

R2DN – Integrating Databases about Natural Disasters

José Veríssimo

PPGI - Graduate Program in Informatics
Universidade Federal do Rio de Janeiro
verissimo@ufrj.br

Marcos R. S. Borges

PPGI - Graduate Program in Informatics
Universidade Federal do Rio de Janeiro
mborges@ppgi.ufrj.br

ABSTRACT

The historical data on disasters serves various purposes, such as managing the risk of future disasters, assessing risk in the insurance sector, and formulating policies to reduce disaster risks, among others. To meet these objectives, there are databases, covering several types of disasters. However, in addition to differences in purposes, scope, access, and coverage, there are variations in the criteria for registering a disaster in the database and the information used for this register. Due to these differences, we have a scenario with discrepancies of information across existing databases. In this study, we propose a repository that combines the main advantages of some global databases, the data needs due to multilateral agreement, and the possibility of integrating regional or specific databases. The proposed Natural Disaster Data Repository (R2DN) contains a data schema and a data mapping process with two of the main global databases (EM-DAT and DesInventar) and S2ID.

Keywords

Disaster databases, database integration, EM-DAT, DesInventar, S2ID

INTRODUCTION

Natural disasters are phenomena that have been part of the Earth's dynamics for centuries, with recorded human casualties and material damages. When they occur in uninhabited areas, they are termed natural events. Earthquakes, floods, tsunamis, epidemics, and droughts are some of these phenomena with distinct characteristics and consequences. In some cases, an event may be associated with one or more of these phenomena, such as the earthquake that triggered the Indian Ocean tsunami in December 2004, affecting 13 countries in the region.

The Centre for Research on the Epidemiology of Disasters – (CRED, 2020b), at the Université Catholique de Louvain in Belgium, estimated that, in the last two decades (2000 to 2019), 7,348 disaster events (excluding biological and technological disasters) were recorded, resulting in approximately 1.23 million lives lost, affecting over 4 billion people, and causing economic losses of around \$2.97 trillion. These figures represent a significant increase compared to the two previous decades (1980 to 1999), which recorded 4,212 events, 1.19 million deaths, over 3.2 billion people affected, and losses of approximately \$1.63 trillion, as summarized in **Figure 1**.

In a scenario where various local databases provide information for the purposes of their regions, sharing information between databases is necessary for more comprehensive studies that include multiple regions. To ensure the agility, quality, and integrity of data, compatibility between the data is necessary, allowing for sharing between databases. However, as mentioned by (Panwar & Sen, 2020), existing major databases still exhibit disparities in disaster recording, making data exchange for comparison and analysis challenging. According to (Koç & Thieken, 2018), the lack of reliable, consistent, and comparable data is seen as a significant obstacle to effective long-term prevention of losses.

The Emergency Events Database (EM-DAT), developed by CRED in 1988, is widely recognized as a reference for the academic community, industry, and government entities in the disaster assessment phase. EM-DAT is maintained by the Université Catholique de Louvain with the support of the World Health Organization (WHO). It is available to non-commercial individuals and organizations upon registration on its website (CRED). A simple search in the Scopus database reveals that over a hundred articles, since 2012, reference EM-DAT in their abstracts. There are other databases, public and private, enabling comparisons between their characteristics.

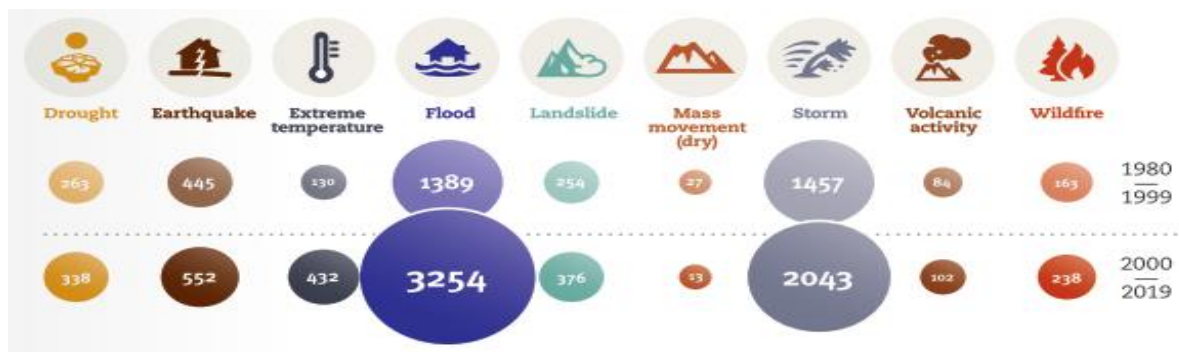


Figure 1. Impact of Disasters: 1980-1999 versus 2000-2019 (CRED, 2020b)

The DesInventar, an initiative of UNDRR (United Nations Office for Disaster Risk Reduction), is a system designed for the acquisition, recovery, querying, and analysis of disaster records. This system is made available to countries in Latin America, the Caribbean, Asia, and Africa.

The S2ID is an Integrated Disaster Information System developed by CEPED/UFSC in 2011/2012 (“Ministério do Desenvolvimento Regional - S2ID”), serving as a support tool for states and municipalities affected by disasters that require federal assistance. The system enables these federative entities to exchange information with the federal government to obtain necessary resources to assist the affected population.

This work supports the proposal that each country or institution should have its own specific database of natural disasters to meet the specific needs of understanding natural phenomena occurring in its territory and their impacts on society. We share the proposals presented by (Wirtz, Kron, Löw, & Steuer, 2014) and (Khankeh, 2022). The purposes of creating and using disaster damage databases, according to (Khankeh, 2022), should include:

1. Promoting disaster relief, recovery, and reconstruction programs.
2. Assessing risks of future disasters.
3. Estimating the economic feasibility of investments made to reduce losses.
4. Monitoring disasters and their impacts to achieve international goals established for disaster risk reduction.
5. Conducting thematic analyses, such as differences in deaths by gender or age group in disasters.

However, we understand that the proliferation of specific databases should consider adhering to some standards already established by other institutions managing disaster databases. This approach can facilitate data sharing among them and enable the exchange and/or crossing of temporal and geographical information necessary for generating more effective analyses supporting governments and organizations in decision-making and the formulation of disaster risk reduction policies. The exchange and sharing of disaster information among countries are explicitly stated in Priority 4 of the Sendai Framework for Disaster Risk Reduction 2015-2030 (UN, 2015), and it is part of the commitment made by 187 UN member countries.

It is also important to emphasize the need to include data in the databases grounded on the demand for new attributes aligned with the disaster risk reduction indicators required by the Sendai Framework. Divergences between databases hinder data sharing and, consequently, limit the potential utilization of disaster databases. This is the main problem that will be addressed in this work.

What is the structure of a data repository that allows the assessment of natural disasters, is compatible with global databases, meets specific demands and complies with multilateral agreements in disaster management?

To answer this question, we established three goals that are described and discussed in this work. They are:

1. To develop a database structure that integrates disaster records from different databases.
2. To include the necessary information for the Sendai Framework indicators.
3. To enable the integration of specific databases aimed to support specific purposes from countries or institutions to assist their policies.

In the following sections, we will present the work context, the methodology employed to develop the study, the description of the three databases selected for the integration exercise, the criteria used for selection, the taxonomy and the R2DN data model, as well as an instance of the integration of the three described databases.

BACKGROUND AND METHODOLOGICAL APPROACH

The growing global demand for actions aimed at disaster risk reduction has driven the need for databases capable of storing and retrieving information related to these events. In response to this scenario, we observe a significant number of published articles addressing the use of disaster databases for various purposes, regardless of the authors' areas of expertise.

In Table 1, we present the results of a search conducted in the Scopus database. This research involved the inclusion of the keyword "database" along with terms representative of each phase of disaster management. We also provide the number of articles found since the year 2013, offering an overview of academic production in this field.

**Table 1. Number of articles found from 2013 onwards, in the Scopus database, using the corresponding sentences
Data retrieved on 10/31/2023**

Search sentence	Number of articles
((“natural disaster” OR “natural hazard”) AND (“prevention” OR “risk reduction”) AND (“database”))	367
((“natural disaster” OR “natural hazard”) AND (“response”) AND (“database”))	347
((“natural disaster” OR “natural hazard”) AND (“assessment”) AND (“database”))	760

The Design Science Research (DSR), widely employed in Information Systems research and advocated by (Hevner, March, Park, & Ram, 2004), is the methodology we have chosen to conduct our research. Its guidelines, which include identifying a problem in society, creating an artifact capable of contributing to the solution of this problem, evaluating the contributions generated by the artifact, and disseminating the achieved results, have shown to be consistent with the planned strategies for the development of our study.

In our study, we identified a problem related to the difficulty of data sharing among existing disaster databases. Additionally, the solution to this problem must meet the requirements for emergency response in Brazil and be in compliance with international agreements in the field of disaster risk reduction. To address this issue, we designed a repository (artifact) that integrates disaster records from three existing databases. The challenges encountered during the integration process are analyzed, and based on these analyses, new requirements are formulated to enhance the repository's design, aiming to overcome the identified challenges.

In Figure 2, we present a diagram containing the stages and deliverables of our methodological approach, aligned with the guidelines of DSR.

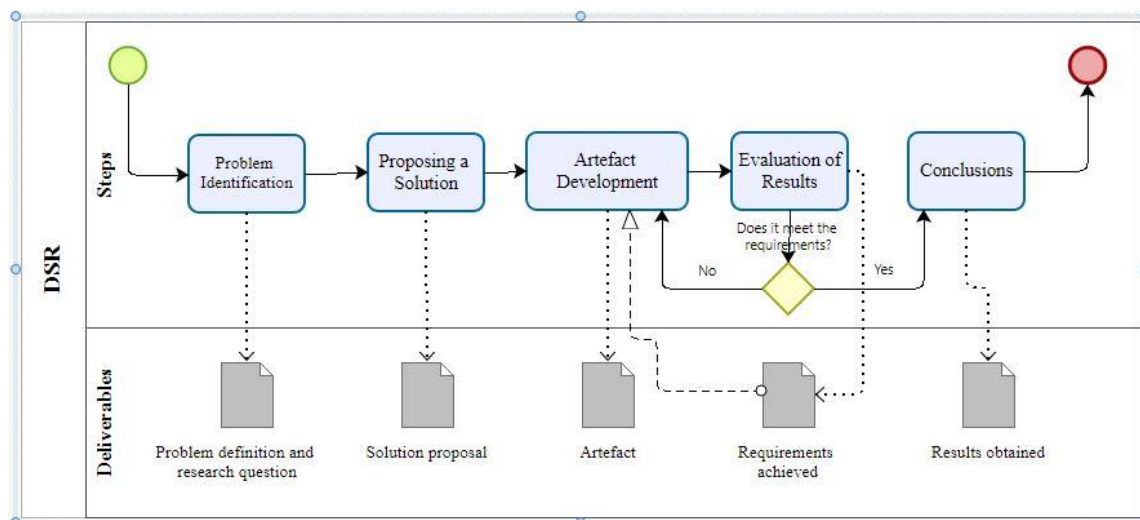


Figure 2. DSR steps and its deliverables

In the "Problem Identification" stage, where we identified a specific problem in society, it was described in this section along with the research question that contains the requirements for the desired solution. In the "Proposing a Solution" stage, we developed a model for a repository that meets the solution requirements, as presented in Section "The Proposed Model for R2DN".

To demonstrate the feasibility of the proposed model, we implemented an instance in the "Artefact Development"

stage, as described in Section “Instantiating R2DN”. The repository development is divided into two parts. In the first part, actions necessary for the instantiation of the artifact are identified in the face of encountered difficulties, and in the second part, the implementation of R2DN is carried out, followed by an "Evaluation of the results". After each evaluation stage, the results are analyzed, and adjustments to the artifact are made as needed to achieve the expected outcomes.

DISASTER DATABASES

This work is based on three distinct disaster databases. Table 2 outlines the justification behind the selection of each one, as well as the criteria adopted for registering catastrophic events.

Table 2. Selected databases with motivation and event recording criteria

Database	Motivation	Registration Criteria
EM-DAT	Widely recognized by the disaster management community and accessible to the public.	10 or more human deaths / 100 or more affected people / Declaration of a state of emergency by the country and/or an appeal for international assistance.
S2ID	Aims to analyze emergencies in Brazilian states and municipalities.	Requesting financial assistance from the federal government.
DesInventar	Records disasters in all South American countries, except Brazil.	Each country using the system has the autonomy to establish its own criteria.

The use of different disaster databases allows for a more comprehensive understanding of disasters due to variations in criteria for recording in each database. However, these discrepancies lead to disparities between databases, making information sharing a challenge and consequently limiting the potential utilization of these disaster databases.

Despite EM-DAT being the globally recognized disaster database, we emphasize a significant discrepancy with the reality in Brazil (the country we chose to perform the data integration). The report on damages and losses resulting from natural disasters in Brazil (1995-2019), produced by CEPED (UFSC) and updated by FAPEU with support from the World Bank (FAPEU, 2020), indicates 64,429 disaster records with economic losses of R\$ 333.36 billion (equivalent today to approximately US\$60 billion) over these 25 years of the study. In the same period, EM-DAT recorded (query conducted on 07/20/2022) 136 disasters, totaling economic losses of approximately US\$ 18 billion.

Several reasons can justify such a discrepancy, including different criteria for recording disasters and distinct purposes of the databases. While EM-DAT aims to register globally impactful disasters for humanitarian purposes, CEPED's database aims to register disasters that states and municipalities require some form of federal assistance.

The numbers presented above highlight the need for a specific disaster database for Brazil that aligns with our objectives and is compatible with major global databases. This would enable data sharing and comparative studies on the losses and damages caused by disasters across countries. We believe that recording losses from disasters is the first step in assessing resilience evolution in different regions.

According to (IRDR, 2014), most disaster databases record human and economic losses in some way, but they lack a standard set of indicators established from a common understanding. An initiative to address this issue is the set of indicators established in the Sendai Agreement.

To initiate the selection of disaster databases, we relied on the systematic review conducted by Khankeh (2022), which provided an analysis of these databases. Based on this review, we selected those with a global presence and a higher number of references on Google Scholar. Table 3 presents some of the characteristics collected from these databases' websites in June 2023. Note that not all information was completed in the websites. In addition, we included the S2ID to the table because it is the official system of the Brazilian government.

The process of selecting R2DN attributes was divided into three stages:

1. Analyzing the data structures of the three selected databases.
2. Analyzing the data needs of the Sendai Framework indicators.
3. Analyzing the data organization of the DaLA methodology.

Table 3. Comparison of the characteristics of disaster databases (the numbers on the website point to the URLs at the end of the table)

	EM-DAT	NatCatService	Sigma	DesInventar	GLIDENumber	Dartmouth	S2ID
Site	1	2	3	4	5	6	7
Administrator	CRED – Belgium	Munich Re – Germany	Swiss Re - Switzerland	LA RED with support from the UN	Asian Disaster Reduction Center (ARDC)– Japan	University of Colorado	National Secretariat for Civil Protection and Defense - Brazil
Start year	1988	1980	Not available	1994	2004	1993	1988
Objective	Support humanitarian actions and provide a basis for assessing vulnerabilities and priorities.	Developed to meet the needs of the insurance industry.	Developed to meet the needs of the insurance industry.	Support analysis of disaster trends and their impacts.	Proposal for a single, global code for identifying disasters.	Remote sensing-based measurements and mapping of the digital map of surface water changes.	The official system for Brazilian states and municipalities to request federal aid.
Disasters since	1900	From 1980 onwards for disasters that meet your criteria.	1970	Varying for each country.	Not available.	Uninformed.	2003
Available	Public upon registration on the website.	Private, but publishes periodic atlases of losses in disasters.	Private, but publishes periodic atlases of losses in disasters.	Public	Public	Public	Public
Registration criteria	> 10 deaths/ > 100 affected/ Declaration of a emergency/ Request for help	Events that result in deaths, injuries, homelessness and/or property damage.	> 20 deaths, > 50 injured, > 2,000 homeless and/or insured losses > \$14M.	At the discretion of the registered information sources.	At the discretion of the accredited countries.	Changing water levels on Earth (captured by remote sensing).	Declaration of a state of emergency or state of public calamity.
Number of records *	25,596 of which 16,398 were natural disasters.	Not available	Not available	Not available	7,589 disasters with Glidenumber.	Not applicable	Not available

1 - <https://emdat.be>2 - <https://www.munichre.com/en/solutions/reinsurance-property-casualty/natcatservice.html>3 - <https://www.sigma-explorer.com/>4 - <https://www.desinventar.net/>5 - <https://www.glidenumber.net/glide/public/about.jsp>6 - <https://floodobservatory.colorado.edu/index.html>7 - <https://s2id.mi.gov.br/>

* Note: All links accessed on 06/13/2023

EM-DAT's data structure is accessible through its website (CRED). The available information covers various aspects of disasters, including identification, classification, location, and human and financial impacts. Among these attributes, five are related to disaster classification, covering nature, group, subgroup, type, and subtype of the event. Regarding the disaster's location, information about the continent and country is provided. However, there is only one attribute to describe the specific location within the country. Information related to financial impacts focuses on total reconstruction costs without providing specific details about the affected economic sectors.

DesInventar provides a document addressing the system's use, giving an overview of its data structure (UN Office for Disaster Risk Reduction (UNISDR), 2019). However, there is a lack of formalism in defining this data. The disaster classification is simplified into a single attribute called "event type" while financial impacts are detailed concerning the affected economic sectors.

S2ID metadata is accessible in the Digital Atlas of Disasters in Brazil (Ceped - Centro Universitário De Estudos E Pesquisas Sobre Desastres, n.d.). Disaster classification information resembles that of EM-DAT, as it served as the basis for S2ID disaster classification. The disaster's location is clearly defined, as municipalities where events occurred register these occurrences to request financial assistance from the federal government. As a result of this process, financial impacts, especially those related to the public sector, are thoroughly documented.

Sendai Framework

The Sendai Framework is a global initiative aimed at promoting disaster risk reduction policies worldwide. It aims to measure the progress of initiatives adopted by countries between 2020 and 2030, compared to the period from 2005 to 2015, through the establishment of seven global targets (UN, 2015). To achieve this goal, the framework seeks to prevent new disasters and reduce risks by implementing integrated and inclusive measures that aim to minimize exposure to risks and vulnerability to disasters. Additionally, it aims to enhance preparedness for disaster response and recovery to strengthen the resilience of affected communities.

These targets, to be achieved by 2030, are divided into:

- a. Reduce global disaster mortality.
- b. Reduce the number of affected people globally.
- c. Reduce direct disaster economic loss in relation to global gross domestic product (GDP).
- d. Reduce disaster damage to critical infrastructure and disruption of basic services.
- e. Increase the number of countries with national and local disaster risk reduction strategies by 2020.
- f. Enhance international cooperation to developing countries.
- g. Increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to people.

Targets a, b, c, and d require data from national disaster databases as evidence of progress in disaster risk reduction policies. Therefore, the attributes to be included in the R2DN data model must align with this need.

THE PROPOSED DATA MODEL FOR R2DN

In this section, we present the answer to our research question. The elements that supported the construction of the model include information sources from the metadata of the three databases, in addition to the indicators of the Sendai Framework.

Figure 3 illustrates the process of modeling the R2DN data schema. The first activity, "Identify disaster data categories" aims to organize the data into categories and subcategories. This approach, combined with the metadata, is crucial for a more precise understanding of the data and assists in defining responsibilities for collecting information about disasters.

Next, in the activity "Select R2DN attributes" we chose the data based on the attributes of the selected databases that will compose the disaster record in R2DN. In the subsequent activity, named "Allocate disaster attributes to their categories" we associate each attribute with its corresponding data category. For the data selected for R2DN, we check if there is a metadata description available in the source database. If it exists, this description is integrated into the R2DN model through the "Search metadata in the source database" activity. However, if it does not exist, in the "Define metadata" activity, metadata is generated for these data.

Finally, in the activity "Improve metadata with real examples" in the face of the absence or low quality of metadata in the source databases, we resort to real examples in the databases to enrich the corresponding metadata and improve its understanding.

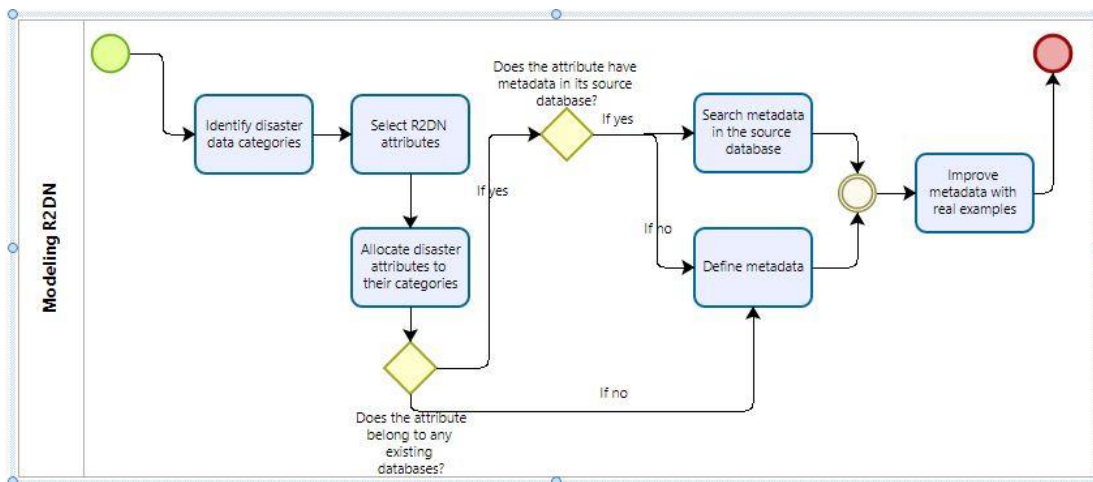


Figure 3. R2DN modeling process

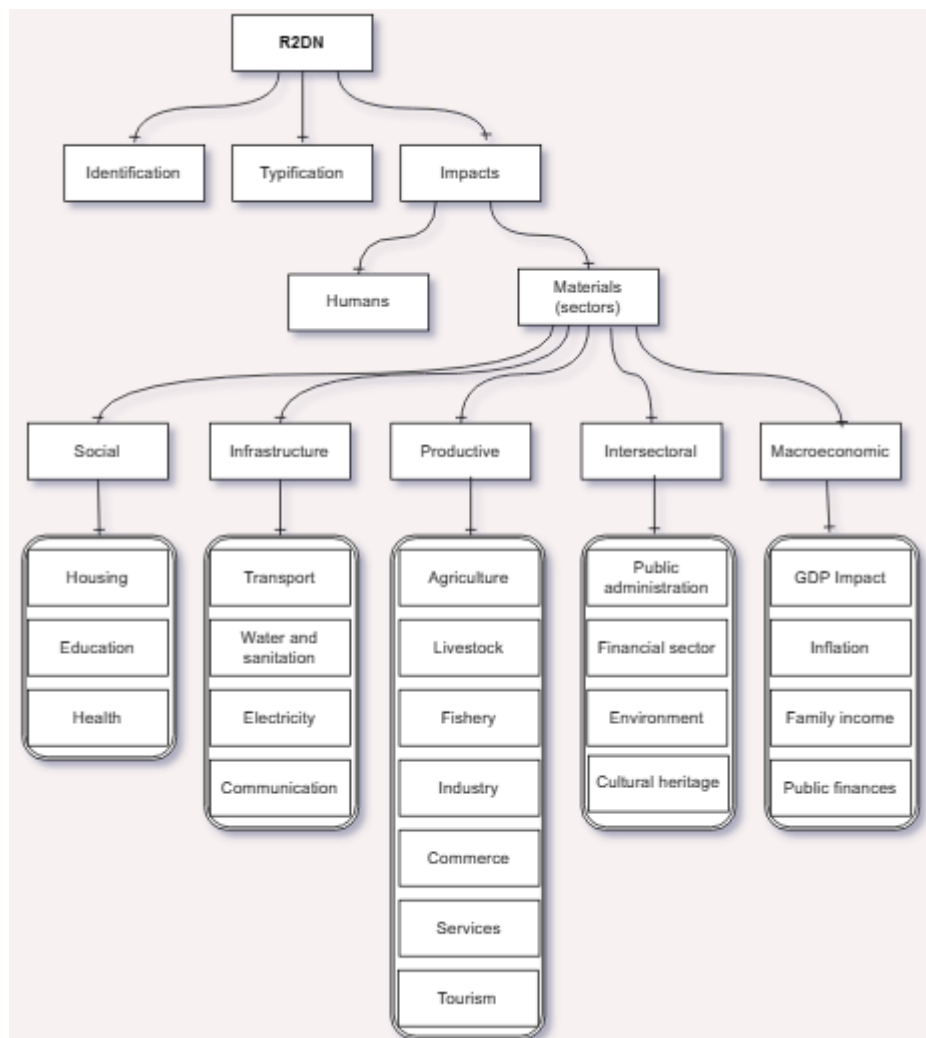


Figure 4. R2DN category and subcategory structure taxonomy

Taxonomy of R2DN Data

As a result of the "Identify disaster data categories" activity represented in Figure 3, we present in Figure 4 a

taxonomy with the structure of identified categories and their subcategories. To construct this taxonomy, we used the DaLA methodology as a reference, which is employed to analyze the impacts generated by disasters on the economy of a society. (World Bank, 2010). The purpose of this representation is to organize the data that will compose the R2DN disaster record, making its understanding and location more accessible. Below, we provide a description of these categories for a better understanding:

- Identification – group all information related to the disaster location, date and time, and event duration.
- Typification – group all information associated with the types and subtypes of the disaster (e.g., hydrological and flooding), disasters associated with the event, and triggering events of the disaster.
- Impacts – human and material impacts caused by the disaster.
- Human – contain the quantity of people directly impacted by the disaster (deaths, missing, injured, etc.).
- Material – group the sectors impacted by the disaster.
- Social – encompass data regarding the quantity of affected housing and educational and health units, whether by destruction or damage, along with associated values for their respective reconstructions.
- Infrastructure – consolidate data related to damages in transportation, water, sanitation, energy, and communication areas.
- Productive – cover damages in production sectors such as agriculture, livestock, fisheries, industry, commerce, and tourism.
- Intersectoral – group damages related to public administration, the financial sector, and the environment.
- Macroeconomic – will contain impacts on inflation, GDP, household income, and public finances.

Table 4. Categories and subcategories and their attributes

Category/subcategory	Attributes
Identification	continent, country, first and second administrative divisions, latitude, longitude, start date and time, duration, name, glide number, source and type of information.
Typification	category, group, subgroup, type, subtype, associated disaster, magnitude value and scale, triggering event, causes and effects.
Humans	dead, injured, sick, missing, displaced, homeless, other affected and people whose livelihoods are affected;
Material impact	description of public and private losses, total public and private losses, total securitized losses and total reconstruction cost;
Impact on housing	destroyed and damaged units, total number of people in destroyed and damaged units and loss value in housing units;
Impact on education	destroyed and damaged units, value to recover the units, value to reestablish teaching and number of days without teaching services;
Impact on health services	destroyed and damaged health units, cost of rebuilding health units, value of reestablishing emergency services, value of loss in vector control and number of days without health services;
Impact on infrastructure	amount of repair and reconstruction works, value of losses for repair and reconstruction, value of losses in public facilities and number of days of interruption in infrastructure services;
Transport	total cost required to restore transport services;
Water and sanitation	losses in the restoration of water supply and restoration of river and sanitary sewage;
Energy	total cost required to restore power generation and fuel distribution;
Communication	total cost to restore telecommunications services;
Agriculture	estimated value of loss of various types of crops;
Livestock	estimated value of losses in dead or sick animals;
Fishery	estimated value of loss in fishing activity;
Industry	estimated value of lost industrial production;
Commerce	estimated value of sales losses from commerce;
Services	estimated value of loss in the services sector;
Other sectors	estimated value of losses in other sectors not previously listed;
Public administration	estimated value to reestablish public safety and urban cleaning services;
Financial sector	estimated value of losses in the financial sector;
Environment	description of environmental damage, population affected by water, air and soil pollution, population affected by water depletion and area affected by fires in parks;
Cultural heritage	estimated value of losses in cultural heritage;

GDP	variation in percentage points of GDP;
Inflation	variation in percentage points in the inflation rate;
Family income	financial impact on family income;
Public finances	financial impact on public resource collection.

Selection of R2DN Attributes And Their Metadata

After analyzing the attributes of each database to address our research question, we adopted the following criteria for selecting the attributes that will be part of the R2DN.

1. **S2ID** – All attributes directly related to disaster events were selected. This selection is based on the R2DN's commitment to providing essential information to the Brazilian federal government to assist states and municipalities affected by a disaster.
2. **EM-DAT** – All attributes related to identification, typification, and impacts associated with the disaster event, as made available on the website, were selected.
3. **DesInventar** – All attributes related to the identification and typification of the disaster were selected. Additionally, among the attributes of human and material impacts, those containing information like S2ID were selected. Also included were those necessary for the indicators of the Sendai Framework.

To meet all the demands of our research question, 101 attributes were chosen, including those from the three selected databases and those necessary for the Sendai Framework indicators. These attributes are distributed across the categories and subcategories presented in Figure 4. In Table 4, we provide a summary of these attributes within their respective subcategories.

The context of this work is the registration of disasters in Brazil. An initiative in another country with the same objectives should encompass items 1 and 2 and adapt item 3 to the context of the new database.

INSTANTIATING R2DN

In this section, we will address the activities necessary to feed the R2DN repository with information about disasters from the selected databases throughout this work. Initially, it is essential to map the R2DN data schema in relation to the schemas of the three databases used in our research: EM-DAT, DesInventar, and S2ID. Integrating disaster records from different databases becomes a valuable tool for comparing different versions of the same disaster.

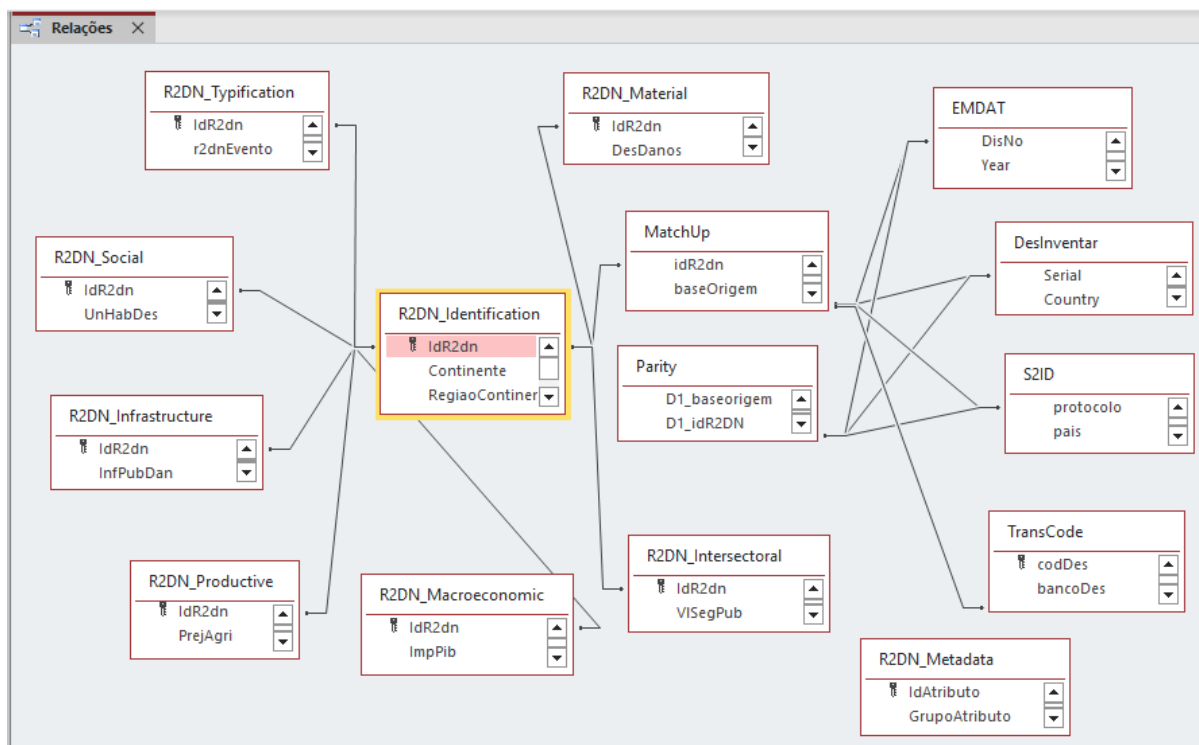


Figure 5. Implementing R2DN in Microsoft Access

Regarding the implementation of the R2DN tables, we decided to organize them based on the data categories established in our taxonomy, as illustrated in Figure 5. In addition to providing a clearer view of the model, this approach was adopted due to the low fill rate in disaster records (CRED, 2021). Therefore, each table will contain records of a disaster only if there is information in the respective data category. The "R2DN_Identification" and "R2DN_Typification" tables, as they contain basic disaster information, will always contain records related to the disaster. Opting for a single-table implementation would have resulted in significant disk space wastage.

The "R2DN_Metadata" table contains the mapping of the attributes of the selected databases with the R2DN attributes. This mapping is necessary to load data from the three databases chosen in R2DN. All other tables with the prefix "R2DN" have information about imported disasters. The additional tables, "MatchUp", "Parity" and "TransCode", play an auxiliary role in the "MatchUp" process, and will be detailed in the "Implementing R2DN" section.

Comparing disasters from different databases can be a challenge, especially when the classifications used are distinct among them. According to (Gall, 2015), it is necessary to establish technical standards for disaster loss databases to ensure comparability and compatibility between databases. The lack of standards hinders the ability to aggregate and consolidate losses over time, space, hazards, and loss metrics, being a critical need in the climate change community. Understanding local, regional, and global conditions is essential for the development of adaptation strategies.

In Table 5 we conducted a comparison of classifications of natural disasters recorded in the databases: EM-DAT (CRED, 2020a), DesInventar (UN Office for Disaster Risk Reduction (UNISDR), n.d.), and S2ID (SEDEC, n.d.). To simplify, we have listed only hydrological and meteorological events, which represent most events in Brazil. Additionally, we limited the presentation to three levels in the classification: Group, Subgroup, and Type.

Table 5. Disaster coding comparison

	EM-DAT		Brazilian disaster code		DesInventar
Group	Subgroup	Type	Subgroup	Type	Type
Hydrological	Flood	Coastal flood	Flooding		Alluvion
		Riverine flood			Avalanche
		Flash flood			Flood

	Landslide	Avalanche, landslide, mudslide, rockfall and subsidence	Flash flood		Landslide
			Flooding		Rain
Meteorological	Storm	Tropical Cyclone	Large-scale/regional-scale systems	Cyclones	Frost
		Extra-tropical storm		Cold fronts / Convergence zones	Electric storm
		Convective storm	Storms	Local / Convective	Snowstorm
	Extreme temperature	Heat wave	Extreme temperature	Heat wave	Heatwave
		Cold wave		Cold wave	Strongwind Tornado
		Severe winter conditions			Hailstorm

Challenges For Data Integration

The challenge in integrating disaster records from different databases lies in identifying similar events across distinct sources. The benefits of identifying similar disasters include:

1. Avoiding distortions in statistical analyses, as some records may refer to the same disaster.
2. Enriching information about a disaster since the three databases contain complementary data.
3. Enabling the comparison of discrepancies in disaster information across various databases.

To identify similarities between disaster records in different databases, three pieces of information are essential: the type of disaster, the location, and the date of the event. The following details how each of these pieces of information will be handled during the R2DN setup process.

As presented in the Table 5, discrepancies in disaster classification among the three databases are noticeable. The difference between S2ID and EM-DAT is relatively smaller, given that the latter based its classification on the former. However, in DesInventar, classification is done by countries, making it more challenging to compare types of disasters. As this work does not aim to exhaust this topic, we choose to restrict the scope to the types of disasters we will address. Thus, we will focus on hydrological, meteorological, and climatological disasters. These categories represent 95.2% of the disasters downloaded from S2ID for the period 1991 to 2022.

Regarding the location of the disaster, greater discrepancies are observed in the information between databases. S2ID considers the location as a discrete domain of states and municipalities, eliminating the possibility of identical locations with distinct spellings. DesInventar also treats the location as a discrete domain, distinguishing administrative divisions by country. For example, in Argentina, it considers province and department, while in Chile, it considers region and province. In EM-DAT, however, the location is a free-text field where various administrative divisions of the country are mixed.

Regarding the date of the disaster, it is necessary to consider a tolerance margin for dates from databases, as there may be a difference of days between dates related to the same disaster in different records. For example, a disaster occurring on 31/12/2013 may be considered the same event in another database but with a date recorded as 01/01/2014. In S2ID, municipalities have a ten-day period to register the disaster in the system from the date of occurrence (UFSC, 2012).

The first activity to instantiate the R2DN involves mapping the proposed data for the R2DN and the data from the selected databases. As the R2DN will be fed with disaster records from the three external databases, it is necessary to identify the source database. From this information, we will use the mapping to find the correspondence between the attributes of the R2DN and the attributes of the source database.

The next step involves generating the files that will contain the disaster records from the three selected databases (CRED, n.d.), (UN Office for Disaster Risk Reduction (UNISDR), n.d.) e (Integração, n.d.), which will be imported into the R2DN. The next section will address the challenges faced when comparing disasters in the R2DN using data from the three databases.

IMPLEMENTING R2DN

The goal of this section is to demonstrate the feasibility of implementing the R2DN through the construction of an artifact, which includes the data structures of the R2DN, the data structures of the three selected databases, and the data structures that will assist in mapping the attributes of each database to the attributes of the R2DN. The main challenge associated with transferring disaster information from other disaster databases to the R2DN is identifying the duplication of disasters. Detecting cases where two or more records from different databases refer to the same disaster is the primary challenge of our experiment, given the disparate ways in which information such as disaster type and disaster location is handled in different databases. This is the challenge we will address during the development of our project.

Since EM-DAT contains disaster records from all countries, DesInventar covers countries in Latin America, excluding Brazil, and S2ID only records events that occurred in Brazil, our goal is to identify duplicate disasters between EM-DAT and DesInventar, as well as between EM-DAT and S2ID. Therefore, it is not expected to find information about an event in all three databases simultaneously.

Regarding the location of the disaster, we note that EM-DAT addresses this information less formally, resulting in a combination of various locations. In the case of Brazil, this may include states and municipalities. On the other hand, S2ID and DesInventar divide the disaster location information into first and second administrative regions, which are pre-registered by countries. Based on this, we conclude that comparing the location in EM-DAT with the other databases should consider both administrative divisions to verify the possibility of similarity of disasters.

Concerning the classification of types of disasters, we observe a significant similarity between the Brazilian disaster code used by S2ID and the disaster classification employed by EM-DAT, as indicated in Table 6. However, the classification used in DesInventar is determined by the country representative when registering the disaster, making it more challenging to find similarities between events. Additionally, we found several instances where the same disaster was registered with different types across databases. For example, as a flood and landslide or a forest fire and drought.

To simplify our observations, we decided to focus exclusively on hydrological, meteorological, and climatological disasters. These categories comprise 95.2% of the disasters recorded in Brazil, as documented in the S2ID Atlas (Integração, n.d.) from 1991 to 2022. Based on our findings, we chose to establish a coding, as illustrated in **Table 6**, for the R2DN. This coding will be used in the comparison between disasters.

Table 6. Disaster Coding Conversion to R2DN

Database	Events	codR2DN
S2ID	Flood, Flash flood, Flooding	1
EMDAT	Flood, Landslide	
DesInventar	Alluvion, Avalanche, Spate	
S2ID	Large scale system, Storm	2
EMDAT	Storm	
DesInventar	Electricstorm, Hailstorm, Snowstorm	
S2ID	Extreme temperature	3
EMDAT	Extreme temperature	
DesInventar	Frost, Heatwave	
S2ID	Seca (Drought)	4
EMDAT	Drought, Wildfire	
DesInventar	Forestfire	

In Figure 6, we present the diagram of the R2DN instantiation process. In the "Load disasters from selected databases" activity, data extracted from the respective databases' websites is imported into auxiliary tables of the R2DN. In the subsequent activity, "Extract data for MatchUp" the "MatchUp" file is generated, containing the data from the three databases necessary for the analysis of similarities between disasters.

The "Analyze similarities between disasters" activity uses the "MatchUp" file to conduct the analysis of disasters. When it identifies one or more similar disasters, it instantiates the "Parity" file with information related to these two disasters. Finally, the "Instantiate R2DN" activity writes the disasters from the three databases into the R2DN.

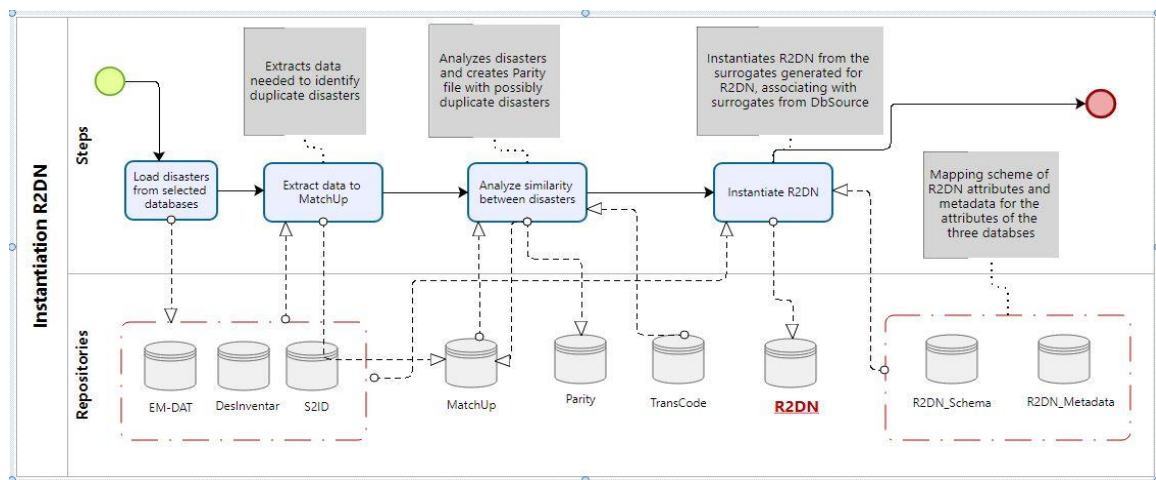


Figure 6. R2DN instantiation process

The "MatchUp" activity, which generates the "Parity" file containing the data of both considered similar disasters, is executed through an SQL command, as illustrated in Figure 7. It is important to note that the "MatchUp" table is invoked twice to enable the comparison of each disaster with all others in search of similarities.

```

ExecMatch X
SELECT d1.baseorigem, d1.idMatch, d1.idbaseorigem, d2.baseorigem, d2.idMatch, d2.idbaseorigem, d1.primDivPais, d2.primDivPais, d1.segDivPais, d2.segDivPais,
d1.data, d2.data, d1.evento, d2.evento, d1.mortos, d2.mortos, d1.feridos, d2.feridos INTO Parity
FROM MatchUp AS D1, MatchUp AS D2
WHERE (D1.pais = D2.pais) AND
(D1.codEvento = D2.codEvento) AND
(D1.baseorigem <> D2.baseorigem) AND
((D1.contadías - D2.contadías) < 10) AND ((D1.contadías - D2.contadías) > -10) AND
(((D2.PrimDivPais LIKE '*' & D1.PrimDivPais & '*') OR (D2.SegDivPais LIKE '*' & D1.SegDivPais & '*')) OR
((D1.PrimDivPais LIKE '*' & D2.PrimDivPais & '*') OR (D1.SegDivPais LIKE '*' & D2.SegDivPais & '*')));
    
```

Figure 7. SQL command executed to pinpoint similar disasters

The conditions for identifying similarities, as can be observed in the "WHERE" clause, include: the same country, the same event code, different databases, a date difference of less than ten days, and similarity in the administrative divisions of the disaster locations. Lastly, the "Instantiate R2DN" activity creates disaster records in R2DN from the "MatchUp" file, incorporating the remaining data from the disasters coming from the auxiliary tables of the three selected databases.

The criteria adopted for selecting disasters to be imported from the three selected databases were as follows:

1. Disasters of hydrological, meteorological, and climatological nature.
2. Occurrence in the period from 2010 to 2015.
3. In EM-DAT, only disasters recorded in Brazil, Argentina, and Chile were considered.
4. In DesInventar, only disasters that occurred in Argentina and Chile were included.

The main consideration guiding all the criteria is the quantity of records available in the Brazilian Disaster Atlas Database, accessible through the website <http://atlasdigital.mdr.gov.br/paginas/downloads.xhtml> (accessed on 10/11/2023). This database covers disasters in Brazil between 1990 and 2022, with a total of 62,273 disaster records. Among these, disasters that fall under the first criterion represent 95.2% of the total.

The disaster records in the DesInventar database for South American countries mostly covered disasters that occurred until 2015. Therefore, we used the criterion of the occurrence period of disasters between 2010 and 2015. The application of this criterion reduced the disasters from the S2ID to 13,143 disaster records.

The choice of Argentina and Chile in South America is due to their presenting similar development indices to Brazil in the region, making the comparison of the impacts of disasters between these countries more equitable. Based on these criteria, the total number of disaster events imported into the auxiliary tables of R2DN was 14,760, as detailed in Table 7.

Table 7. Number of records imported into R2DN

Number of imported disasters in the period 2010-2015	
S2ID	13,143
EM-DAT – Brazil	33
EM-DAT – Argentina	15
EM-DAT – Chile	13
DesInventar – Argentina	1,345
DesInventar – Chile	211
Total	14,760

When considering the criteria for inserting EM-DAT disaster records, which include more than 9 deaths, more than 99 people affected, or the need for international assistance, see Table 8, we assume that all EM-DAT records should be found in the other databases. With this assumption, we face the challenge of locating the 33 EM-DAT disasters that occurred in Brazil in the S2ID, as well as the 28 EM-DAT disasters that occurred in Argentina and Chile in DesInventar.

Table 8. Final result of the MatchUp process

Numbers after running the MatchUp process	
Total records in the Parity file	1,746
Number of similar disasters	873
Similar disasters between EM-DAT and S2ID	686
Similar disasters between EM-DAT and DesInventar	187
Number of disasters not found in S2ID (spelling problem)	3
Number of disasters not found in S2ID	1
Percentage of disasters located in S2ID (except spelling error)	90.63%
Number of disasters not found in DesInventar (spelling problem)	5
Number of disasters not located in DesInventar	7

Percentage of disasters located in DesInventar (except spelling errors)	69.56%
-------------------------------------------------------------------------	--------

Regarding the 33 Brazilian disasters registered in EM-DAT, one was not identified in our search in S2ID, three were not located due to differences in the spelling of locations, and the remaining 29 show similarities with 686 disasters in S2ID. The way disasters are recorded in S2ID, by municipality, creates situations where, for example, a disaster in the Serrana Region of Rio de Janeiro in 2011, resulting in 900 fatalities, has a single record in EM-DAT, while in S2ID, it is recorded in 19 municipalities.

In another example, the floods that occurred in Rio Grande do Sul in July 2014, due to the overflowing of the Uruguay River, impacted 117 cities, as reported in the media. (G1 - O Globo, 2014). This disaster has only one record in EM-DAT, and our process identified 98 similar disasters in S2ID. This highlights that the information from EM-DAT complements that of S2ID, providing a broader view of disasters, while the latter offers a more detailed perspective on disasters.

Another noteworthy result is that when applying EM-DAT criteria - more than nine fatalities or more than ninety-nine people impacted - for including disasters in its database, we note that, of the 13,143 disaster records in S2ID, 1,924 meet these criteria. According to these parameters, these disasters should be included in the EM-DAT database, which only contains 33 disasters, as shown in **Table 7**. Even considering that one EM-DAT disaster can represent dozens of disasters in S2ID, a considerable disproportion is evident. This reinforces the need for countries to develop regional databases suitable for their realities and demands.

The "MatchUp" process found 90.63% of Brazilian disasters from EM-DAT in S2ID, excluding cases with different spellings. In the case of DesInventar, this percentage was 69.56%. After a preliminary analysis, we were unable to explain the discrepancy between the percentages in the two databases. However, we believe that the greater formality in the use of S2ID may, in part, explain this difference in percentages.

CONCLUSIONS

This study introduced the R2DN data repository model as a response to the research question, which was designed to meet the established requirements. These requirements include the ability to assess disasters, compatibility with major global disaster databases, compliance with multilateral agreements aimed at reducing disaster risks, and meeting the specific needs of a country or organization to support its disaster management policies.

The strategy adopted to achieve the research objectives was as follows:

1. Define a methodological approach for the research.
2. Investigate the characteristics of major global databases, considering citations in academic works.
3. Understand the needs of the Brazilian disaster system, S2ID.
4. Analyze data requirements to meet the demands of the Sendai Framework.
5. Select EM-DAT, DesInventar, and S2ID as reference databases for the study.
6. Identify data needs to meet the requirements of the Sendai Framework.
7. Develop a taxonomy, based on the DaLA methodology, to classify the data selected for the proposed repository model.
8. Implement an artifact to demonstrate the feasibility of the proposed solution.
9. Evaluate the obtained results and generated contributions.
10. Indicate possible directions for future research within the scope of this study.

The **first contribution** of this study lies in the data taxonomy, the data model, and the metadata of R2DN. Our initial encounters with the files of some of the studied databases proved to be quite challenging in understanding the information contained in the data. All databases, without exception, provide their information in a disconnected list of data, significantly complicating its comprehension. Additionally, the available metadata did not offer adequate assistance. We believe that this work will be of great help to those wishing to delve into this subject. The classification of data through the developed taxonomy, coupled with the model that exposes the relationships between the data and the associated metadata, will facilitate the work of anyone dealing with this topic.

The **second contribution** we highlight comes from the development of the artifact that implements an instance of R2DN in a commercial database system. The steps in the development, such as mapping attributes from a database to R2DN, describing data import, unifying disaster type codes, and the procedures necessary to identify similar disasters, facilitate the inclusion of disaster records from other databases into R2DN. We believe we have opened a door for anyone interested in the topic, allowing them to expand R2DN with records from other disaster databases.

Regarding the limitations of this work, we acknowledge that using the three databases restricts the scope of the problem, as the comparison between their data is binary in nature. To elaborate, EM-DAT records need to be compared either with S2ID disasters or with those from DesInventar, given the specificity that S2ID exclusively contains records related to Brazilian disasters, while DesInventar has no records of disasters in Brazil. To overcome this limitation, we intend to include a fourth and perhaps a fifth database in a future version of the work.

Future Work Perspectives

The set of categories does not cover all types of disasters, but it can be easily extended to include disasters such as earthquakes, common in certain parts of the world. As we wanted to examine the occurrence of duplicate events, we left out types of events that are rare in Brazil and therefore are not part of the S2ID database. The taxonomy can also be extended or even replaced by a more complete one.

We believe that handling the location information of the disaster is the aspect that needs the most improvement. The way EM-DAT describes the same disaster occurring in various locations in a single record, with only one attribute for this representation, poses challenges. Furthermore, the spelling of locations in this attribute often contains inaccuracies. Since EM-DAT is the most widely referenced disaster database globally, any initiative to integrate disaster data must include it. In our work, whenever we needed to precisely locate an EM-DAT disaster, we resorted to internet searches using information such as country, date, and type of disaster, and were successful in most cases. Automating this search process represents an opportunity for future work.

Disaster management encompasses three basic phases: prevention, response, and recovery. Our work focused on databases that support the response phase, aiming to restore essential activities, and the recovery phase, to return to pre-event normality in the society affected by the incident. We consider that our proposed disaster data is comprehensive regarding the human and material impacts caused by these events. However, it lacks information on risks in the affected locations, as none of the selected databases in the study address this aspect. Therefore, we believe that integrating a database that includes information on disaster risk into the R2DN represents an opportunity.

Finally, to evaluate our solution, it should be used by researchers and professionals in the field of disasters to demonstrate its efficiency and effectiveness. In our initial attempt to produce such evaluation we didn't get enough returns to produce a relevant analysis. We intend to pursue this evaluation soon.

ACKNOWLEDGMENTS

The work of Marcos R. S. Borges was partially supported by the Rio de Janeiro Research Foundation (FAPERJ) under grant # E-26/202.876/2018.

REFERENCES

- Ceped - Centro Universitário De Estudos E Pesquisas Sobre Desastres. (n.d.). *Base De Dados Do Atlas Digital de Desastres do Brasil (Brazilian Digital Disaster Atlas Database)*. (In Portuguese)
- CRED. (n.d.). EM-DAT website. Retrieved from <https://public.emdat.be/>
- CRED. (2020a). *EM-DAT Guidelines : Data Entry , Field Description / Definition*. 1–11.
- CRED. (2020b). Human Cost of Disasters. *Human Cost of Disasters*. <https://doi.org/10.18356/79b92774-en>
- CRED. (2021). Missing data on economic losses variables from EM-DAT. *Cred Crunch*, (63), 1–2. Retrieved from <https://www.emdat.be/cred-crunch-63-missing-data-economic-losses-variables-em-dat-0>
- FAPEU. (2020). *Relatório de danos materiais e prejuízos decorrentes de desastres naturais no Brasil: 1995 – 2019 /Banco Mundial.Global Facility for Disaster Reduction and Recovery*. (In Portuguese)
- G1 - O Globo. (2014). O Globo. *O Globo*, p. 1. Retrieved from <https://g1.globo.com/rs/rio-grande-do-sul/noticia/2014/07/196-mil-pessoas-estao-fora-de-casa-apos-chuva-no-rs-diz-defesa-civil.html>
- Gall, M. (2015). The suitability of disaster loss databases to measure loss and damage from climate change. *International Journal of Global Warming*, 8(2), 170–190. <https://doi.org/10.1504/IJGW.2015.071966>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Integração, M. da. (n.d.). Atlas dos desastres brasileiros (Atlas of Brazilian disasters). Retrieved from <http://atlasdigital.mdr.gov.br/paginas/downloads.xhtml> (In Portuguese)

- IRDR. (2014). *IRDR DATA Reports - Perils Classification and Hazards Glossary*.
- Khankeh, S. A. M. F. N. N. all 8 authors Hamid R. K. R. (2022). Worldwide disaster loss and damage databases: A systematic review. *Journal of Education and Health Promotion*. https://doi.org/10.4103/jehp.jehp_1525_20
- Koç, G., & Thieken, A. H. (2018). The relevance of flood hazards and impacts in Turkey: What can be learned from different disaster loss databases? In *Natural Hazards* (Vol. 91). <https://doi.org/10.1007/s11069-017-3134-6>
- Ministério do Desenvolvimento Regional - S2ID (“Ministry of Regional Development”). (n.d.). Retrieved November 29, 2022, from <https://www.gov.br/mdr/pt-br/assuntos/protecao-e-defesa-civil/sistema-integrado-de-informacoes-sobre-desastres#:~:text=O> (In Portuguese)
- Panwar, V., & Sen, S. (2020). Disaster Damage Records of EM-DAT and DesInventar: A Systematic Comparison. *Economics of Disasters and Climate Change*, 4(2), 295–317. <https://doi.org/10.1007/s41885-019-00052-0>
- SEDEC. (n.d.). COBRADE. Retrieved from https://www.gov.br/mdr/pt-br/centrais-de-conteudo/publicacoes/protecao-e-defesa-civil-sedec/DOCU_cobrade.pdf
- UFSC. (2012). Curso de Capacitação dos Gestores - S2ID (“Manager Training Course”). In *Curso de Capacitação dos Gestores de Defesa Civil para uso do Sistema Integrado de Informações sobre Desastres (S2iD)*. Retrieved from <https://s2id.mi.gov.br/>
- UN. (2015). Sendai Framework 2015-2030. Retrieved from https://www.preventionweb.net/files/43291_sendaiframeworkfordren.pdf
- UN Office for Disaster Risk Reduction (UNISDR). (n.d.). DesInventar Website. Retrieved from <https://www.desinventar.net/DesInventar/>
- UN Office for Disaster Risk Reduction (UNISDR). (2019). *Desinventar Sendai 10.1.2*. (February), 57. Retrieved from https://www.desinventar.net/documentation/DISendai_DataManagement_2019-2.pdf
- Wirtz, A., Kron, W., Löw, P., & Steuer, M. (2014). The need for data: Natural disasters and the challenges of database management. *Natural Hazards*, 70(1), 135–157. <https://doi.org/10.1007/s11069-012-0312-4>
- World Bank. (2010). *Damage, Loss and Needs Assessment: Guidance Notes. Volume 1: Design and Execution of a Damage, Loss and Needs Assessment*. 86. Retrieved from www.worldbank.org