensive experience of work carried out for public administrations and could couple experience in developing databases and in GIS.

This unique mix created an environment in which the IS has been the object of continuous work, iterative improvement and checking at each stage of the project. As already explained several times, the information system that has been developed is constituted by an advanced relational database that provides the possibility to store data and to retrieve information for several societal sectors that are affected by disasters. A “sector” is clearly an abstraction of reality, each sector is constituted by sub-sectors, components, individual elements each with their own features. Physical damage does not affect “the sector” but its physical elements and components, functional and systemic damage may affect sub-sectors or the sector itself. As explained in Deliverable 3.1 each sector has been analysed in detail and decomposed in entities and attributes, abstracting from single cases to account for those constant and more frequent aspects that can be recognized in each disaster and lead to damage and losses that can become part of a statistics. The reader can easily see the conceptual advancement that has been accomplished from the first sector that has been treated, namely agriculture, and the last one, buildings that will be briefly presented in the following. Each sector that has been developed in between, roads, power and telecommunication, provided insight and permitted to advance in the path towards standardization and grouping of elements so as to improve the overall structure of the information system. Not only the knowledge on each system and on what constitutes damage to the system has evolved, but also the skill to translate the latter into performing components of the database as well as the identification and the capacity to represent the relationships between the different components, sub sectors and sectors.

In the following the process that has been followed for buildings will be briefly presented, reminding that a more thorough discussion is provided in a submitted paper by Faiella et al.

1.1. Steps of the process to develop the database that are relatively easily replicable

The overall process can be subdivided in three main steps that are somehow subsequent but cyclic revisions are necessary to adjust and to account for aspects and elements that are not always easy to consider and remember right at the beginning devoted to the elicitation of requirements as shown in figure 1.

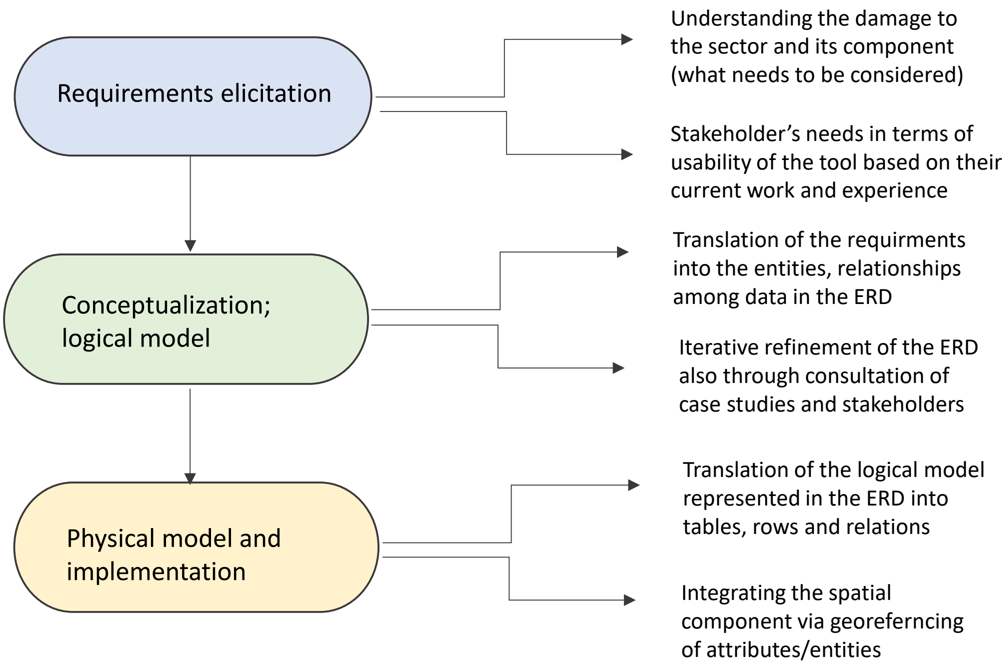

Figure 1. Process to develop a relational database system

1.1.1. Elicitation of requirements and structuring the "sector"

The first step is the identification of requirements. As mentioned this has been done in the LODE project according to what can be labelled as an Agile process, iterative and cyclic through which partners with the help of stakeholders identified both technical and procedural requirements for the database development and also having in mind its subsequent use in real life situations. As the database has been developed in a modular way, after ending one sector, the following one permitted to introduce some novel abstractions and grouping of entities that was then used for the subsequent development. Each sector among the following has been elaborated and requirements have been identified as shown in figure 2.

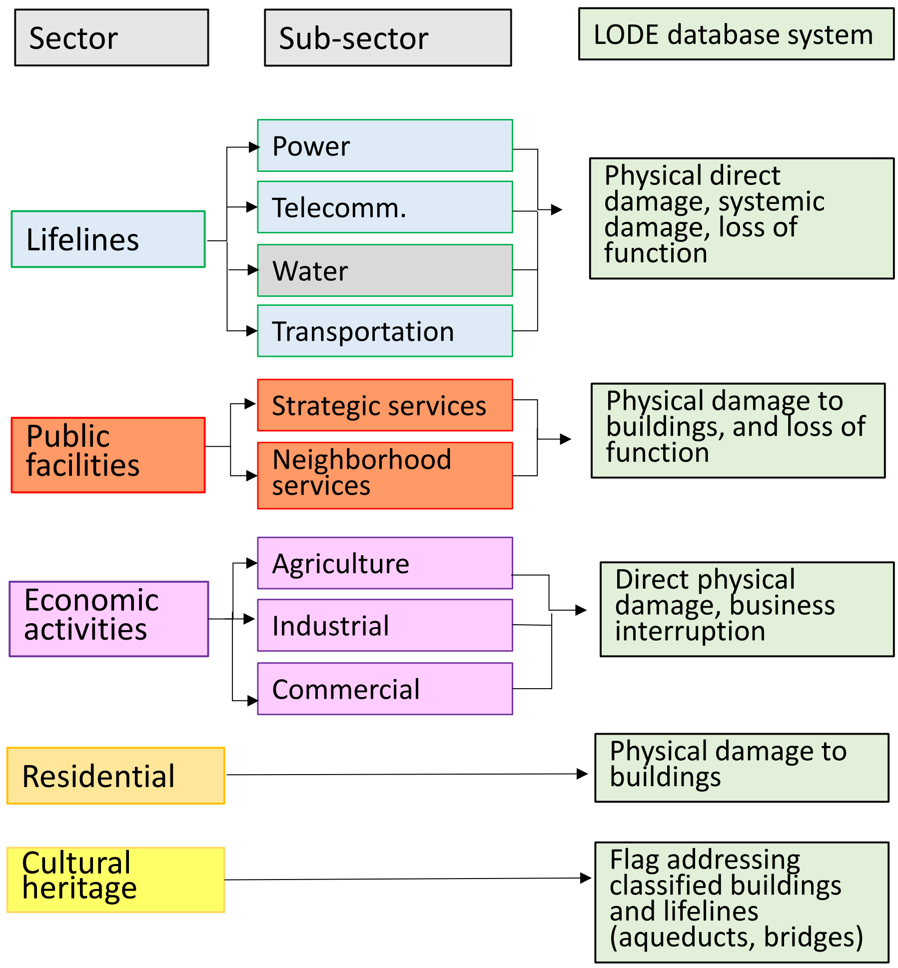

Figure 2. In color the sectors for which the database has been developed, in green it is specified the type of damage that has been considered and that can be retrieved as an information from the database

It must be highlighted that speaking about "sectors" is a rather generic label referring to a rather broad categorization of societal systems providing certain functions to a community. In the LODE project the following sectors have been considered (colored in the diagram): power, telecommunication and roads for lifelines; agriculture, commercial and industrial for economic activities; residential buildings; public facilities and cultural heritage.
As for public facilities and cultural heritage only those hosted in a "ordinary" structure (see figure 4) have been included in the database, whilst special facilities and structures, such as a hospital have not been included. Cultural heritage have not been considered in their entirety and complexity in the development of the database, only for movable assets there is the possibility to flag if the asset is or not a cultural heritage and how listed and classified (Unesco, national, or other classification).

1.1.2. Conceptualization of the logic model of the database system. Example related to buildings

The proposed conceptual model discretizes the main physical, functional characteristics of the damage that might occur after hazardous events to buildings across relevant spatial and temporal scales, to account for potential dynamic changes of the latter as damage unfolds. The model is designed to record and monitor damages in different stages, both in the immediate emergency as well as in the recovery phase due to the possibility to record the dates of both the survey and the damages. Having data regarding different phases after the disaster allows to track the history and dynamics of the damages, which is useful also to monitor the recovery and reconstruction effectiveness.

In the proposed model, buildings are represented as "*entities*", characterized by the so called *"attributes"*, which have been selected in accordance to the knowledge acquired through the showcases of the LODE project, and *"relationships"* representing the interactions between the sector's elements allowing to maintain the link between damages and the time of occurrence.

Entity Relationship Diagrams have been chosen to represent the conceptual model of the database (ERD). The basic elements of an ERD are called entities (represented as rectangles), each entity is described through relative attributes. From a graphical point of view, entities are generally related to each other through rhombus (Chen's notation). Relationships among two entities have cardinality ratios which specify the maximum number of relationship instances that an entity can participate in (..).

The relational database proposed and showed in Figure 3 comprises specific characteristics of the different impacting mechanisms in case of diverse natural hazards, since it is the result of detailed sectoral understanding of damage and its mechanisms (direct physical damage and incurred costs due to repair and retrofitting; systemic consequences due to loss of functionalities placed in the damaged buildings like commercial activities and services).

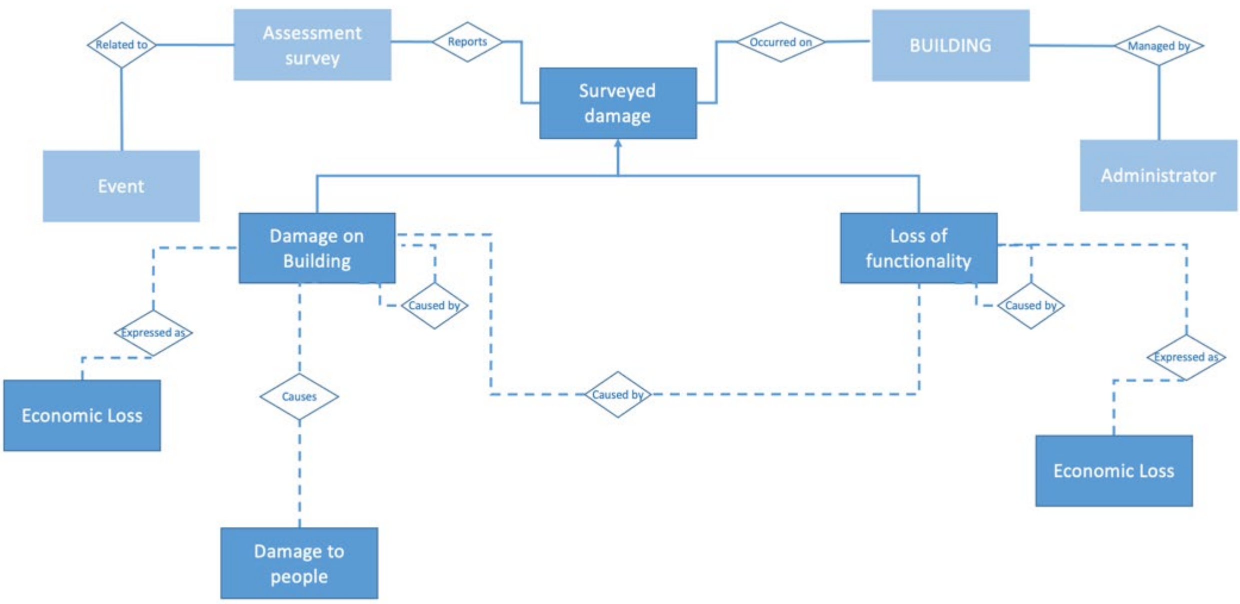

Figure 3. ERD for buildings

In the case of residential buildings the attributes are as represented in Figure 4.
Such attributes relate to the features of the building (that can and should be retrieved from already existing databases, such as cadastral), or prior vulnerability assessment (as for public facilities in Greece and Italy). A category of attributes relate to the use; in fact different floors may be used in different ways, with upper floors residential and ground level where shops and offices are located.

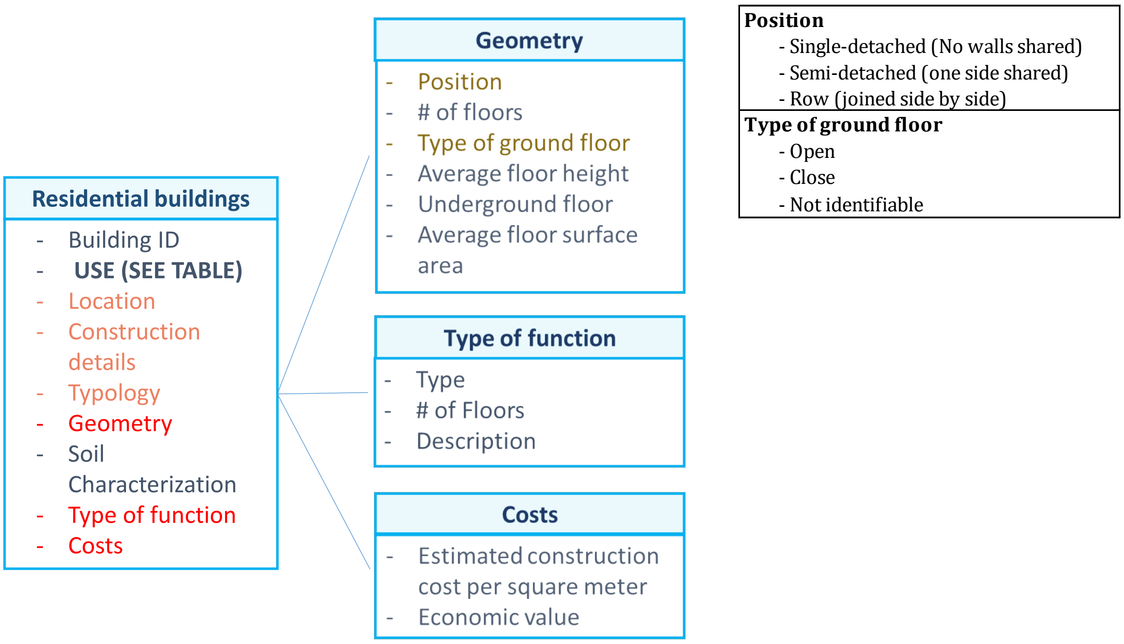

Figure 4. Attributes related to a residential building

In Figure 5, attributes related to the damage to buildings is shown. They are subdivided in overall damage to the structure (partial, total collapse, damage to nonstructural elements) and damage to the individual damaged components.
This information is generally collected by civil protection or by insurance to estimate the overall costs that will be allocated for repair/substitution. Attributes related to the costs are then more general in terms of abstraction and attributes have been already defined as shown in Figure 6.
As for the damage, also a qualitative classification or grading is usually adopted, for example for usability purposes. Based on consultation with the stakeholders working on reporting damage to buildings and Greece and Italy it has been decided to keep this information, also as a link to first level rapid assessments, such as those provided by Copernicus, that also do provide a damage grading.

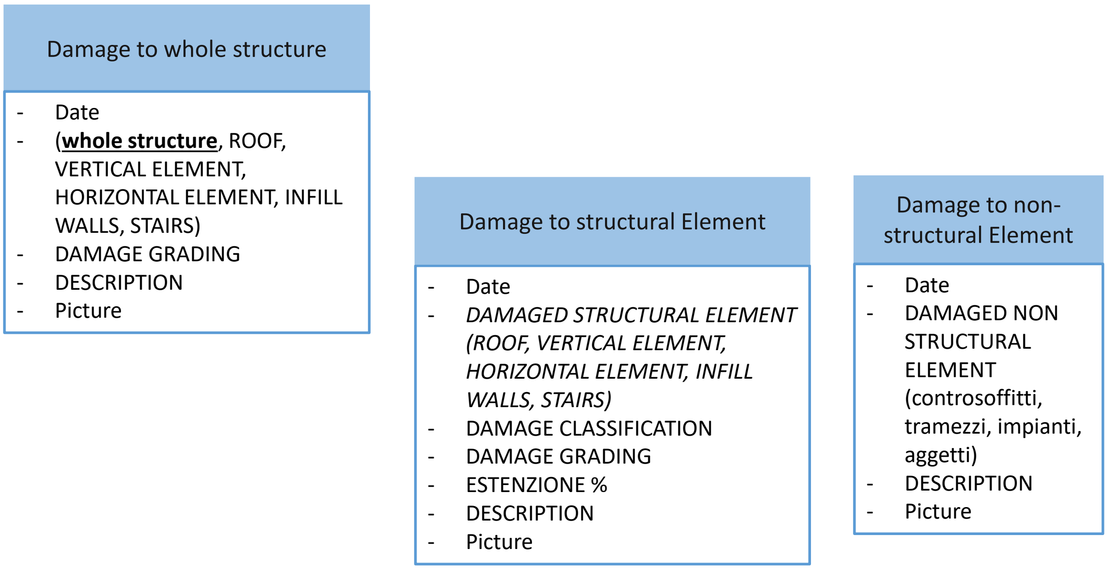

Figure 5. Damage to buildings attributes

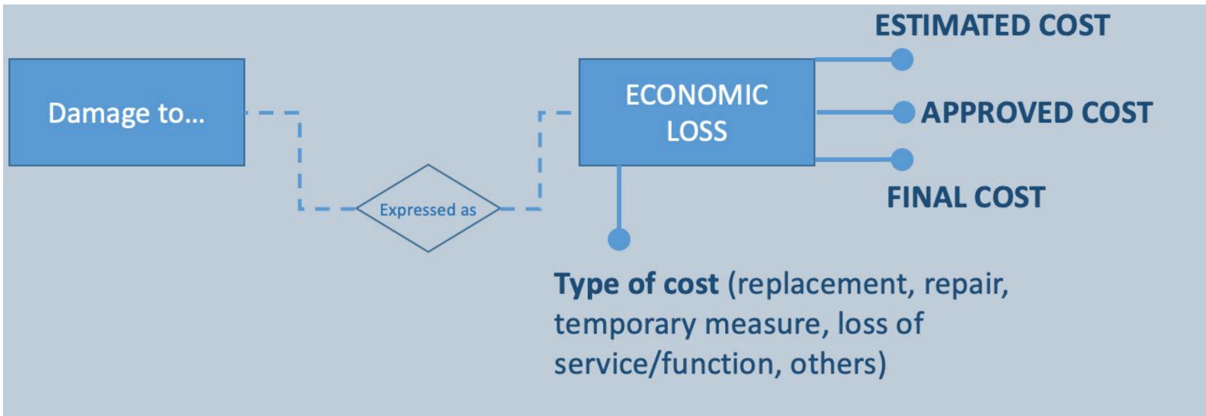

Figure 6. Economic loss corresponding to the surveyed damage (need to repair/substitute)

An important aspect of abstraction and structuring of the database has been carried out for "buildings" as can be seen in figure 7. In fact, we may well have residential buildings, commercial buildings, services offered in buildings etc. Furthermore, often the situation is such that the same buildings hosts different functions at different floors, a condition that has been included in the database system. As shown in figure 7, in many sectors you have components that are building (even in the case of lifelines you have the control centre for example). Such buildings can be ordinary or complex structures (a hospital or a theater is a complex structure from multiple perspectives, meaning also it cannot be easily standardized). In the LODE project as for now the database has been developed for ordinary buildings, thus covering residential almost completely (we may still have special residential buildings that are complex structure but this is an exception), most commercial activities and several services (such as for example a bank or a post office).

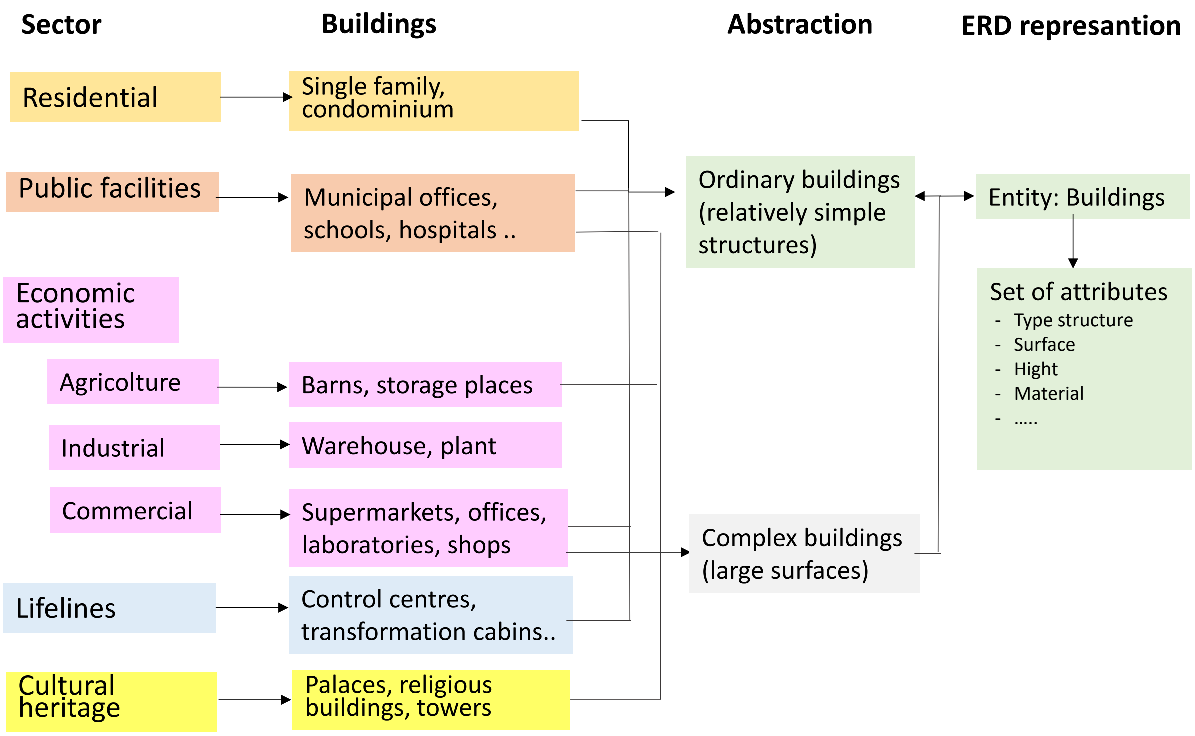

Figure 7. Conceptualization of the database on buildings and its relation to the different sectors

2. Technical aspects of the developed system and its replicability

The LODE system has been implemented used on Model View Controller (MVC) paradigm on the server side to allow the separation of software modules used to store, elaborate, and present the data related to survey and damages.
The Client – Server communication is instead based on Representational State Transfer (REST) architecture that allow the implementation of stateless Graphic User Interface GUI.
The main components of the IT system are the follow:

- **GUI**: is a web CRUD (Create, Read, Update, Delete) application coded using ReactJS framework and related libraries (as Redux and Router). The GUI code is served as static resource SPA (single page application) using Nginx server as reverse proxy.

- **Application Server:** is a component that manage the request coming from GUI**,** check the validity and store or request the data from DB. The server application is created in Java using ORM (object to relation mapping) Hibernate implementation to translate native query to OO (object oriented) entities and Spring framework to reduce the boilerplate code and enforce programming best practices. The Spring framework creates a REST API interface that allow to query the system by generic client (our GUI is an example of web client implementation).

- **Relational DB:** is a component that store the data and allow to query it using SQL language. We use a PostgreSQL engine empowered by GIS extension to manage geospatial data and support topology query.

In Figure 8 we can see the high-level view of the data model structure with the evidence of relation between logical entities

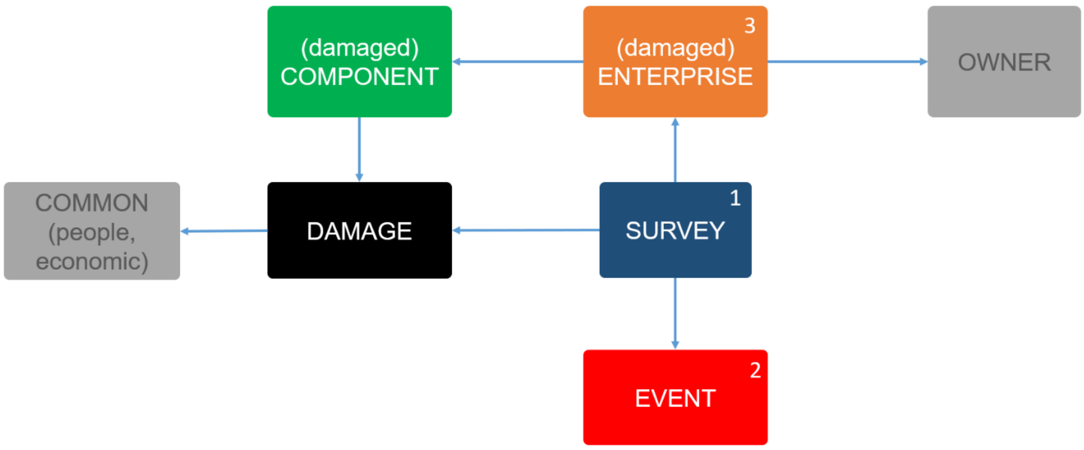

Figure 8. High level application workflow.

In the following Figure 9. we can see the high-level view of the application workflow:

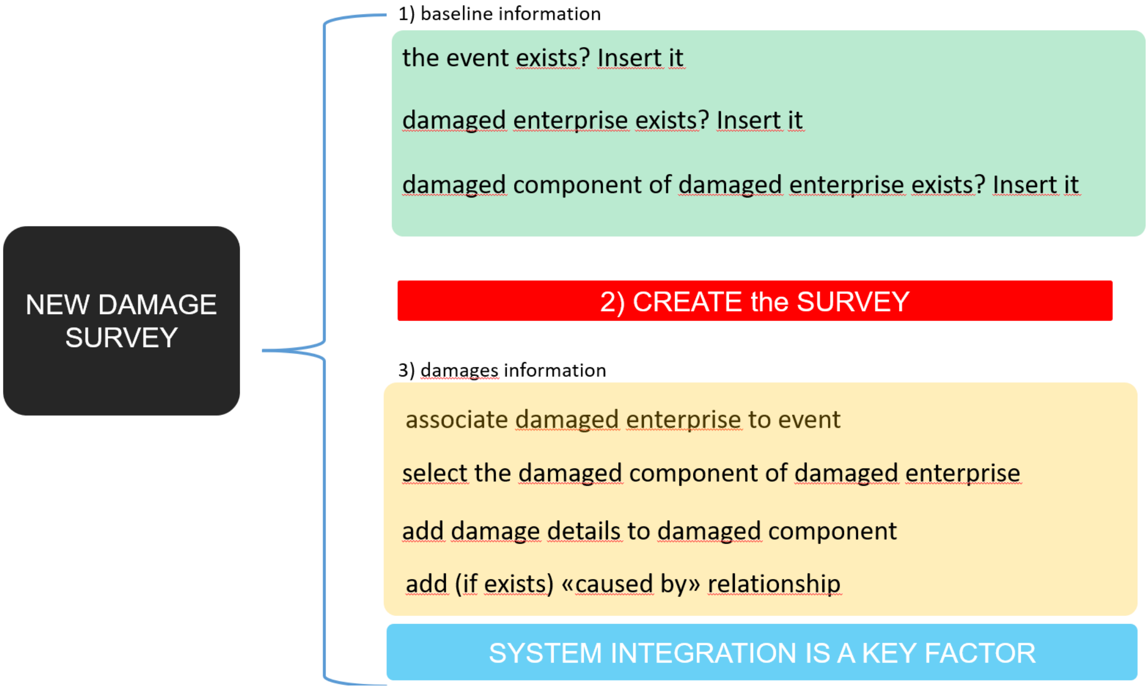

Figure 9. High level view of the application workflow

At the core of the entire activity is the creation of a survey that is triggered by the declaration of an emergency, an occurred event and its peculiar specific circumstances depend on the legislation in each country. In Italy for example following an earthquake, victims ask for a survey in case their building has not been inspected yet by civil protection authorities. As can be seen in the figure, first information on the surveyed entity must be inserted in case it is not already available and then could be captured from another database. Then the damage is inserted based on the survey.

In Figure 10, the high level view of the software architecture is provided.

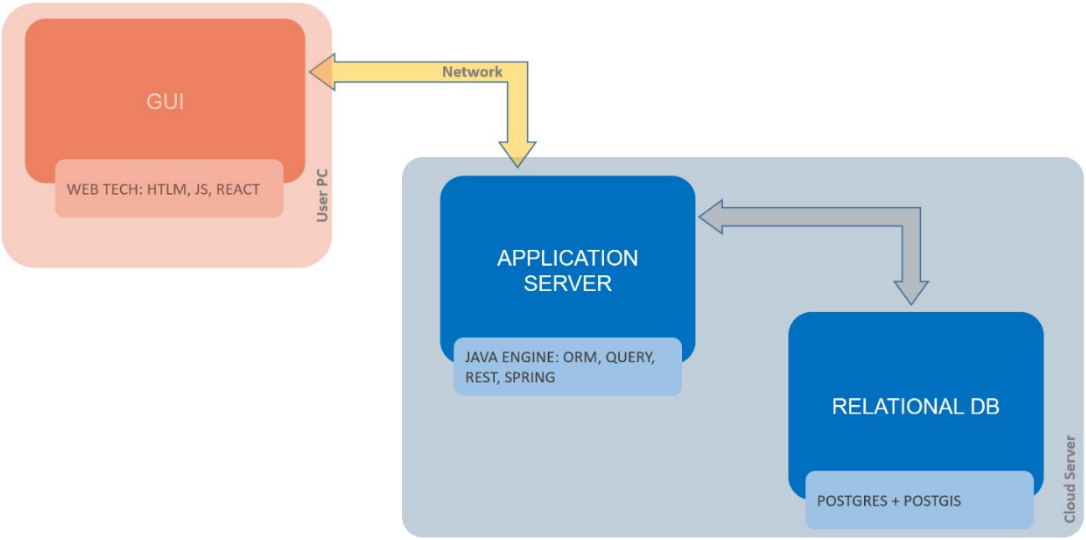

Figure 10. High level view of the software architecture

All the component was released in production under Digital Ocean cloud server and available in public domain at address https://lode.world/lode/ with a minimal hardware requirement of 1 shared vCPU with 1GB RAM.

3. Interoperability with other databases: governance and technical

In order to understand the needed interoperability with other databases and datasets produced by a large number of public a step back must be done in order to better characterize the life cycle of disaster data and what is the use of the data at each phase.
According to what has been also highlighted in the White Paper, the damage and loss data can be useful across all the disaster cycle.
The life cycle of damage data begins with research and planning of which types of data will be useful after the occurrence of a catastrophic event. The questions that arise during this

phase of the data life cycle concern the preparation needed for what kind of data are relevant and how they can be recorded after the occurrence of the event.
During the collection phase of the data, it is important for detailed guidelines on collection methods and timing. The collection of data in multi hazard, cascading and/or interconnected events complicates the procedure. Further questions arise on who collects the data, targeted training for data collection methods, the possibility of citizen's contribution and the issues that arise around self-reporting of damage and loss data.
The following phase, data processing and analysis calls for questions around data homogenization and data aggregation in order to understand the direct, indirect, tangible and intangible disaster impacts.
There is therefore an aspect of lifecycle related to the production and management of the disaster data and to the design and programming its collection; there is also an aspect of the use of the data to produce distinct analytical results. In both cases other datasets and other databases must be used. In fact, even to collect the data in surveys, one needs to know the area to be covered and should be guided by a map that pre-exists the event. For example in the case of buildings, it is sometimes the case and would make a lot of sense in general, to provide surveyors already with maps and data related to cadastral information on the properties to be inspected. This saves time but also guarantees a better result in terms of data quality and reliability.

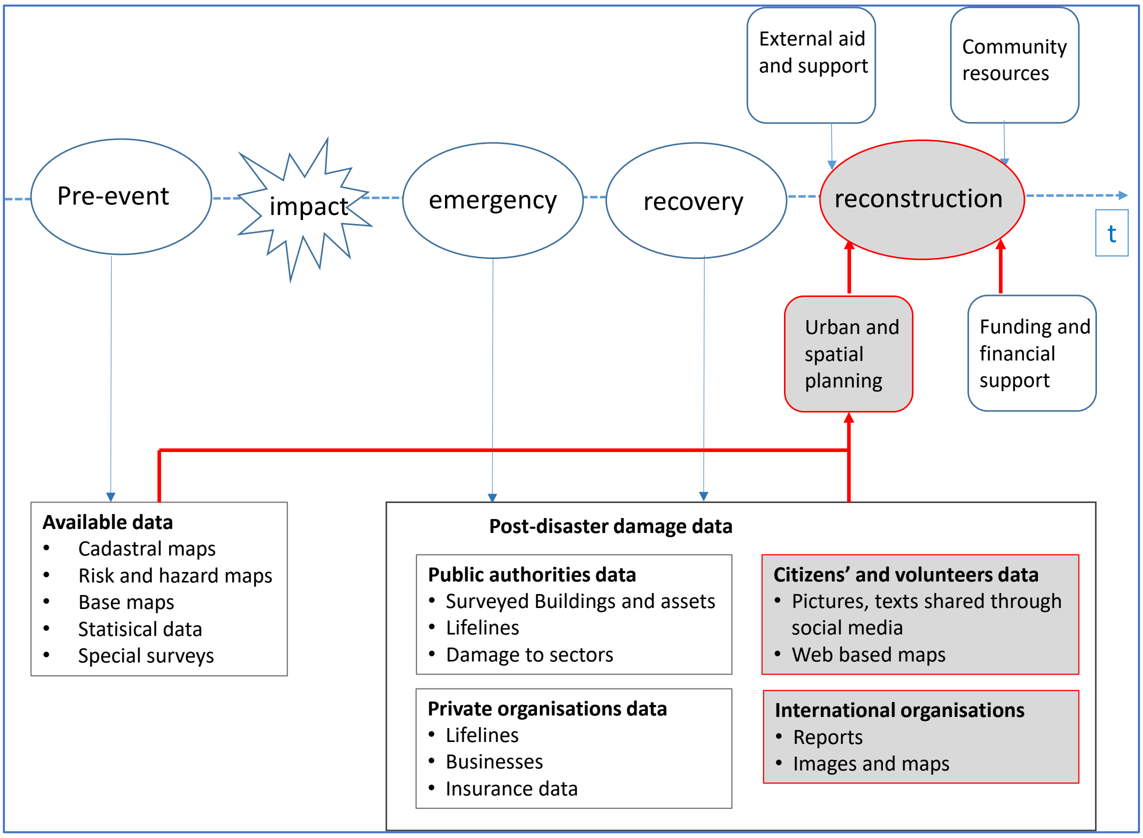

Figure 11. Production, use and reuse of damage and loss data

In an article we wrote at the end of the IDEA project (Mejri et al., 2017) the following figure was produced and also included in the proposal of LODE (Figure 11) that represents the different kind of data that must “interact” to produce meaningful results supporting disaster risk reduction in all phases. Both the data and their use are occurring at different timescales and with different purposes.

In the pre-impact phase different data layers are used to produce risk assessments, scenarios and forecasts of probable damage. Such data layers include georeferenced statistical data about population and urban areas, infrastructures, public facilities, industrial plants, hazards of different types etc.

In the aftermath of a disaster, the same layers and also the produced scenarios are used to identify epicentral areas, areas more in need of aid and rescue (because they are known to be isolated or poorly accessibile) etc. During the emergency some damage data are produced, for example through Copernicus or with fast reconnaissance surveys. At a later stage, during the recovery, better damage data are collected and in some cases (in ideal cases) mapped. Once this occurs, such data can be used for reconstruction purposes, and in this case the comparison with the situation prior to the event in terms of urban fabric, facilities etc. must be considered in order to prioritize interventions and also to make sure that mistakes are not done twice. For research purposes, but also with the aim of better supporting ministries and regions in charge of urban planning, long term mitigation etc., damage data can be used to produce ex post-scenarios to compare with pre-event risk assessments and scenarios to verify to what extent the latter two differ and if so for what reasons. The ultimate aim being improving risk assessment models.

This discourse regarding the multiple uses and re-use of different types of data, including but not limited to damage and loss data, raises the issue of interoperability. Interoperability means then, according to the experience gained in the LODE project through the showcases and in the consultations with stakeholders the following two must be considered:

- Technical interoperability;
- “Organisational” interoperability;

As for the technical interoperability it is often taken too much for granted that it is easy to obtain. However matching for example regional maps georeferenced according to the most common systems and cadastral data is often very difficult because of intrinsic differences in the georeferencing that derive from the purpose for which they are developed. In other cases, the grid that is used for representing the data is difficult to match, as the case of statistical data that are represented by census area that do not match with ordinary regional or city layers.

As for the organizational interoperability, it must be reminded that in cases where technically it is feasible, to obtain the exchange of data and the extraction of datasets from another database, an agreement must be established among organisations. This is not per se impossible but requires collaboration and defining modalities of sharing and in some cases overcoming privacy and other issues related to sensitive data (as may be persons and businesses that have been damaged in an event).

4. Limits of usability of the developed database

The developed database system as mentioned several times is apt to serve an enhanced damage data collection for the future, it is not conceived to be used for historic data regarding damage and losses recorded in past events.

The relational database system requires some re-engineering or re-processing of the current procedures of damage data collection but more importantly of damage data storage and management, it requires an entity that coordinates such storage in a way that permits filling the different records on the different sectors for most parts. This re-engineering or revision of current procedures and responsibilities is in line with the mandate of Open Data according to June 19 EU Directive (EU) 2019/1024 regarding the usability and reusability of data, especially those produced by the public sector. Also in the Recovery Fund Next Generation EU one of the key priorities is on digitalization of the public sector. In this regard the revision of current practices is of extreme importance to comply with the Directive and strengthen the accessibility of critical data for the public administrations involved.

The additional effort that is required to compile a database such as the LODE one is not as large as it may be thought of without testing it. We made sure that most of the data that are required are actually already now collected by some entity during or after a disaster. Few are really additional data. However what is dramatically different is the way the data are actually stored and managed and the request of an office that coordinates both the access to data from different sources and the subsequent storage, management and of extreme importance the longer term sustainable maintenance and update of the system. This coordination requires for sure an allocated budget and specialized personnel that at the moment is not mandatory. However some of the public administrations we worked with already have this type of competence and personnel specifically allocated, but often using inapt and inefficient software (Excel tables).

As for the historic data, our experience with existing partial databases show that a re-adaptation of the LODE system would be necessary to use it effectively for historic data. So, a two fold additional work would be required. On the one hand some re-adaptation of the LODE system to match with the majority of existing historic databases or paper documents. This to avoid having many records simply empty. The other side of the needed effort relates to unpack information that is available on past events according to the formats of the adapted database. Unpacking for the sake of being able later to query the system for multiple purposes. Such unpacking, perhaps even more than adapting an existing database system, requires additional, dedicated specialized personnel that will systematically input data for all events for which information exists.

Whilst data that is available for events dating back ten or more years ago is certainly very fragmented and so the challenges are even greater, for more recent events some structuring and organization of data has already occurred. However, even for the LODE showcases for most of which we had at our disposal the best data available, still many gaps existed or the format of the data was such that it made it very hard to populate the developed database system. Only a very few records and tables could be actually filled in.

References

Mejri O., S. Menoni, K. Matias, N. Aminoltaheri, (2017) Crisis information to support spatial planning in post disaster recovery, in International Journal of Disaster Risk Reduction, 22, pp. 46–61.

Teorey, T.J., Lightstone, S., Nadeau, T., Jagadish, H. V. (2011) Database modeling and design. Logical design. 5th Edition, MK.

To be informed and updated about **LODE**'s activities access:

www.lodeproject.polimi.it