Position Paper
Cascading Effects Analysis Enabled by Semantic Interoperability in the Resilience Data Space

Authors: Federico Gutt [1] (Speaker), Martin Huschka [1], Alexander Stolz[1, 2].

[1] Fraunhofer Institute for High-Speed Dynamics, Ernst-Mach-Institut, EMI, Freiburg, Germany;
[2] Albert-Ludwigs-Universität Freiburg, Institut für nachhaltige technische Systeme, Emmy-Noether-Straße 2, 79110 Freiburg

1 Motivation

In recent years there has been an increase in the number of catastrophes and crises that have negatively affected society, its systems, and processes. Some examples of these events are the global pandemic of COVID-19 from 2020, the earthquake in Turkey and Syria in February 2023, as well as the floods and extreme rainfall in Germany in 2021 and in 2024.

One of the most important components of today's society, and at the same time most vulnerable to these crises, have been the critical infrastructure (CI) systems. These systems provide essential services such as health care, electricity, telecommunications, transportation, etc. Hence, experts in safety and security engineering as well as crisis management have recently been interested in analyzing these CI systems' resilience [1].

However, one of the main characteristics of CI systems is their high level of interdependence at different levels and facets [2]. This means that failures in one system can propagate and cause unexpected cascading effects in other systems. For instance, the surge in COVID-19 cases had a ripple effect on various critical infrastructure systems. It significantly affected the job market as companies downsized or closed, impacting the economic system. This workforce reduction, along with health protection measures taken due to rising cases, disrupted other systems, such as transportation, decreasing air travel.

Therefore, modeling and simulating interdependencies and cascading effects in these complex CI systems has become vital for resilience analyses [3]. However, these analyses demand high-quality and trustworthy data of various kinds [1]. Following the previous example, an exhaustive resilience analysis requires diverse data from various actors, such as COVID-19 case numbers, reported job positions, short-time workers, unemployed individuals, and flight departures and arrivals. This diverse data and its interoperability are crucial for a comprehensive understanding of the situation.

Currently, crisis managers and resilience experts face significant challenges due to the absence of standardized methodologies for collecting, managing, and aggregating the heterogeneous and decentralized data essential for their analyses. Therefore, for municipal resilience analysis, a trusted network for data sharing and interoperability of the data provided is paramount.

To overcome these challenges and enable cascading effects analysis, this position paper describes how semantic interoperability within the resilience data space is exploited for a specific application.

2 State of the art in cascading effects analysis

As highlighted in the previous section, CI systems become increasingly vulnerable when disruptive events trigger cascading effects, affecting a larger number of their components [4]. Traditional risk analysis methods, which typically focus on single systems, prove insufficient under these circumstances due to the necessity of domain expert knowledge and their lack of consideration for complex, interconnected systems [2]. Therefore, it has become essential to research more advanced methods that allow modeling the interdependencies of CI systems for resilience analysis [5].

Research into modeling methods for interdependent CI systems has been extensive [6], [7], [8], yet few studies incorporate the concept of resilience as comprehensively as the classification proposed by Ouyang [3]. Ouyang groups these methods into six categories: *(1) Agent-based approaches* that simulate CI systems as networks of interacting agents [9], *(2) System dynamics-based approaches* that utilize causal-loop diagrams [10], *(3) Economic theory-based approaches* employing models like input-output and computable general equilibrium [11], *(4) Network-based approaches* that depict CI systems through nodes and interlinks [3], *(5) Empirical approaches* which derive insights from historical data and expert experiences [3], and *(6) Other statistical methods* like Bayesian networks and Granger causality tests to describe the causality and dynamic behavior of interdependencies within CI systems.

Since each method has its advantages and weaknesses, comparing them to each other and choosing the most appropriate one depends on the application context. However, researchers such as Ouyang [3], Trucco et al.[13], and Dao et al. [1], identify three limitations and challenges faced by the methods previously mentioned:

1. **Incomplete modeling:** Many approaches focus only on subsets of CI systems, not capturing all necessary components and their interconnections comprehensively.

2. **Integration and co-simulation difficulties:** Integrating different models from various approaches is complex due to methodological differences and the need for domain expertise.

3. **Heterogeneous and decentralized data:** Data-driven methods face challenges in accessing and integrating diverse data from various domains which are often stored and owned by different stakeholders.

The development of a Resilience Data Space powered by semantic technologies provides effective solutions to these complex challenges. On the one hand data spaces enable the automated sharing of diverse decentralized data from trusted actors, facilitating the analysis of potential cascading effects across different CI systems. On the other hand, semantic technologies provide "a systematic representation of heterogeneous systems in terms of entities and their interdependencies for study and simulation purposes" [14]. By employing semantic-based structures, such as ontologies and knowledge graphs, semantic interoperability within the data space can be ensured, effectively formalizing, and modeling the meaning and context of data, regardless of their heterogeneity or decentralization.

3 Semantic Interoperability in the Resilience Data Space

The primary goal of the Resilience Data Space is to enhance data-driven decision-making in scenarios that demand high reliability and rapid response, such as crisis management and resilience analyses of critical infrastructure systems. A critical element in achieving this is trust and semantic interoperability, ensuring that data from diverse sources can be integrated and understood in a unified manner. For systems interoperability among the various actors, the technologies and standards of the International Data Spaces are employed within the Resilience Data Space. The Eclipse Data Space connector [15] enables technical interoperability of the systems, facilitating data exchange and consumption between public institutions at different governmental levels —country, state, and municipality— for resilience analysis.

Semantic interoperability is crucial in any data space, enabling actors to effectively share, interpret, and use data while maintaining consistency and accuracy across various datasets and domains. It is particularly vital in crisis management, where it fosters trust in data for decision-making. Within the Resilience Data Space, a semantic data catalog serves as a data fabric, interconnecting and aggregating data while adhering to the following requirements regarding the knowledge provided by the data:

i. **Origin:** Clear documentation of where and how data is sourced, ensuring authenticity and reliability.
ii. **Data Structure:** For AI-ready data, the relevant dimensions as well as the timely and geographical resolution of data must be machine-readable.
iii. **Data Quality:** To ensure trustworthy analysis results, information about the quality of a dataset using metrics like completeness and consistency is needed.

A standardized metadata schema is adopted in response to these requirements and in alignment with the International Data Spaces Information Model [16]. This schema is based on the Data Catalog Vocabulary (DCAT) [17], a RDF vocabulary designed to enhance interoperability between web-published data catalogs. Specifically, the DCAT Application Profile for data portals in Germany (DCAT-AP.de [18]) is employed as the foundation for the semantic data model in the Resilience Data Space.DCAT-AP.de provides comprehensive metadata that enhances the general understanding and usability of data, describing key aspects of the datasets like timely and geographical resolution, to trace back the origin and context of the data (i.). Moreover, the RDF Data Cube Vocabulary [19] defines the dataset's dimensions and structure (ii.). Finally, the use of the Data Quality Vocabulary [20] ensures that all datasets in the Resilience Data Space are enriched with detailed information on data quality (iii.), modeling the consistency or completeness of the data.

By leveraging IDS components in combination with standardized and interconnected vocabularies, the Resilience Data Space ensures that data from various sources is not only technically and semantically interoperable but also adheres to high standards of quality and reliability, crucial for the contexts in which it is intended to be used.

4 A knowledge graph-based approach to analyze cascading effects in CI systems

Building on the aforementioned semantic interoperability foundations within the Resilience Data Space, a knowledge-graph-based approach was developed for modeling CI interdependencies discovered using statistical methods on heterogeneous and decentralized data from these systems.

This approach comprises three key steps: (1) data selection, (2) statistical interdependency analysis, and (3) semantic modeling of interdependencies into a knowledge graph. This semantic graph subsequently serves as a robust platform for knowledge retrieval for identifying and visualizing cascading effects within CI systems, providing crisis managers with a comprehensive and explainable view of potential impacts and vulnerabilities. To illustrate this approach, let's revisit the COVID-19 pandemic scenario from the motivation section, where the surge in cases affected the job market and air transportation.

In the data selection (1), relevant metadata of the datasets related to a specific CI is retrieved from the semantic data catalog. Taking the "Infektionen" dataset as an example, which reflects the number of COVID-19 infections in Germany, the use of semantic vocabularies in the data space discloses that the dataset is in CSV format, covers Germany at the federal level, and updates daily (see Figure 1). This semantically represented metadata provides crucial insights about the dataset's origin and context.

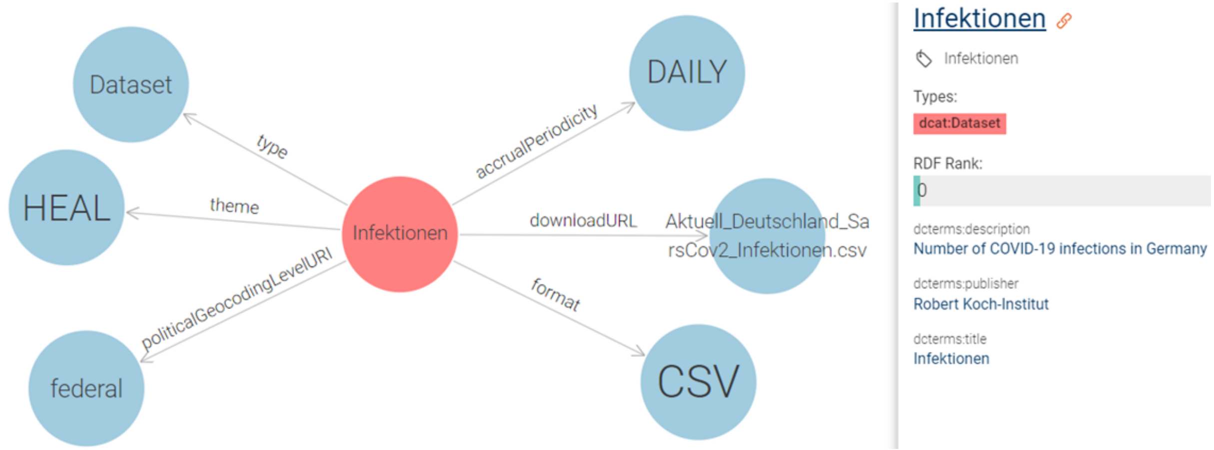

Figure 1: Snapshot from GraphDB [21] showing metadata retrieval for the "Infektionen" dataset, including format, geographical and temporal resolution, access URL, and theme.

The first step of the statistical interdependency analysis (2) is data acquisition and cleaning. The interdependencies between the datasets are then quantified using statistical methods such as the Pearson correlation [22] and the Granger causality test [23] in an automatic fashion. The resulting pearson-correlation-coefficient and granger-test-p-value for each interdependency are semantically modeled in a knowledge graph (3) and interlinked with the metadata from the data catalog related to these datasets.

In this knowledge graph, the interdependencies calculated statistically are represented as RDF triples, indicating that a Dataset1 has an impact on a second Dataset 2 (see left side of Figure 2). For quantification of this impact between the nodes, RDF-star [24] is used to add the information of the concrete value of this interdependency using the pearson-correlation-coefficient and the and granger-test-p-value (see right side of Figure 2).

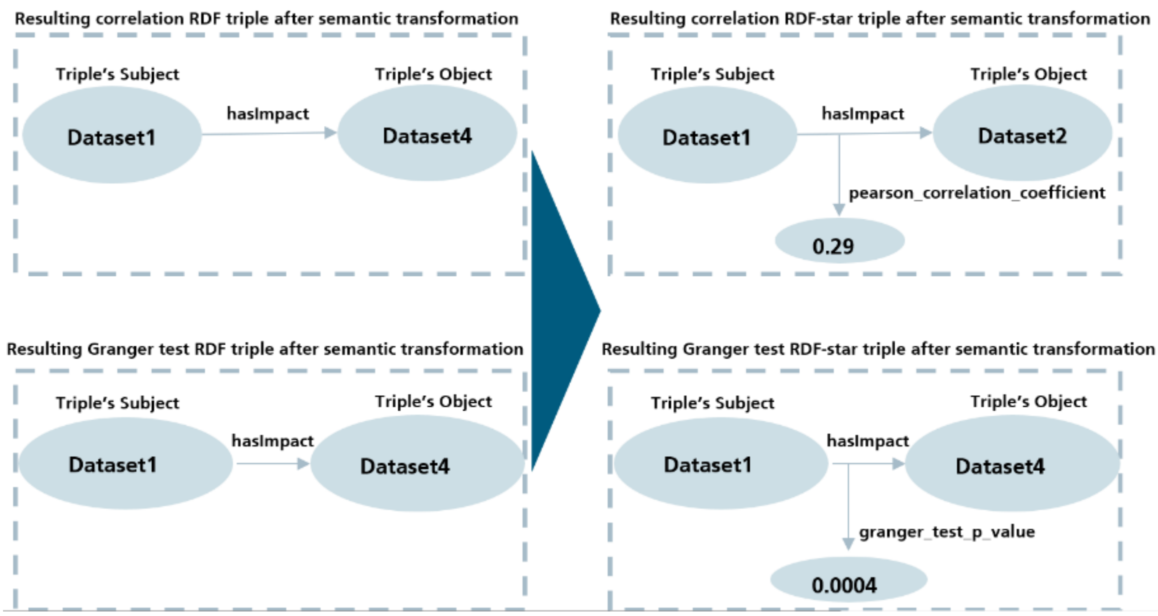

Figure 2: RDF (left) and RDF-star (right) model of interdependencies between datasets from CI systems.

By using queries that implement graph path search algorithms to traverse the knowledge graph, potential paths reflecting the sequence of repercussions between source, intermediate, and/or destination nodes can be identified. When combined with the enriched information about the interdependencies and metadata of the CI datasets, this provides crisis managers with valuable insights for identifying and visualizing cascading effects and their context. Additionally, the calculated interdependency values can be used to filter query the results within a significance range defined by a Pearson correlation or Granger causality test.

Following the example in this section, a query on the knowledge graph using the mentioned algorithms, with the number of COVID-19 infections, "Infektionen", as the source node and Frankfurt Airport departures, "F_FRA", as the destination node, generates an automatic cascade of interdependencies. This chain, which has a maximum path length of 3 and filters interdependencies based on a Granger causality test p-value significance of $\alpha = 0.05$, is shown in Figure 3.

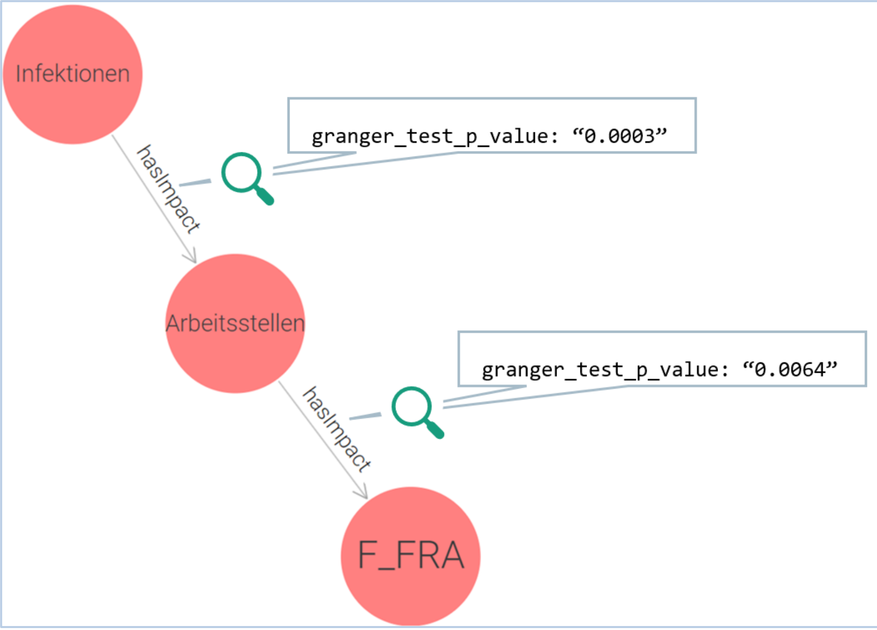

Figure 3: Snapshot from GraphDB showing the cascade of interdependencies between COVID-19 infections and Frankfurt Airport departures, filtered by Granger causality test p-value significance of $\alpha = 0.05$. Values are manually added due to visualization limitations in the software.

The visualization shown in Figure 3 allows the crisis manager to identify the cascading effects triggered by the rise in COVID-19 cases. It reveals that this source event impacted the number of job positions, which in turn affected the number of flight departures at Frankfurt airport. The information contained in the triple's edges, added manually to Figure 3 for explanation purposes due to RDF-star visualization limitations in graph database GraphDB [21], provides statistical evidence of the interdependencies based on the Granger causality test. This enhances trust and explainability of the displayed interdependencies. Lastly, the provided interactive visualization allows intuitive graph navigation and node expansion to gain more context on the datasets by visualizing the metadata values (see Figure 2)

5 Conclusions and Outlook

The Resilience Data Space enables cascading effects analysis by ensuring semantic interoperability. A knowledge graph-based approach built on this foundation effectively identifies and visualizes potential impacts in critical infrastructure systems, as demonstrated in the COVID-19 scenario. This approach enhances decision-making in crisis management, contributing to a more explainable CI resilience analysis.

Since the technologies presented in this paper are still under development, there is room for improvement. Enhancements could include better visualization of cascading effects using Labeled Property Graphs (LPG), research on more advanced algorithms for quantifying interdependencies, and further development of the Resilience Data Space's architecture and its components.

6 Acknowledgements

This position paper is based on the author's bachelor thesis [25], developed under the scope of HERAKLION project [26] (funding code 13N16293). This project is funded by the Federal Ministry of Education and Research (BMBF) within the framework of the "Research for Civil Security" program of the federal government and is an approved project under the special Corona action measure.

7 References

[1] J. Dao, S. T. Ng, Y. Yang, S. Zhou, F. J. Xu, and M. Skitmore, "Semantic framework for interdependent infrastructure resilience decision support," *Automation in Construction*, vol. 130, p. 103852, Oct. 2021, doi: 10.1016/j.autcon.2021.103852.

[2] S. Creese, M. H. Goldsmith, and A. O. Adetoye, "A logical high-level framework for Critical Infrastructure resilience and risk assessment," in *2011 Third International Workshop on Cyberspace Safety and Security (CSS)*, Sep. 2011, pp. 7–14. doi: 10.1109/CSS.2011.6058564.

[3] M. Ouyang, "Review on modeling and simulation of interdependent critical infrastructure systems," *Reliability Engineering & System Safety*, vol. 121, pp. 43–60, Jan. 2014, doi: 10.1016/j.ress.2013.06.040.

[4] P. Rogers, J. Coaffee, and D. Murakami Wood, *The Everyday Resilience of the City: How Cities Respond to Terrorism and Disaster*. 2008.

[5] K. Thoma, B. Scharte, D. Hiller, and T. Leismann, "Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches," *European journal for security research*, vol. 1, no. 1, Art. no. 1, 2016, doi: 10.1007/s41125-016-0002-4.

[6] Idaho National Laboratory (INL), "Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research," INL/EXT-06-11464, 911792, Aug. 2006. doi: 10.2172/911792.

[7] I. Eusgeld and W. Kröger, "Comparative Evaluation of Modeling and Simulation Techniques for Interdependent Critical Infrastructures".

[8] C. Griot, "Modelling and simulation for critical infrastructure interdependency assessment: a meta-review for model characterisation," *International Journal of Critical Infrastructures*, Dec. 2010, Accessed: Jun. 14, 2024. [Online]. Available: https://www.inderscienceonline.com/doi/10.1504/IJCIS.2010.037453

[9] "Analyzing maintenance strategies by agent-based simulations: A feasibility study - ScienceDirect." Accessed: Jun. 14, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832009000350

[10] N. Kollikkathara, H. Feng, and D. Yu, "A system dynamic modeling approach for evaluating municipal solid waste generation, landfill capacity and related cost management issues," *Waste Management*, vol. 30, no. 11, pp. 2194–2203, Nov. 2010, doi: 10.1016/j.wasman.2010.05.012.

[11] A. Rose, "Economic Principles, Issues, and Research Priorities in Hazard Loss Estimation," in *Modeling Spatial and Economic Impacts of Disasters*, Y. Okuyama and S. E. Chang, Eds., Berlin, Heidelberg: Springer, 2004, pp. 13–36. doi: 10.1007/978-3-540-24787-6_2.

[12] Y. Yang, S. T. Ng, F. J. Xu, M. Skitmore, and S. Zhou, "Towards Resilient Civil Infrastructure Asset Management: An Information Elicitation and Analytical Framework," *Sustainability*, vol. 11, no. 16, p. 4439, Aug. 2019, doi: 10.3390/su11164439.

[13] P. Trucco, B. Petrenj, S. Bouchon, and C. D. Mauro, "Ontology-based approach to disruption scenario generation for critical infrastructure systems," *IJCIS*, vol. 12, no. 3, Art. no. 3, 2016, doi: 10.1504/IJCIS.2016.079022.

[14] L. Galbusera and G. Giannopoulos, "Exploiting Web Ontologies for Automated Critical Infrastructure Data Retrieval," in *Critical Infrastructure Protection XI*, vol. 512, M. Rice and S. Shenoi, Eds., in IFIP Advances in Information and Communication Technology, vol. 512. , Cham: Springer International Publishing, 2017, pp. 119–136. doi: 10.1007/978-3-319-70395-4_7.

[15] "eclipse-edc/Connector." Eclipse Dataspace Components, Jun. 14, 2024. Accessed: Jun. 14, 2024. [Online]. Available: https://github.com/eclipse-edc/Connector

[16] "International Data Spaces Information Model." Accessed: Jun. 15, 2024. [Online]. Available: https://international-data-spaces-association.github.io/InformationModel/docs/index.html#

[17] "Data Catalog Vocabulary (DCAT) - Version 3." Accessed: Jun. 13, 2024. [Online]. Available: https://www.w3.org/TR/vocab-dcat-3/

[18] J. Zedlitz, E. Priefer, and et al., "DCAT-AP.de Spezifikation 2.0." Accessed: Jun. 13, 2024. [Online]. Available: https://www.dcat-ap.de/def/dcatde/2.0/spec/

[19] "The RDF Data Cube Vocabulary." Accessed: Jun. 13, 2024. [Online]. Available: https://www.w3.org/TR/vocab-data-cube/

[20] "Data on the Web Best Practices: Data Quality Vocabulary." Accessed: Jun. 13, 2024. [Online]. Available: https://www.w3.org/TR/vocab-dqv/

[21] "General — GraphDB 10.1.0 documentation." Accessed: Jun. 15, 2024. [Online]. Available: https://graphdb.ontotext.com/documentation/10.1/

[22] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesth Analg*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.

[23] G. Xu and X. Zhang, "Statistical analysis of resilience in an air transport network," *Frontiers in Physics*, vol. 10, 2022, Accessed: Jan. 31, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphy.2022.969311

[24] D. Arndt, J. Broekstra, B. DuCharme, and et al., "RDF-star and SPARQL-star." Accessed: Jun. 13, 2024. [Online]. Available: https://www.w3.org/2021/12/rdf-star.html

[25] F. Gutt, M. Huschka, and A. Stolz, "A Knowledge-Graph-Based Approach for Modeling Interdependencies Discovered Using Statistical Methods on Heterogenous and Decentralized Data for Quantitative Resilience Analysis - Bachelor's Thesis in Embedded Systems Engineering at the Albert-Ludwigs-Universität Freiburg, Faculty of Engineering," May 2023.

[26] "HERAKLION | Heuristische Resilienzanalysen für Kommunen mittels Datenraumfunktionalitäten." Accessed: Jun. 14, 2024. [Online]. Available: https://www.heraklion-projekt.de/