2024

Contributions of AI to advance interoperability with data mediators

Torben Ukena
University Leipzig, Germany, torben.ukena@uni-leipzig.de

Rainer Alt
University Leipzig, Germany, rainer.alt@uni-leipzig.de

Contributions of AI to advance interoperability with data mediators

Research in Progress

Torben Ukena[1], Rainer Alt[1]

[1] Leipzig University, Information Systems Institute, Leipzig, Germany
{torben.ukena,rainer.alt}@uni-leipzig.de

Abstract. This study presents an innovative approach to advancing interoperability in information systems through the development of an Artificial Intelligence (AI)-based data mediator. Although standards have contributed to interoperability among disparate systems, the lack of universal standards still requires tools for data mediation. To reduce the substantial need for manual configuration of these systems, this paper outlines a strategy for translating data between two systems with different data schemas automatically. Unlike traditional methods, the proposed data mediator leverages recent advancements in AI to facilitate automatic mapping of heterogeneous data.

Keywords: Artificial Intelligence, Data Mediation, Interoperability, Standards

1 Introduction

In our interconnected world, standards serve as foundational blocks for complex systems. They ensure consistency, reliability, and safety across domains, from the dimensions of a bolt to the protocols governing international telecommunications. Standards are key as they enable disparate systems and components to work seamlessly together, fostering innovation and efficiency (Shivakumar, 2022). This is illustrated by the healthcare sector, where interoperability potentials are estimated at $77.8 billion per year for the United States alone (Walker et al., 2005). A more recent study emphasized that "the efficiency potential currently lost in the fragmented interplay of stakeholders, sectoral boundaries, and limited care coordination […] account for up to 25% of healthcare spending in Europe and the US" (Pidun et al., 2021, p. 1). Although interoperability has been a persistent problem for decades, the evolution of standards and mediation technologies still requires a significant amount of manual activity. To increase automation in this area, a promising area to explore is Artificial Intelligence (AI) that contributes facilities in pattern recognition to enhance interoperability. From this background, this research analyzes existing interoperability concepts and proposes to create an AI-based solution that is capable of managing heterogeneous data to contribute to the broad field of interoperability. For the sake of consistency, this paper repeatedly

19th International Conference on Wirtschaftsinformatik,
September 2024, Würzburg, Germany

refers to examples from the healthcare sector. However, it should be noted that the proposed AI model can be applied across other domains where networking among heterogeneous systems is widespread, e.g. retailing, banking or the automotive industry.

2 Levels and Status of Interoperability

2.1 Definitions

Interoperability is defined by the European Commission and the European Interoperability Framework as “the ability of information and communication technology systems and of the business processes they support to exchange data and to enable sharing of information and knowledge” (IDABC, 2004). In order to apply this definition, it is necessary to break down interoperability to different levels, as shown in Figure 1.

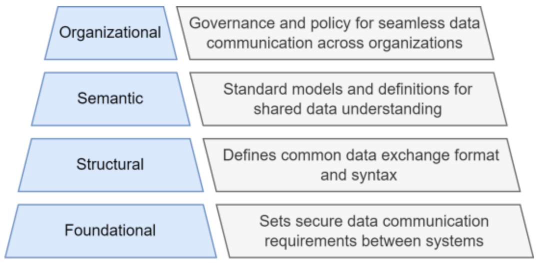

Figure 1. Interoperability levels (from Adebesin et al. (2013))

Data mediation is an approach designed to solve the data interoperability problem by enabling the translation of data between two systems with different data schemas. The difference between data mediators and a mapping is the ability to automatically generate such translations (Renner et al., 1999).

2.2 Technical Interoperability

Technical interoperability at the foundational level has been extensively researched. With the introduction of ISO/OSI and TCP/IP standards, fundamental interoperability is now rarely an issue (Sunyaev et al., 2023). Similarly, many well-established paradigms for structural interoperability exist. Representational State Transfer (REST) is probably the most widely used architecture style. The HTTP protocol is defined by its standard methods for communication, and exchanging data is done using standard formats, typically JSON or XML. REST has become the de-facto standard for offering a service on the web due to its ease of use and its long-term presence (Neumann et al., 2021). In healthcare, Fast Healthcare Interoperability Resources (FHIR) is evolving as a promising standard which uses REST under the hood.

2.3 Semantic Interoperability

With foundational and structural interoperability in place, the next step is to get a common understanding of the meaning of the data, i.e. semantic interoperability. Currently, each organization has its own REST Application Programming Interface (API) with different endpoints and data models. Consider the following example of API responses from their respective sleep endpoint from Fitbit and Oura[1]. Although both provide sleep data via the API in JSON format, the data model is not interoperable. For example, they both use a day key but require different access methods. If an analytics platform wanted to use sleep data from both vendors (e.g., to determine a wellbeing score), it would have to manually map the data and refer to each vendor's documentation for data ranges.

Table 1. Example of sleep API responses from wearables

Oura Sleep API	FitBit Sleep API

{
 "id": "string",
 "efficiency": 69,
 "awake_time": 0,
 "time_in_bed": 0,
 "day": "2019-08-24",
 "total_sleep_duration": 0
}
	{
	 "sleep":
	 [{
	  "dateOfSleep": "2020-02-21",
	  "efficiency": 42,
	  "duration": 27720000,
	  "infoCode": 0
	 }]
	}

To address the lack of semantic interoperability, many domain-specific standards have been developed. These standards often define the entire data representation and include terminologies used within the domain. In healthcare, the Fitbit and Oura API response could be fully represented as a FHIR resource. This would allow for a universal data representation with clear units, value ranges, and terminology. However, in healthcare, as in any industry, FHIR is not the only standard in use. In fact, in the U.S. healthcare sector alone, there are more than 40 different standards development organizations creating standards for interoperability (HIMSS, 2024, p. 1).

Similarly, other sectors such as IoT are facing the same issues as Noura et al. (2019, p. 807) claims that “there are currently several different academia, industry, and standardization bodies aiming to solve IoT system interoperability. It is not likely that a common set of standards will be universally accepted which will allow IoT devices and platforms to work together.” They further state that implementing interoperability should not require major changes to existing systems.

2.4 Mapping Approaches

Therefore, instead of designing another standard for semantic interoperability, a new layer could leverage existing standards and systems (Braunstein, 2018, p. 36). In doing

[1] Further information about the API responses from Fitbit and Oura can be found in the Fitbit API Documentation (https://dev.fitbit.com/build/reference/web-api/sleep) and Oura API Documentation (https://cloud.ouraring.com/v2/docs#tag/Sleep-Routes).

so, the improvement of AI architectures in recent years is promising to help address interoperability issues. The conducted research therefore focuses on the development of an AI data mediator that is trained using interoperable data schemas to automatically map heterogeneous data into a standard.

Table 2. Interoperability levels for different data mapping approaches

<table>
<tr><th></th><th>Manual mapping (Dos Reis et al., 2015, p. 1473)</th><th>Ontology-based mapping (Amrouch and Mostefai, 2012, pp. 2–3)</th><th>AI data mediator</th></tr>
<tr><td>Foundational</td><td>achieved</td><td rowspan="2">achieved</td><td rowspan="3">achieved</td></tr>
<tr><td>Structural</td><td rowspan="2">limited to use case</td></tr>
<tr><td>Semantic</td><td>limited to use case, occurrence of lexical and semantic mismatches</td></tr>
<tr><td>Organizational</td><td>not achieved</td><td>not achieved</td><td>limited, depends on placement within the organization (see section 3.1)</td></tr>
</table>

The concept of a data mediator per se is not innovative. Various approaches to data and process mediation have been proposed and evaluated. Pessoa et al. (2008) examined different methodologies and highlighted the notable disadvantages of manually creating and managing ontologies for mediation purposes. Ali and Chong (2019) attempted to address the issue of manual mapping by introducing a semantic annotation algorithm that utilizes deep representation learning to align different ontologies. However, focusing on ontologies requires the translation of structured data into semantic ontology models which not only introduces additional complexity but also requires the standards to have a readily available ontological representation. Besides this, little research has been conducted to overcome the weaknesses of current solutions. Table 2 provides a brief overview of the level of interoperability each approach covers and illustrates the disadvantages of the state-of-the-art methods.

3 Evaluation of an AI Data Mediator

3.1 Example of a Practical Use Case

Consider the following practical example of such an AI data mediator. A new system enables a home sleep lab that sends key data from wearables and other sensors to an analytics platform. The analytics platform evaluates the data and makes recommendations to improve sleep quality. The patient consults a physician regarding sleep problems, the doctor prescribes the sleep lab, and the analytics system sends the data to the doctor's system for evaluation via the Gematik Telematic Infrastructure (TI) using the

Communication in Medicine (KIM) standard. Even within this simplified example there are many different data formats as shown in Figure 2.

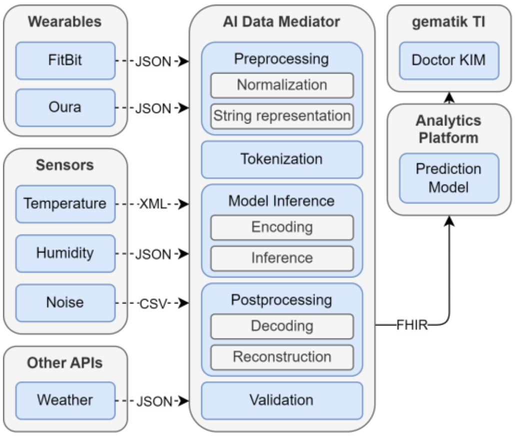

Figure 2. High level concept of the AI data mediator

In this example each provider uses custom proprietary formats. Therefore, both the analysis platform and the doctor's system need to develop connectors to map the data. With an AI as a data mediator, this process could be automated by sending the HTTP requests from the API via the data mediator, which converts the data into a standardized format. There are several ways in which the data mediator could be integrated. Firstly, it could act as middleware, as shown in Figure 2. This would be comparable to a star network topology, with the data mediator act as a proxy for all necessary requests. Another method is to include the data mediator as part of each API. This approach would allow for a fully connected network, but each node would require its own instance of the data mediator. Identifying the optimal placement of the data mediator is crucial for practical implementation and part of the work plan as outlined in section 4.

3.2 Proposed AI Architecture

ChatGPT proves that the implementation of an AI that enables mapping to standards is possible. With appropriate prompts, ChatGPT can effectively convert relevant information into FHIR. However, a dedicated AI is required for practical implementation. After considering several architectures, including traditional machine learning models, rule-based systems, and deep learning approaches, the most suitable AI architecture appears to be the Transformer-based models, specifically tailored for sequence-to-sequence tasks. Among these, the Text-to-Text Transfer Transformer (T5) model stands out for its versatility and capability to handle complex structured data translation tasks (Raffel et al., 2019).

In the context of developing a data mediator, the concept of sentinel tokens has been considered. Sentinel tokens are predefined, unique markers or symbols used to indicate specific structural elements, boundaries, or types of content within a sequence of data

(Raffel et al., 2019, pp. 12–13). However, the use of sentinel tokens would require both the source and target formats to be serialized with tokens. Since the model should be able to handle a variety of standards, it would not be possible to serialize every single input and output data for training. In addition, the real-world data that the model would work with would also need to be serialized. Therefore, it was decided to not implement sentinel tokens and instead encode the format as string representation.

Another important aspect of the evaluated AI architecture of the data mediator is the validation layer. For major standards like FHIR, there are Command Line Interface (CLI) or API validators available to check conformance with the specification. The output from the T5 model will serve as the input for the validation layer. The validation layer will first perform a syntactical validation of the underlying data format, if possible. If the data format passes the syntactic validation, the corresponding validation API or CLI will be used to check if it meets the standard specification. If either of these steps fails, the T5 model will be used iteratively to fix the mismatches.

4 Workplan

To further advance the development of the AI data mediator, several follow-up activities are planned, and the next steps are outlined below.

1. Evaluation of a proper placement of the AI data mediator as outlined in section 3.1. If placed well, it could provide pseudo-interoperability, e.g. by transforming HL7 v2 into FHIR and being addressed as an interface rather than the underlying system.
2. A systematic literature review is planned to gain a detailed understanding of existing standards. The next step involves systematically collecting the detected standards, including specification schemas, example files, and definitions.
3. The T5 architecture for the data mediator must be defined in accordance with its placement.
4. To prepare the training data, it is necessary to create and validate examples of mappings from various input formats to the collected standards. This process can be partially automated by utilizing existing manual mappers.
5. The T5 Model will be trained, validated, and fine-tuned using all collected data and a test set. The model's design, performance, and (hyper-)parameters will be evaluated in a comprehensive computational study.

The research aims to evaluate a model for enhancing interoperability in practice. To achieve this, it may be necessary to focus on a specific domain and integrate the data mediator more accurately. However, it is important to note that the AI data mediator can be applied beyond the respective domain. The healthcare sector is a suitable domain for studying interoperability challenges. We maintain close communication with experts from the Medical Center of the University of Leipzig to develop a theoretically sound and practically applicable model.

References

Adebesin, F., Foster, R., Kotzé, P. and van Greunen, D. (2013), "A Review of Interoperability Standards in E-health and Imperatives for their Adoption in Africa", *South African Computer Journal*, Vol. 50.

Ali and Chong (2019), "Semantic Mediation Model to Promote Improved Data Sharing Using Representation Learning in Heterogeneous Healthcare Service Environments", *Applied Sciences*, Vol. 9 No. 19, p. 4175.

Amrouch, S. and Mostefai, S. (2012), "Survey on the literature of ontology mapping, alignment and merging", in *2012 International Conference on Information Technology and e-Services, 24.03.2012, Sousse, Tunisia*, IEEE, pp. 1–5.

Braunstein, M.L. (2018), "Health Care in the Age of Interoperability: The Potential and Challenges", *IEEE pulse*, Vol. 9 No. 5, pp. 34–36.

Dos Reis, J.C., Pruski, C. and Reynaud-Delaître, C. (2015), "State-of-the-art on mapping maintenance and challenges towards a fully automatic approach", *Expert Systems with Applications*, Vol. 42 No. 3, pp. 1465–1478.

HIMSS (2024), "Interoperability in Healthcare", available at: https://www.himss.org/resources/interoperability-healthcare.

IDABC (2004), "European Interoperability Framework for Pan-European e-Government Services (EIF) Version 1.0.", *European Communities*.

Neumann, A., Laranjeiro, N. and Bernardino, J. (2021), "An Analysis of Public REST Web Service APIs", *IEEE Transactions on Services Computing*, Vol. 14 No. 4, pp. 957–970.

Noura, M., Atiquzzaman, M. and Gaedke, M. (2019), "Interoperability in Internet of Things: Taxonomies and Open Challenges", *Mobile Networks and Applications*, Vol. 24 No. 3, pp. 796–809.

Pessoa, R., Quartel, D. and van Sinderen, M. (2008), "A Comparison of Data and Process Mediation Approaches", *Journal of Vacuum Science & Technology A - J VAC SCI TECHNOL A*.

Pidun, U., Knust, N., Kawohl, J., Avramakis, E. and Klar, A. (2021), "The untapped potential of ecosystems in health care", *Boston Consulting Group*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J. (2019), *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*.

Renner, S., Rosenthal, A. and Scarano, J. (1999), "Data Interoperability: Standardization or Mediation".

Shivakumar, S. (2022), "Securing Global Standards for Innovation and Growth", *Center for Strategic and International Studies (CSIS)*.

Sunyaev, A., Dehling, T., Strahringer, S., Da Xu, L., Heinig, M., Perscheid, M., Alt, R. and Rossi, M. (2023), "The Future of Enterprise Information Systems", *Business & Information Systems Engineering*, Vol. 65 No. 6, pp. 731–751.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D.W. and Middleton, B. (2005), "The value of health care information exchange and interoperability", *Health affairs (Project Hope)*, Suppl Web Exclusives, W5-10-W5-18.