

Improving Health Information Exchange by Combining Interoperable Resources and Digital Collection Integration Tools

Félix Buendía

*Universitat Politècnica de València
Valencia, Spain*

fbuendia@disca.upv.es

Joaquín Gayoso-Cabada

*Universidad Politécnica de Madrid
Madrid, Spain*

j.gayoso@upm.es

José-Luis Sierra

*Universidad Complutense de Madrid
Madrid, Spain*

jlsierra@ucm.es

Abstract

Health information exchange plays a key role in any kind of healthcare system. An important barrier in this exchange is the lack of mechanisms and tools able to fully integrate information models that underlie existing healthcare systems. The current paper presents an approach to gathering multiple sources of clinical data by taking advantage of the HL7 FHIR (Fast Healthcare Interoperability Resources) format. The use of this standard format together with the *Clavy* tool enables a powerful approach to managing digital health collections that can easily be exchanged in different healthcare domains. In this sense, several content items for healthcare training, based on e-learning standards, have been generated from a clinical dataset that combines FHIR resources and DICOM images. Such a generation process shows the capability of the approach presented to coping with the exchange of health information based on multiple multimedia formats and controlled medical vocabularies.

Keywords: information exchange, healthcare systems, interoperable resources, data integration tools.

1. Introduction

Health Information Exchange (HIE) plays a key role in any kind of healthcare system. Electronic HIE allows “doctors, nurses, pharmacists, other health care providers and patients to appropriately access and securely share a patient’s vital medical information electronically” [16]. This exchange is usually supported by interoperability mechanisms providing “timely and seamless portability of health information” [7]. Such a need for exchanging or using data in a coordinated manner is particularly important in Medical Information Systems (MIS), which require the communication and interchange of clinical information coming from different healthcare platforms [6]. Multiple health data sources can be collected from patient registries, medical records or clinical reports. Most of them have to be exchanged among healthcare professionals in order to support multidisciplinary clinical routines. EHRs (Electronic Health Records) are typical examples of systems addressed to storing patient information in which interoperable capabilities are still a challenge [20]. Clinical registries also benefit from using standardized formats that aid the interoperable exchange of health information [19]. Thus, interoperability has become an essential factor in every ambit of digital medicine [12]. In this sense, several levels of interoperability can be defined that include structural, syntactical or more advanced semantic issues. Semantic levels refer to “common underlying models and codification of the data including the use of data elements with standardized definitions” [7]. These

common models have been extensively reviewed during the last years to evidence the increasing interest in offering a uniform health data storage interface [24]. Moreover, the spread of standardized vocabularies to represent and codify healthcare data [22] has presented a remarkable challenge to integrating the huge volume of health information that can be exchanged and shared. Therefore, we need to find mechanisms and tools able to fully integrate the information models and semantic notations that underlie existing healthcare systems. In this sense, HL7¹ has become one of the organizations most actively pursuing semantic interoperability in the healthcare domain. The current paper presents an approach to integrating multiple sources of healthcare data by taking advantage of the HL7 FHIR (Fast Healthcare Interoperability Resources) format. There are several publications that have proposed the application of FHIR resources in healthcare data integration. The OHDSI (Observational Health Data Sciences and Informatics) Common Data Model has been harmonized with FHIR to facilitate electronic health records-based data exchange [10]. An FHIR-based system has been used to integrate structured and unstructured EHR Data [9]. SMITH (Smart Medical Information Technology for Healthcare) implements a federated approach to share health information among a set of partners [25]. This approach is based on HL7 communication standards such as CDA (Clinical Document Architecture) and FHIR. This latter one has also been used as a meta-model to integrate Common Data Models and transform them into FHIR formats [17]. Summing up, FHIR represents a remarkable option for exchanging well-structured information within an interoperable health context. However, the main challenge still lies in the integration of unstructured data in such a context.

The approach proposed takes advantage of the *Clavy* [5] tool, which provides users with mechanisms to obtain and integrate information from different data sources using plug-in components which can be adapted to different data schemata (e.g. organized as FHIR resources) or unstructured multimedia formats such as annotated text reports or images. The *Clavy* tool has proven its ability to retrieve and organize these multiple types of information sources around medical collections [2]. In the current context, this tool has been used to extract information from FHIR repositories together with files associated to related imaging studies and then, to transform the resources retrieved into interoperable e-learning content which can be stored in well-known e-learning platforms with instructional or training purposes. The main contribution of the *Clavy* tool consists of providing a versatile reconfiguration and mapping process between the FHIR resources and data schemata that enables linking with other multimedia formats and their exportation under standard specifications. Moreover, *Clavy* is able to connect with existing controlled vocabularies and healthcare coding notations.

The remainder of the work is organized as follows. Section 2 refers to some related work about the integration of either structured or unstructured data sources in a healthcare context. Section 3 describes the approach presented organized as a set of steps addressed to ingesting, transforming and integrating these different healthcare data sources. Section 3 shows a case study used to check the suitability of this approach and how it enables the generation of new resources, which can be exchanged with a health instructional purpose. Finally, some conclusions and future work are drawn in section 4.

2. Related works

Promoting a better integration of health information systems has been increasingly pursued during recent years [15]. The exchange of information in these systems can be considered crucial since it enables enhanced decision-making support or contributes to improved patient care in clinical settings [14]. As commented before, one of the biggest challenges in this integration process consists of merging health data items coming from multiple sources, whether structured or unstructured. In the past, the search for technological platforms to integrate several sources of healthcare information promoted the development of PHR (Personal Healthcare Records) systems such as Google Health or Microsoft HealthVault [23] However, the discontinuity of these systems highlighted the complexity

¹ <https://www.hl7.org/implement/standards/>

of exchanging this kind of healthcare data sources. PHRs have been reviewed in the context of medical and health record integration [8] and their application can be linked to the use of smartphones and data collected from wearable devices [4]. The scope of the current work is more focused on overcoming difficulties when trying to integrate structured and unstructured data in the healthcare domain. Scheurwegs et al [21] researched how to link both types of data sources when dealing with EHRs and the assignment of controlled clinical codes in these records. They compiled a dataset from 14 medical specialities, which was composed of unstructured data sources such as surgery text reports or discharge letters and structured data such as lab results, oncological pathologies or demographic data. Linking data elements, structured and unstructured, to different procedure or diagnosis codes required a mapping process among these data items, similar to the approach proposed in the current work. Also related is the hybrid solution addressed to extracting structured medical information from unstructured data in medical records [13]. Its authors developed a method to manually curate structured data elements from unstructured clinical data such as medical images or patient reports. More recently, a knowledge-guided integration of structured and unstructured data to facilitate health decision making was presented [18]. This knowledge integration was based on a medical domain ontology that helped experts to select the most relevant attributes of structured and unstructured information.

3. Approach proposal with *Clavy*

The approach proposed is based on the idea of retrieving interoperable data sources using standard formats and generating different kinds of outcomes that can be exchanged and adapted to certain purposes. Figure 1 shows an overview of the approach architecture that displays data inputs for diverse types of healthcare information, which can be stored, for example, on HL7 FHIR resources, DICOM² images or text documents. This data input can be processed by means of the *Clavy* tool to generate several outputs such as patient-oriented reports for personalized user care, pathology-oriented cases for clinical treatment conditions or diagnostic studies addressed to medical research.

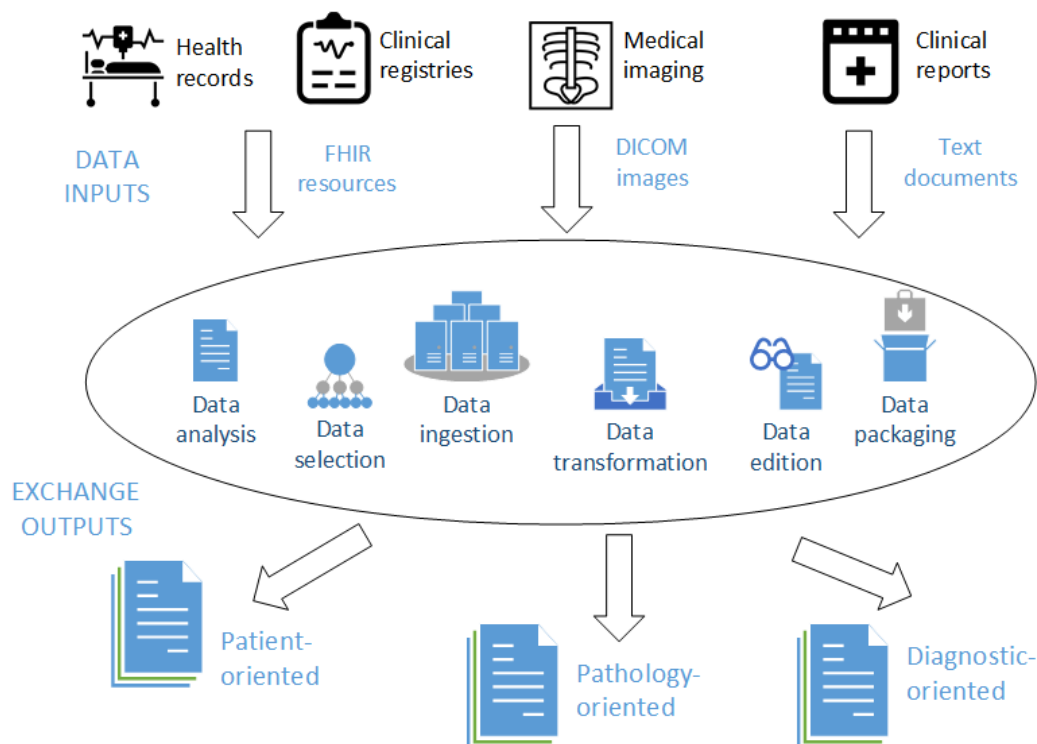


Figure 1. Architecture overview of the proposed approach.

² <https://www.dicomstandard.org/>

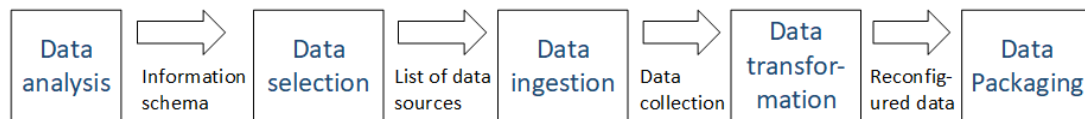


Figure 2. Workflow template for the approach.

Such a global architecture can be used to implement a set of different workflows according to the available healthcare data sources and the final target to be generated as output. Figure 2 shows an example of workflow template that displays a sequence of steps starting with the analysis of data inputs and their selection in order to fulfil a given purpose or need. Of these stages, data ingestion, data transformation and data packaging can be supported by the *Clavy* tool. Next subsections address the details of the workflow. Subsection 3.1 describes the data analysis and data selection stages. Subsection 3.2 introduces the *Clavy* tool, which is used to support data ingestion and data transformation (subsection 3.3), as well as data edition and packaging (subsection 3.4).

3.1. Analysis and selection of healthcare data sources

In the first stage, an analysis of the data sources available and their further selection are performed. As mentioned in the previous section, several perspectives can be analyzed in terms of healthcare information exchange, ranging from the patient point of view to a condition or a diagnostic arrangement. These sources come from different data types, structured or not, such as electronic health records, medical images or text clinical reports. Structured data can be represented by database systems, spreadsheets, or XML files while text documents or image files are examples of unstructured formats. In this approach, structured data sources are based on FHIR resources that offer a homogenous interface to access this kind of interoperable data. In the case of unstructured data, there is a wide range of multimedia formats to represent medical images or text reports.

Therefore, experts usually face multiple challenges when dealing with the wide range of existing data sources and they have to select the most suitable sources in order to fit a specific purpose, either in a clinical context or oriented towards training healthcare practitioners. The current approach attempts to provide the right mechanisms and tools that will allow these experts to identify the best data types for a given purpose. In this sense, the *Clavy* tool offers a powerful platform to manage all the potential data sources that can be deployed in a healthcare context as described below.

3.2. *Clavy*

As described in [5], *Clavy* is a reconfigurable data management tool that is able to get pieces of content that are usually poorly structured in external sources and arrange them in highly specialized digital collections. For this purpose:

- *Clavy* enables the aggregation of content from multiple external sources using *import plug-ins*, which makes it possible to import data from standard platforms and formats (XML, JSON, relational or NoSQL databases, etc.).
- *Clavy* also supports the addition of different *transformation plug-ins*, which makes it possible to enrich data with domain-specific semantics.
- In addition, *Clavy* supports powerful edition capabilities, allowing domain experts (physicians, in our case) to both edit and enrich individual information items and also to reorganize the overall structure of the digital collection. These capabilities in turn can be specialized using suitable *edition plug-ins*.
- Finally, *Clavy* can be extended with suitable *export plug-ins*, which can be used to convert the collections transformed and edited to different standard formats.

In order to harmonize all these elements, *Clavy* provides a unified digital collection model inspired in generalized markup languages like SGML and XML [3]. Indeed, a digital collection in *Clavy* integrates a set of *documents* conforming to a set of *document schemata*. Documents in *Clavy* allow for aggregating and organizing resources in terms of *element-value* pairs. In addition, a document schema describes the hierarchical organization of the elements in the documents that follow this schema. Using *Clavy*'s edition facilities, it is possible to *reconfigure* schemata by adding new element types, by

renaming and erasing existing elements, and by reorganizing the hierarchical relationships among all the elements in the schema. In consequence, these changes in the schemata are automatically mirrored in the collection documents. This feature lets domain experts reorganize the whole collection by solely editing the collection's document schemata. In addition, they can use the document editing facilities of the tool to make more fine-grained adjustments in the content, as well as to enrich this content with new information items (e.g., text and image synchronized annotations, as described in [1]).

3.3. Data ingestion and transformation

The list of data sources chosen in the selection stage can be ingested by means of *Clavy* import plug-ins. These plug-in components are based on a generic importation engine that can be adapted according to the features of the data types to be extracted. In addition, *Clavy* transformation plug-ins enable the automatic transformation of the information ingested to add *semantic layers* to this information. Adding semantic layers means altering the internal structure of those documents addressed to storing the ingested data sources by unifying them in a common organization or enabling *semantic link* elements to enhance the navigation through these documents and their connection with multimedia contents.

3.4. Data edition and final exportation

The last steps in the approach proposed consist of editing the documents that store the healthcare data sources retrieved and exporting them in a format that can be easily exchanged in new environments according to the needs defined. The edition process takes advantage of the *Clavy* tool to add new element types that make some clinical requirements clear or to erase those elements that can be considered irrelevant in this context. Finally, *Clavy* provides mechanisms to export the results of this data edition in the form of standard interoperable packages, as for example, e-learning packages conforming to the IMS-CP³ specification, which can be used to show healthcare training procedures from an instructional point of view [2].

4. Case study

An application of the approach proposed is addressed in this section in form of a case study that shows its feasibility when dealing with heterogeneous data sources, which can be exchanged and integrated into a context of healthcare interoperability. This case study is based on a well-known example of a dataset used by the Society of Imaging Informatics in Medicine (SIIM) in a collaborative computing experience celebrated every year since 2014 [11]. The SIIM dataset is composed of clinical data, imaging files, and radiology reports for five patient profiles with different features and conditions. It is publically available and represents a remarkable reference in the use of interoperable formats such as FHIR resources or DICOM files together with LOINC-RadLex⁴ and SNOMED-CT⁵ codes.

The case study itself shows how the available information sources are analyzed, how the relevant information is selected and the target organization is envisioned, and how the *Clavy* tool is used to ingest the multiple data sources that are part of the SIIM dataset and to transform them to produce learning content compliant with the IMS CP specification. Figure 3 outlines the workflow stages followed during the whole process. Next subsections describe these stages.

4.1. Data Analysis and selection

In the first stage of the approach, the expert plays a crucial role in the analysis of the data sources available. For the current case study, a *patient-oriented* perspective has been chosen in order to verify how the approach proposed is able to represent and manage an integral view of patient information. In this context, given the set of resources composing the SIIM dataset, their arrangement is focused on getting a unified perspective of the

³ <http://www.imsglobal.org/content/packaging/index.html>

⁴ <https://loinc.org/committee/loinc-radlex/>

⁵ <https://www.snomed.org/>

patient’s clinical conditions, the reports obtained from their diagnosis or the image studies supporting these reports. Consequently, as made apparent in Figure 4, we focus our attention on the following FHIR data entities to be selected:

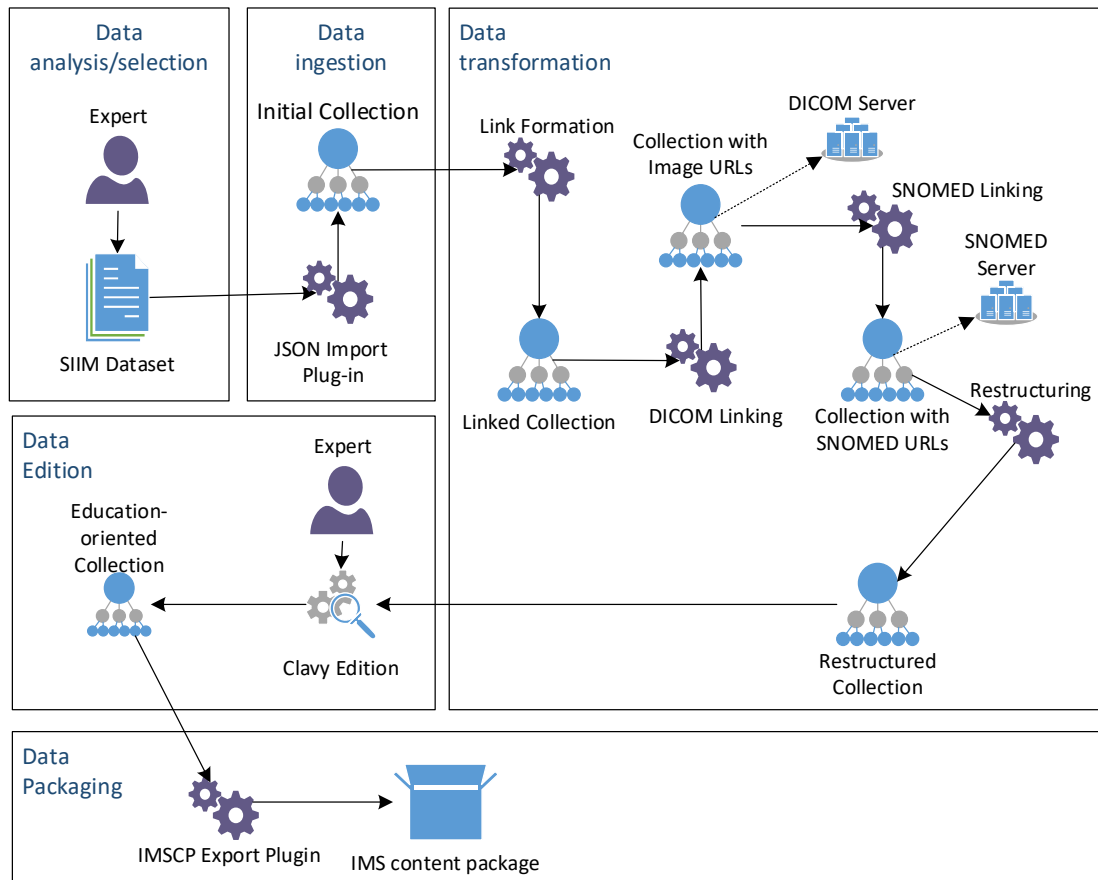


Figure 3. Workflow stages from FHIR resources to IMS CP-compliant contents.

- *Patient*: provides relevant demographic data such as birth date, gender, address, contact information or marital status.
- *Condition*: makes it possible to represent a clinical pathology, a health dimension or any other issue relevant from a healthcare point of view.
- *ImagingStudy*: is able to store the set of DICOM images, organized in series and instances, concerning a particular modality such as X-Ray, MR, CT or ultrasound.
- *DiagnosticReport*: addressed to aggregating those findings or interpretation of diagnostic tests concerning a given subject or a group of them.

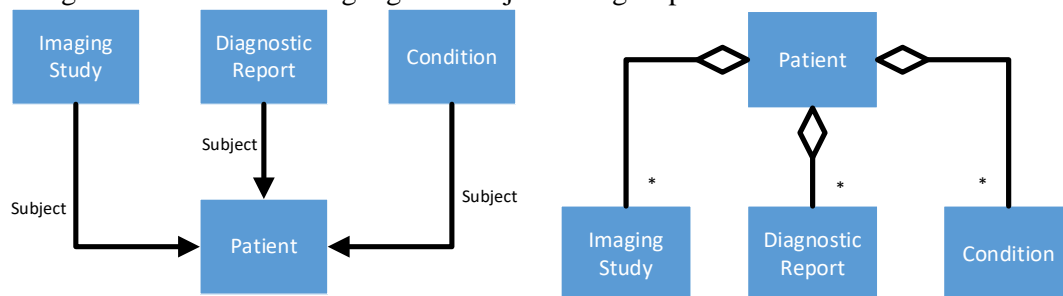


Figure 4. Selected FHIR entities and their organization.

Originally, these FHIR entities are linked through a unidirectional connection as displayed on the left image of Figure 4. This feature makes it apparent that the current FHIR specification fails to provide direct links from a *Patient* instance to resources such as *Condition*, *DiagnosticReport* or *ImagingStudy* entities, including access to related DICOM files. The alternative to this setting is displayed on the right image of Figure 4, which shows a more suitable organization, according to the different conditions, imaging studies or diagnostic reports, which can be aggregated as items of a patient-oriented

transformation. The effect of this transformation is illustrated in Figure 6, where two examples of FHIR resources are displayed in the left area, the first one referring to a *Patient* entity and the second one representing a related *DiagnosticReport*. The right area on Figure 6 displays the transformed documents, which include an explicit semantic link element between the two resources.

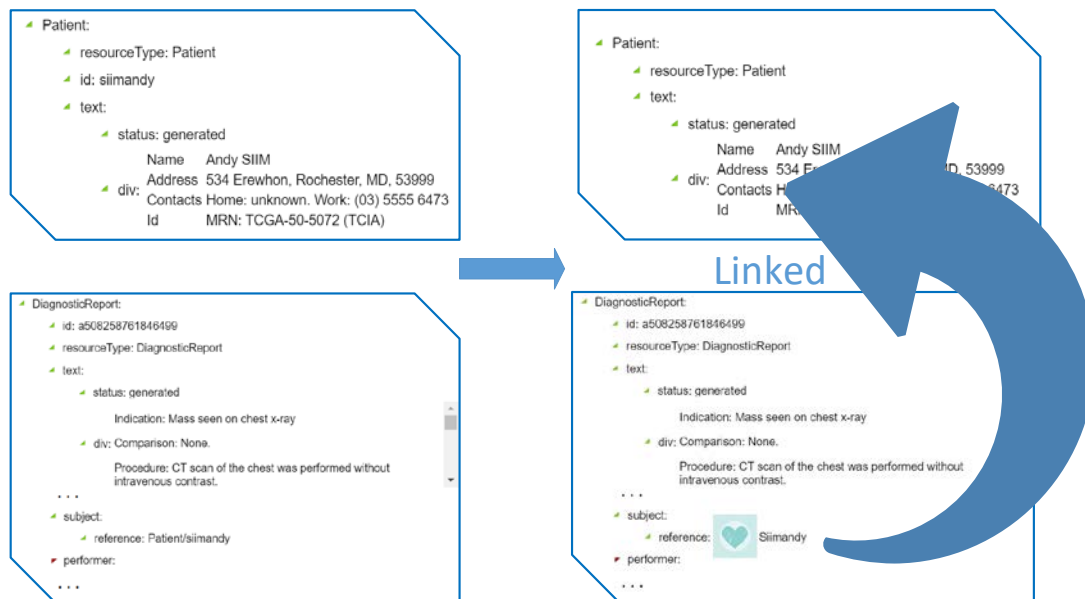


Figure 6. Clavy document transformation of FHIR resources.

Likewise, specific semantic link elements can be added to connect imaging studies (based on FHIR notations) with DICOM images. For this purpose, a DICOM *linking* transformation is used. For instance, Figure 7 shows the effect of these transformations on the original document (left area of the figure), which represents an example of *Entry* within an *ImagingStudy* resource, to produce the final document displaying the link to the final image accessible through a DICOM viewer⁶ in the right area. A similar process is followed to link *condition* documents with SNOMED terms, via a SNOMED *linking* transformation.

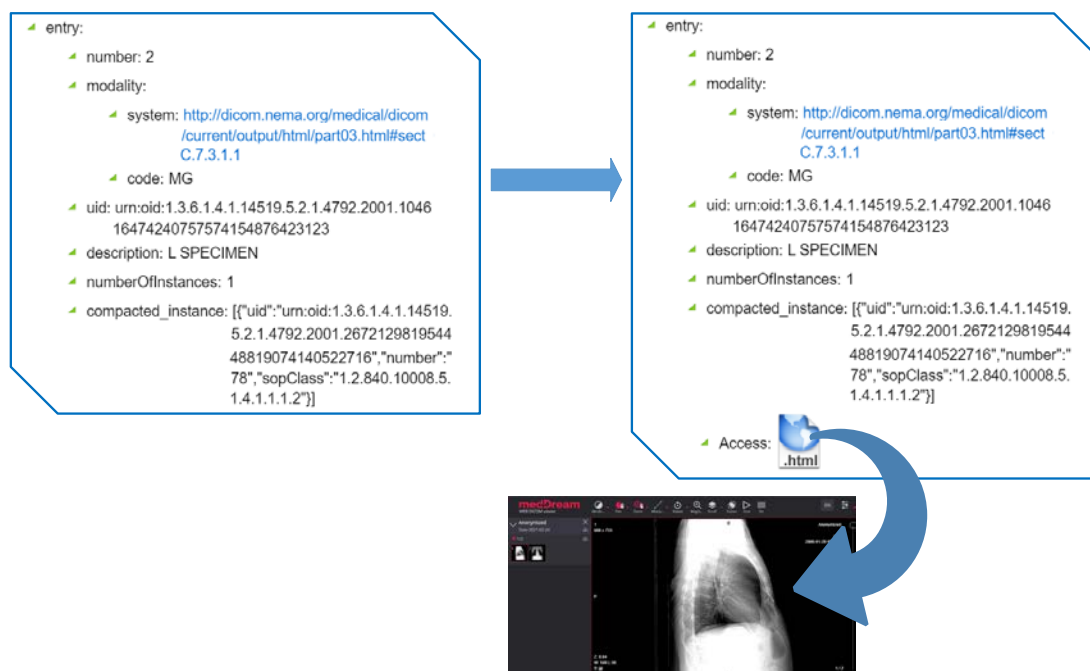


Figure 7. Clavy document transformation of *ImagingStudy* instances.

⁶ <https://www.softneta.com/products/meddream-dicom-viewer/>

Finally, the patient-oriented perspective required in this case study is achieved through a *restructuring* transformation, which reorganizes the collection to aggregate all the information related to each patient into a single document (see Figure 8).

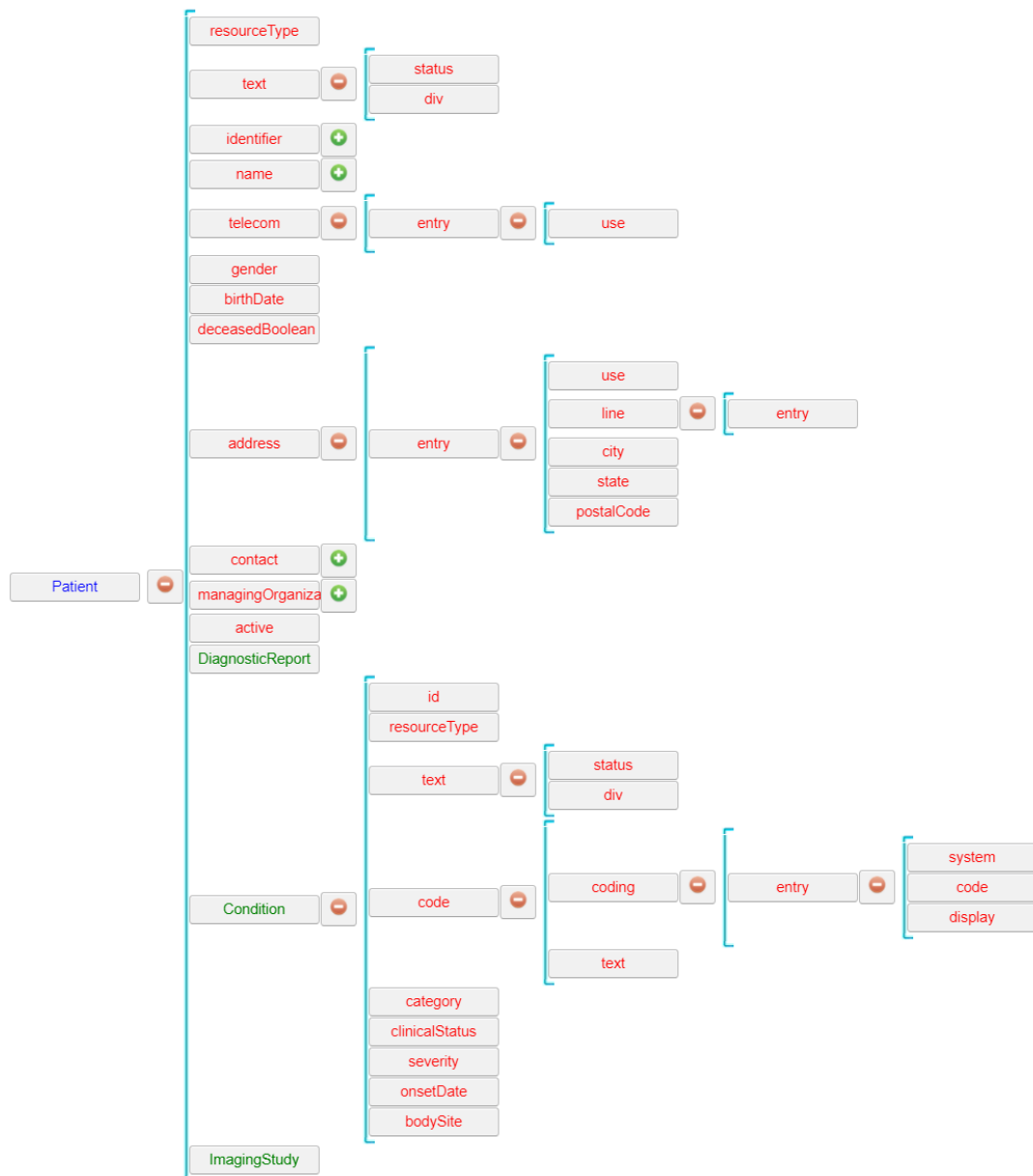


Figure 8. Excerpt of the single document schema for the collection that results of the transformation pipeline.

4.4. Data Edition

For certain purposes, the collection resulting from the transformation pipeline described in the previous section can exhibit many pieces of information that are irrelevant or unnecessarily detailed. This is, for example, the situation in our case study, where our ultimate goal is to produce educational content packages. To that end, *Clavy's* editing functionalities can be used to simplify the schema, eliminating all elements that are deemed irrelevant, and reorganizing the structure to better suit our specific objectives. Figure 9 shows the collection schema after the simplification and restructuring process. As can be appreciated, much of the information in FHIR records considered redundant (e.g., system and code elements for SNOMED terms in *conditions*, as it is now possible to access the complete entry in the SMOMED server) or irrelevant (e.g., much of the information related to DICOM series, as these can now be accessed directly in the DICOM server) to our specific educational goal has been removed. In addition, further reconfigurations have been applied to the schema.

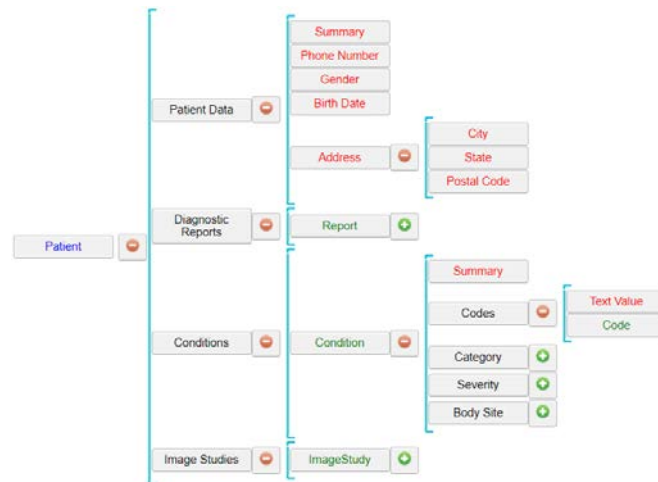


Figure 9. Excerpt of the edited schema.

4.5. Data Packaging

As a result of the edition process, interoperable medical information from different sources (FHIR resources, DICOM images or SNOMED codes in the current case study) is integrated into a digital collection with a specific educational purpose. Ultimately, it can be demonstrated that is possible to export this information in an equally interoperable format.

In our scenario, each patient document represents a significant clinical case, which makes it possible to generate e-learning packages containing sets of patient-oriented clinical cases. To do so, it is sufficient to use the *Clavy's IMS CP export plug-in*, which makes it possible to generate IMS content packages from a set of objects in a collection. The IMS content packages exported can, in turn, be incorporated into any learning management environment that supports the IMS CP specification. For example, Figure 10 shows an excerpt from a *Moodle*⁷ course showing IMS content extracted from our *Clavy* collection.

Figure 10. Screenshot of the IMS CP contents deployed in a Moodle LMS.

5. Conclusions

The current work has presented an approach that is able to deal with the integration and

⁷ <https://moodle.org/>

exchange of a wide variety of healthcare data sources. Such an approach has been organized in a set of stages, which allow the ingestion, transformation, edition or curation of these data sources to finally export them to new interoperable formats. In this case, a publicly available dataset composed of patient data, imaging files, and radiology reports has been used to show the power and versatility of the proposed approach. This dataset combines FHIR resources representing clinical text information together with DICOM images and SNOMED codes. The approach application enables the generation of an IMS-CP compliant package that integrates all these data items from a patient-oriented perspective. Such an IMS-CP package can be exchanged among several instructional platforms such as *Moodle* or *Cloud Scorm*⁸ and shows the potential of the approach to link and integrate multiple types of health data sources.

Further works plan to test the transformation and exportation capabilities of the proposed approach in new clinical scenarios gathering other sources of healthcare data for educational purposes. For example, Personal Health Records (PHR) can provide specific data sources with a closer patient perspective in a context of personalized medical attention. Moreover, these future works will explore the possibility of using those standard educational packages in e-learning platforms, helping to spread their knowledge among different healthcare communities.

Acknowledgements

This work was funded by Research Project TIN2017-88092-R.

References

1. Buendía, F., Gayoso-Cabada, J., Sierra, J. L.: An Annotation Approach for Radiology Reports Linking Clinical Text and Medical Images with Instructional Purposes. In: Proceedings of Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality Salamanca, Spain, pp. 510-517 (2020)
2. Buendía, F., Gayoso-Cabada, J., Sierra, J.-L.: Generation of Standardized E-Learning Content from Digital Medical Collections. *Journal of Medical Systems*, 43, 188 (2019)
3. Coombs, J.H., Renear, A.H., DeRose, S.J.: Markup systems and the future of scholarly text processing. *Communications of the ACM* 30(11): 933–947 (1987)
4. Dameff, C., Clay, B., Longhurst, C. A.: Personal health records: more promising in the smartphone era? *Jama*, 321(4), 339-340(2019)
5. Gayoso-Cabada, J., Rodríguez-Cerezo, D., Sierra, J.-L. Browsing Digital Collections with Reconfigurable Faceted Thesauri. In: Proceedings of International Conference on Information Systems Development (ISD), pp. 378-389 Katowice, Poland (2016)
6. Hammami, R., Bellaaj, H., & Kacem, A. H.: Interoperability for medical information systems: an overview. *Health and Technology*, 4(3), 261-272 (2014)
7. Healthcare Information and Management Systems Society (HIMSS): Interoperability in Healthcare. <https://www.himss.org/resources/interoperability-healthcare> (online access on March 18th, 2021)
8. Heart, T., Ben-Assuli, O., Shabtai, I.: A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. *Health Policy and Technology*, 6(1), 20-25 (2017)
9. Hong, N., Wen, A., Shen, F., Sohn, S., Liu, S., Liu, H., Jiang, G.: Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. In: Proceedings of AMIA Summits on Translational Science, 74 (2018)
10. Jiang, G., Kiefer, R. C., Sharma, D. K., Prud'hommeaux, E., Solbrig, H. R.: A consensus-based approach for harmonizing the OHDSI common data model with HL7 FHIR. *Studies in health technology and informatics*, 245, 887 (2017)
11. Kohli, M., Morrison, J. J., Wawira, J., Morgan, M. B., Hostetter, J., Genereaux, B., Langer, S. G.: Creation and curation of the Society of imaging informatics in medicine Hackathon dataset. *Journal of digital imaging*, 31(1), 9-12 (2018).

⁸ <https://cloud.scorm.com/>

12. Lehne, M., Sass, J., Essenwanger, A., Schepers, J., Thun, S.: Why digital medicine depends on interoperability. *NPJ digital medicine*, 2(1), 1-5 (2019)
13. Luo, L., Li, L., Hu, J., Wang, X., Hou, B., Zhang, T., Zhao, L. P.: A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system. *BMC medical informatics and decision making*, 16(1), 1-14 (2016)
14. McCarthy, D. B., Propp, K., Cohen, A., Sabharwal, R., Schachter, A. A., Rein, A. L.: Learning from health information exchange technical architecture and implementation in seven beacon communities. *eGEMs*, 2(1), 6 (2014)
15. Michelsen K, Brand H, Achterberg P W, Wilkinson J R.: Promoting better integration of health information systems: best practices and challenges. Copenhagen: WHO Regional Office for Europe; Health Evidence Network (HEN) synthesis report (2015)
16. National Alliance for Health Information Technology: What Is Interoperability? <https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie> (online access on March 18th, 2021)
17. Pfaff, E. R., Champion, J., Bradford, R. L., Clark, M., Xu, H., Fecho, K., Ahalt, S.: Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR medical informatics*, 7(4), e15199 (2019)
18. Radaoui, M., Ben Abdallah Ben Lamine, S., Baazaoui Zghal, H., Ghedira Guegan, C., Kabachi, N.: Knowledge guided integration of structured and unstructured data in health decision process. In: Proceedings of 28th International Conference On Information Systems Development, 28, Toulon, France (2019)
19. Rastegar-Mojarad, M., Sohn, S., Wang, L., Shen, F., Bleeker, T. C., Cliby, W. A., Liu, H.: Need of informatics in designing interoperable clinical registries. *International journal of medical informatics*, 108, 78-84 (2017)
20. Reisman, M.: EHRs: the challenge of making electronic data usable and interoperable. *Pharmacy and Therapeutics*, 42(9), 572 (2017)
21. Scheurwegs, E., Luyckx, K., Luyten, L., Daelemans, W., Van den Bulcke, T.: Data integration of structured and unstructured sources for assigning clinical codes to patient stays. *Journal of the American Medical Informatics Association*, 23(1), 11-19 (2016)
22. Schulz S., Stegwee R., Chronaki C. Standards in Healthcare Data. In: Kubben P., Dumontier M., Dekker A. (eds) *Fundamentals of Clinical Data Science*. Springer, Cham (2019)
23. Sunyaev, A.: Evaluation of Microsoft HealthVault and Google Health personal health records. *Health and Technology*, 3(1), 3-10 (2013)
24. Weeks, J., Pardee, R.: Learning to Share Health Care Data: A Brief Timeline of influential common data models and distributed health data networks in US health care research. *eGEMs*, 7(1), 4 (2019)
25. Winter, A., Stäubert, S., Ammon, D., Aiche, S., Beyan, O., Bischoff, V., Löffler, M.: Smart medical information technology for healthcare (SMITH): data integration based on interoperability standards. *Methods of information in medicine*, 57, e92 (2018)