Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250578

Implementing FAIR Principles in Health Terminology Servers: The HeTOP Case Study Hosting ODH-UT

Arriel BENIS[a,1], Julien GROSJEAN[b,c], Kevin EYER[b], Flavien DISSON[b], and Stéfan DARMONI[b,c]

[a] *Department of Digital Medical Technologies, Holon Institute of Technology, Israel*
[b] *Department of Digital Health, Rouen University Hospital, France*
[c] *Analyse Intégrée Multimodale en Santé (AIMS) lab, University of Rouen Normandy, France*

ORCiD ID: Arriel BENIS https://orcid.org/0000-0002-9125-8300, Julien GROSJEAN https://orcid.org/0000-0002-7446-644X, Flavien DISSON https://orcid.org/0000-0001-8793-4305, Stéfan DARMONI https://orcid.org/0000-0002-7162-318X

Abstract. The compliance of health terminology servers with the FAIR principles is crucial for establishing an interoperable digital health ecosystem that facilitates health data exchange. These servers must maintain high-quality standards while providing persistent, multilingual, and cross-lingual management of health terminologies. Implementing the FAIR principles is vital for creating consistent connections within One Digital Health, leading to better global healthcare outcomes and decision support systems by supporting international collaboration and facilitating interactions between all the actors involved in health data usage. Thus, this paper focuses on aligning the Health Terminology/Ontology Portal (HeTOP) with the FAIR principles, hosting the One Digital Health - Unified Terminology (ODH-UT).

Keywords. FAIR, Semantic interoperability, Controlled vocabulary, Terminology server, Languages, Digital Health, eHealth, Electronic Health Record, Clinical Datawarehouse, One Health, One Digital Health.

1. Introduction

Semantic interoperability is crucial for reusing healthcare data in Clinical Data Warehouses (CDWs). Utilizing terminologies or ontologies as a common data structure is key to achieving semantic interoperability for biomedical data. Health terminology servers (HTSs) manage ontologies, terminologies, and coding systems at a semantic level. They are essential for Electronic Health Records (EHR) and CDWs, enhancing the quality of data collected in EHR for various purposes such as research, hospital management, decision support systems, and coding optimization in CDWs [1,2].

Healthcare management organizations (HMOs) often use different software and terminologies to record health information, which can lead to discrepancies in electronic

[1] Corresponding Author: Arriel Benis, Department of Digital Medical Technologies, Holon Institute of Technology, Golomb St. 52, Holon, 5810201, Israel; E-mail: arrielb@hit.ac.il.

health records. HTS can help by standardizing an HMO's health-related terminologies and mapping them to national or international reference terminologies.

Accordingly, several HTSs have been created to store, search, and use multiple TOs simultaneously to allow mapping from one to the others. Among them are some internationally used, such as UMLS [3], the EBI Ontology Lookup Service [4], the National Center for Biomedical Ontology - BioPortal [5], and HeTOP [6]. HTSs also exist at the national level in several countries, such as the NHS (National Health Service) England Terminology Server, the Ontoserver in Australia, and SMT (a.k.a. Server Multi Terminologies) in France [7,8]. HeTOP is developed since 2007, at the Department of Digital Health of the Rouen University Hospital, Normandy, France. HeTOP is a cross-lingual health termino-ontologies server [9], comprising in October 2024, over 100 health terminologies in 55 languages. In contrast, other HTSs (e.g., BioPortal) comprise more than 1,000 ontologies, mainly in English, without integrating natively multilingual or crosslingual approaches.

This complexity must be standardized to be persistent over time. Therefore, the ISO 26162 standard comprises a model to manage terminology, knowledge, and content, particularly designing, implementing, and maintaining terminology management systems (or terminology servers). To put these recommendations into action, it is important to consider the FAIR (i.e., "Findability," "Accessibility," "Interoperability," and "Reusability") principles providing a framework for organizing research outputs so that they can be easily accessed, understood, exchanged, and reused digital resources, including datasets, codes, workflows, and research objects [10]. Few HTS developers and owners have analyzed and published their compliance with FAIR principles [11].

We aim to investigate how HTSs comply with FAIR principles and metrics in the context of One Digital Health (ODH) [12]. ODH uses digital technologies to create consistent links between education, environment monitoring, human and veterinary medicine, industry 4.0, and citizen engagement. Understanding HTSs' compliance with FAIR principles is crucial for establishing an interoperable digital health ecosystem capable of seamless, secure health data exchange and processing, considering the five ODH dimensions highlighted above [13]. Moreover, HeTOP hosts the ODH – Unified Terminology (ODH-UT) [14] as a part of the Medical Informatics and Digital Health Multilingual Ontology (MIMO) [15].

After introducing the HTSs and the FAIR principles above, we present our methodology for assessing compliance with the HTSs' properties and FAIR principles. Then, we present the example of the HeTOP cross-lingual terminology server in a synthetic and use case-based way.

2. Methods

To create effective terminology servers, it is crucial to ensure high-quality data that adheres to certain components such as availability, usability, reliability, relevance, and presentation quality. Additionally, data must align with various quality elements, including accessibility, timeliness, authorization, credibility, definition/documentation, metadata, accuracy, consistency, integrity, completeness, auditability, fitness, readability, and structure. These elements must also be in line with the FAIR principles.

The FAIR guiding principles emphasize the need to enhance computational systems' capabilities to manage digital resources' monitoring, collection, processing, and stewardship with minimal human intervention. This is especially important in the health and medicine field, where the "big data" and "cloud computing" revolutions have increased data volume, velocity, and variability.

3. Results

HTSs should inherently align with FAIR principles, being findable, accessible, interoperable, and reusable. However, each principle requires careful assessment. For instance, HeTOP, a cross-lingual terminology server, meets most FAIR criteria. It ensures identifiers crucial for human-machine interoperability, supporting the "Open Science" vision, and offers rich metadata, searchable resources, and more. Yet, HeTOP does not fully adhere to all FAIR principles. Notably, it lacks versioning for terminologies (typically for ICD10), providing only the latest version online. This strategic decision was made to optimize response times and ease the end-user experience, although older versions are stored offline for historical relevance in clinical data.

Table 1. FAIR principles applied to the HeTOP cross lingual terminology server

FAIR principles	HeTOP
F1: (Meta)data are assigned a globally unique and persistent identifier.	Each health concept integrated into HeTOP has its own Uniform Resource Identifier (URI), which includes one identifier for each terminology and a code for each language.
F2: Data are described with rich metadata (defined by R1 below).	(+) HeTOP manages data with rich metadata, including rich relations for concepts and mappings between different terminologies. The metadata comprises internal and external descriptions for imported controlled vocabularies and metathesaurus.
F3: (Meta)data clearly and explicitly include the identifier of the data they describe.	(+) In HeTOP, each piece of data has a global and unique identifier: HID (HeTOP ID). When available, the ID from the original data source is also included, such as CUI for UMLS concepts or IDs for reference terminologies like MeSH or NCIT.
F4: (Meta)data are registered or indexed in a searchable resource.	(+) HeTOP is a terminology server that indexes searchable resources of imported and house-made metadata sets. Its internal model requires analysis to integrate new terminologies or ontologies. The data is reusable and can be indexed on a platform for medical informatics and digital health academic teaching resources. Additionally, HeTOP's searchable resource is based on a semantic search engine previously designed to index health resources.
A1: (Meta)data are retrievable by their identifier using a standardized communications protocol.	(+) HeTOP metadata are retrievable using a standardized communications protocol (http, ftp) and XML-based standard data exchange formats (e.g., OWL, JSON).
A1.1: The protocol is open, free, and universally implementable.	(-) HeTOP hosts some terminologies under a commercial license. Thus, access to HeTOP is partially free, and for some parts of HeTOP, it is restricted to authorized/granted users.
A1.2: The protocol allows for an authentication and authorization procedure, where necessary.	(+) In some terminology, Access to HeTOP is restricted to authorized users. After registering and granting access, users are required to use an ID and a password.

A2: Metadata are accessible, even when the data are no longer available.	(-) HeTOP does not allow automated access to old and not up-to-date metadata. To do this, a special request is necessary to restore a specific backup in a dedicated area.
I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	(+) HeTOP metadata are accessible, sharable, and (re)usable regarding the use of well-known knowledge representation languages such as OWL, JSON, RDF, and CSV.
I2: (Meta)data use vocabularies that follow FAIR principles.	(+) HeTOP manages controlled vocabularies that are FAIR compliant *per se*, such as SNOMED and LOINC.
I3: (Meta)data include qualified references to other (meta)data.	(+) This is a fundamental principle of HeTOP and, by extension, of multi-terminologies servers. The HeTOP metadata include qualified references to other metadata.
R1: (Meta)data are richly described with a plurality of accurate and relevant attributes	(+) This is a fundamental principle of any multi-terminologies server.
R1.1: (Meta)data are released with a clear and accessible data usage license.	(+) HeTOP is freely accessible for freely accessible and non-licensed terminologies. The final user must own a commercial license for license-based terminologies, whereas HeTOP owns a specific research license. A login is necessary to access HeTOP for these last ones. The released (meta)data include clear and accessible data usage information for each license adaptation.
R1.2: (Meta)data are associated with detailed provenance.	Each metadata is clearly stated in HeTOP (e.g., there is a clear difference between metadata from terminologies and HeTOP or its partners (e.g., the French National Digital Health Agency).
R1.3: (Meta)data meet domain-relevant community standards.	(+) The metadata meets the medical informatics and digital health standards by linking 1-N terminologies and 1-N languages (e.g., English, French, Spanish).

4. Discussion and Conclusions

HTSs are crucial in healthcare for exchanging health information. The FAIR principles emphasize that health terminologies should be findable, accessible, interoperable, and reusable. HTSs are essential tools for managing and providing access to terminological resources. HTSs centralize access to terminologies, making it easier for all the actors involved in health data usage to discover needed information through advanced search functions and comprehensive metadata. High-quality metadata and integration with wider search systems are essential for enhancing findability. HTSs improve accessibility by offering various access methods and ensuring interoperability through standardized formats and protocols. However, without adequate access controls or support for diverse data formats, the accessibility of sensitive resources could be compromised. Compatibility issues and inconsistencies in interpreting terms across domains can also reduce the benefits of TSs. In the future, we can expect a strong focus on standardization and harmonization to improve interoperability among HTSs. Semantic technologies like Linked Data and artificial intelligence processes will enhance HTS's FAIRness. The user interfaces will also become more intuitive, personalized, and integrated with knowledge graphs to provide intelligent and contextualized terminological services. FAIR metrics will emerge to evaluate adherence to these principles. Adhering to these principles is also essential for internationalizing new healthcare concepts, facilitating harmonized cross-border collaboration, and promoting global interoperability in using and exchanging health data. Applying FAIR principles improves HTSs usability and interoperability by promoting vocabularies (e.g., ODH-UT [14]) dealing with each one of the five ODH dimensions: education, environmental monitoring, human and veterinary medicine, industry 4.0, and citizen engagement in real-world applications [16].

References

[1] Grosjean J, Soualmia LF, Bouarech K, Jonquet C, Darmoni SJ. An approach to compare bio-ontologies portals. Stud Health Technol Inform. 2014;205:1008–1012.

[2] Wiedekopf J, Ulrich H, Drenkhahn C, Kock-Schoppenhauer A-K, Ingenerf J. TermiCron - Bridging the Gap Between FHIR Terminology Servers and Metadata Repositories. Stud Health Technol Inform. 2022;290:71–75. doi:10.3233/SHTI220034

[3] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32:D267–70. doi:10.1093/nar/gkh061

[4] Perez-Riverol Y, Ternent T, Koch M, Barsnes H, Vrousgou O, Jupp S, Vizcaíno JA. OLS Client and OLS Dialog: Open Source Tools to Annotate Public Omics Datasets. Proteomics. 2017;17:1700244.

[5] Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res. 2011;39:W541–5.

[6] Darmoni SJ, Pereira S, Sakji S, Merabti T, Prieur É, Joubert M, Thirion B. Multiple Terminologies in a Health Portal: Automatic Indexing and Information Retrieval. In: Combi C, Shahar Y, Abu-Hanna A, editors. Artificial Intelligence in Medicine. AIME 2009. Lecture Notes in Computer Science. 2009;5651:255–9. doi:10.1007/978-3-642-02976-9_37

[7] National Health Service. The NHS England Terminology Server [Internet]. NHS Digital; 2023. Available from: https://digital.nhs.uk/services/terminology-servers#the-nhs-england-terminology-server

[8] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. J Biomed Semant. 2018;9:24. doi:10.1186/s13326-018-0191-z

[9] CISMeF. HeTOP, the Health Terminology/Ontology Portal [Internet]. 2024. Available from: https://www.hetop.eu/hetop/en/

[10] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. doi:10.1038/sdata.2016.18

[11] Krauss P, Touré V, Gnodtke K, Crameri K, Österle S. DCC Terminology Service—An Automated CI/CD Pipeline for Converting Clinical and Biomedical Terminologies in Graph Format for the Swiss Personalized Health Network. Appl Sci. 2021;11:11311. doi:10.3390/app112311311

[12] Benis A, Tamburis O, Chronaki C, Moen A. One Digital Health: A Unified Framework for Future Health Ecosystems. J Med Internet Res. 2021;23:e22189. doi:10.2196/22189

[13] Tamburis O, Benis A. One Digital Health for more FAIRness. Methods Inf Med. 2022;(Ahead of print):a-1938-0533.

[14] Bauberg H, Nehama D, Blinder C, Davidman R, Cohen E, Tamburis O, et al. Initiating a One Digital Health Unified Terminology (ODH-UT) to Facilitate Community Expansion. Stud Health Technol Inform. 2024;316:1569–73. doi:10.3233/SHTI240719

[15] Benis A, Grosjean J, Billey K, Montanha G, Dornauer V, Crișan-Vida M, et al. Medical informatics and digital health multilingual ontology (MIMO): A tool to improve international collaborations. Int J Med Inform. 2022;167:104860. doi:10.1016/j.ijmedinf.2022.104860

[16] Benis A, Grosjean J, Disson F, Crisan-Vida M, Weber P, Stoicu-Tivadar L, et al. Empowering healthcare education: A multilingual ontology for medical informatics and digital health (MIMO) integrated to artificial intelligence powered training in smart hospitals. Health Inform J. 2024;30:14604582241287010. doi:10.1177/14604582241287010.