Journal of Medical Systems (2019) 43: 188
https://doi.org/10.1007/s10916-019-1330-5

EDUCATION & TRAINING

Generation of Standardized E-Learning Content from Digital Medical Collections

Félix Buendía [1] · Joaquín Gayoso-Cabada [2] · José-Luis Sierra [2]

Received: 9 January 2019 / Accepted: 8 May 2019 / Published online: 18 May 2019

Abstract
In this paper, we describe a new approach to generating standardized e-Learning content from existing medical collections. The core of this approach is a tool called *Clavy*, which makes it possible to retrieve information items from medical collections, to transform these items into meaningful learning units, and to export them in the form of standardized e-Learning packages. In addition to describing the approach, we assess its feasibility by applying it to the generation of IMS Content Packages from *MedPix*, an online database of medical cases in the domain of radiology.

Keywords Medical knowledge · Digital collections · Information management tools · Data repositories · Instructional resources · E-learning platforms

Introduction

In recent years, there has been a proliferation of online digital medical collections that contribute to organizing medical knowledge in many different areas. To wit, initiatives such as TCIA [1] or BioLINCC [2] represent some examples of public sharing repositories that are the basis for health-care information collections. Although they are mainly focused on research goals, many of these collections have undoubted educational value, as evidenced by experiences on the use of data repositories for medical student education [3], recommendations oriented towards building them [4], or actual collections like *MedPix* [5]. Therefore, to take full advantage of this educational value, it would be convenient to have systematic methods that allow the integration of these materials in standardized e-Learning scenarios [6]. In this paper, we address these concerns. The main contributions of this paper to the medical training area are:

- To formulate a non-invasive method that makes it possible to produce standardized e-learning content packages from existing medical collections that lack support for generating such standardized material.
- To support the method proposed by means of a tool called *Clavy* [7], which facilitates the importation of data and content from existing medical collections, the transformation and reorganization of all this information to meet specific instructional criteria, and the exportation of the resulting structures as standardized content packages.

In order to assess the feasibility of the approach, we describe its application to *MedPix*, which provides a huge amount of medical cases in the radiology field. Thus, we show how, by using *Clavy*, it is possible to generate IMS Content Packages [8] from these *MedPix* medical cases, which can be deployed in widely-used e-Learning platforms. Preliminary stages of this work are described in [9, 10].

The generation process model

As we have already mentioned, there are many medical collections endowed with high educational value, but

This article is part of the Topical Collection on *Education & Training*

✉ José-Luis Sierra
jlsierra@ucm.es

Félix Buendía
fbuendia@disca.upv.es

Joaquín Gayoso-Cabada
jgayoso@ucm.es

[1] Escuela Técnica Superior de Ingeniería Informática, Universitat Politècnica de València, Valencia, Spain

[2] Facultad de Informática, Universidad Complutense de Madrid, Madrid, Spain

which lack specific mechanisms to export instructional content in standard e-Learning formats. Two representative examples are the aforementioned TCIA and *MedPix* collections: although they provide web-service REST-based APIs to retrieve content, this content is structured in collection-specific terms, without any resemblance to e-Learning standards. It is also applicable to many other medical collections available on-line (e.g., those datasets collected from the BioLINCC initiative). While simple URL-based linking can be used to incorporate content from these existing collections in on-line courses [11], it hinders more advanced integration features promoted by e-Learning standards: content reorganization and repurposing, learner activity tracking, finer-grain learning objects, etc. Therefore, the deployment of e-Learning standard-compliant content from these collections is of utmost importance to enhancing their educational value [12]. Thus, we begin by modeling a process to address these issues.

Actors

Figure 1 shows the model proposed to represent the generation process. Since each medical collection can exhibit a particular, collection-specific organization, it is not possible to provide a universal solution for generating standardized e-Learning content from any existing medical collection. Therefore, our process model envisions the active collaboration of two different actors:

- *Developers*, who are in charge of providing the tools required for transforming the medical information data into standard learning packages during development activities: *importation / reconfiguration / exportation engine development* activities.
- *Medical instructors*, who, in turn, use the tools provided by developers to actually generate the learning packages during *generation* activities: *medical collection importation / reconfiguration* and *learning content exportation* activities.

Canonical representation

A key concept in the process model is the *canonical representation* of medical collections. Its rationale is to cope with the heterogeneity of medical collections by providing a unified framework to map the different collections. For this purpose, such a canonical representation conceives of a medical collection as one made of:

- A set of *resources* (X-ray images, free-text clinical reports, etc.), which represent the local items in the collection, as well as external items represented by their URLs.
- A set of *metadata documents*. A metadata document describes how to aggregate resources and other simpler metadata documents to yield significant learning units, and also provides additional descriptive metainformation concerning these learning units. For example, a clinical case can be described with a metadata document by aggregating a free-text report or a set of X-ray images. Additionally, it can add metadata concerning the report's author, patient data, etc.
- A *metadata schema*, which is an explicit representation of the structure followed by the metadata documents (e.g., the aforementioned structure for clinical cases).

Generation workflow

The input to the generation model is an existing *medical* collection, and the process is divided into the next steps:

- First, by using a suitable *importation engine*, which is provided by developers, medical instructors can generate the intermediate *canonical representation*. To do so, the importation engine can provide several customization mechanisms, allowing instructors to decide the specific information to be imported and to fine-tune the way in which this information will be reflected in the canonical representation.
- Then, medical instructors can transform this canonical representation with a suitable *reconfiguration engine*, also provided by developers, to make the structure required to generate the learning packages explicit. The reconfiguration engine will let instructors edit not only the resources and metadata documents, but also the metadata schema, in order to better adapt the organization of the collection to specific educational settings.
- Finally, developers can offer instructors a suitable *exportation engine*, which lets them generate *standardized learning content* from the reconfigured canonical representation. As with the importation engine, the exportation engine can also provide a set of features that the instructor can customize to generate and package the learning content. For example, when a content package concerning a set of clinical cases is generated, its components can be organized according to the instructor's needs.

As a final remark, it is worthwhile to point out that the process model is iterative and incremental in nature, as it promotes the continuous improvement of both the materials generated and the generation tools.

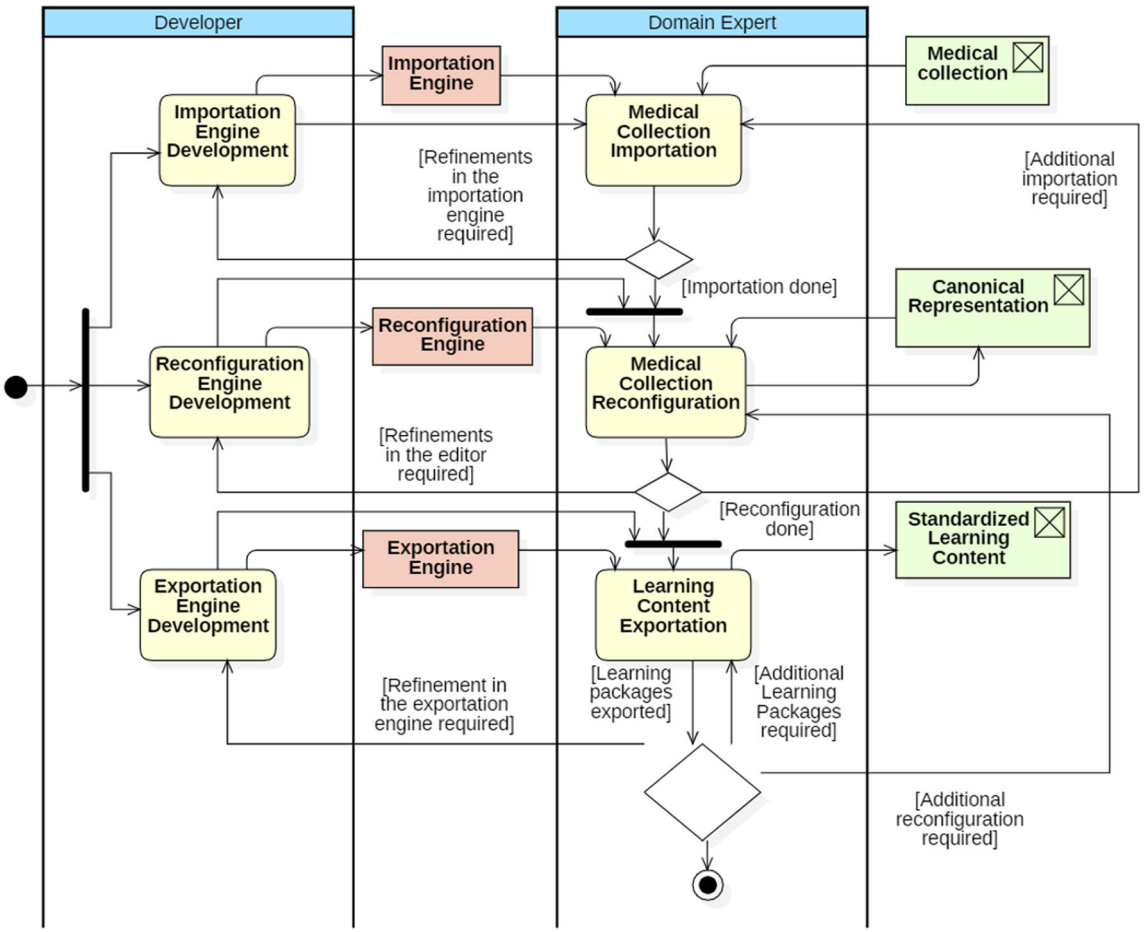


Fig. 1 The generation process model

Supporting the generation process with *Clavy*

As aforementioned, our generation process model is supported by a tool called *Clavy*. In this section we briefly review the main features of *Clavy* that are relevant from the medical context perspective during the process of generating standardized learning content.

Importation

The first step in this generation process consists of selecting a suitable medical collection in order to retrieve its information, while adding, in a further step, the corresponding instructional value. *Clavy* is equipped with an extensible importation system supported by an *importation plug-in architecture* that allows medical instructors to extract medical information from different sources such as research papers, clinical reports, or imaging databases. These plug-in elements enable the importation of data from standard platforms and formats (XML, JSON, relational databases, etc.) or their adaptation to specific extraction protocols by filtering the information required. Such features let instructors retrieve information items from medical collections and arrange them in terms of the aforementioned canonical representation.

Reconfiguration

In a second step, *Clavy* provides an extensible reconfiguration engine, which integrates:

- On the one hand, a user-friendly *schema editor* oriented to medical instructors, which does not assume any specific knowledge on data management or computer science. Indeed, since all the documents in a *Clavy* collection share the common structure described by its schema, *Clavy* lets instructors reconfigure this structure by editing only the schema (schemata in *Clavy* are conceived as hierarchical arrangements of *element types*, which model the information items in the documents). Thus, Fig. 2 shows a screenshot of this schema editor that enables instructors: (i) to rename element types whose medical denomination may be unclear; (ii) to remove useless or non-relevant elements from an instructional perspective; (iii) to merge two semantically equivalent element types; and (iv) to

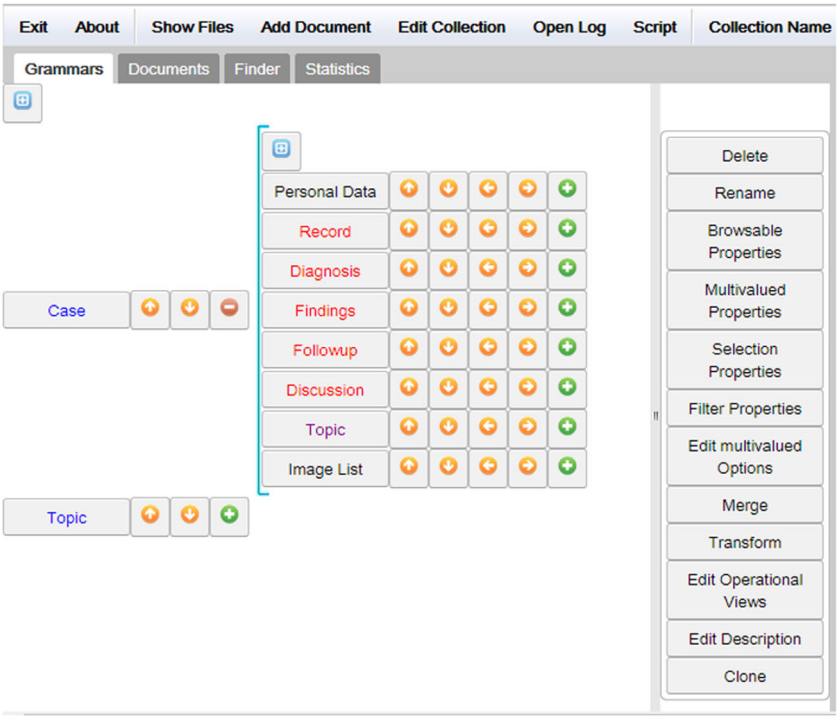


Fig. 2 Screenshot of the schema editor in the *Clavy* platform

change the hierarchical organization of elements (e.g., in Fig. 2, to organize the *patient's* data items). Once the schema is reconfigured, *Clavy* takes care of the required adjustments in the internal collection's organization in a seamless and automatic way.

- On the other hand, an extensible *document editor*, which lets instructors tweak individual documents in order to carry out more in-depth content-specific curation activities. By default, this editor is a form-oriented one as

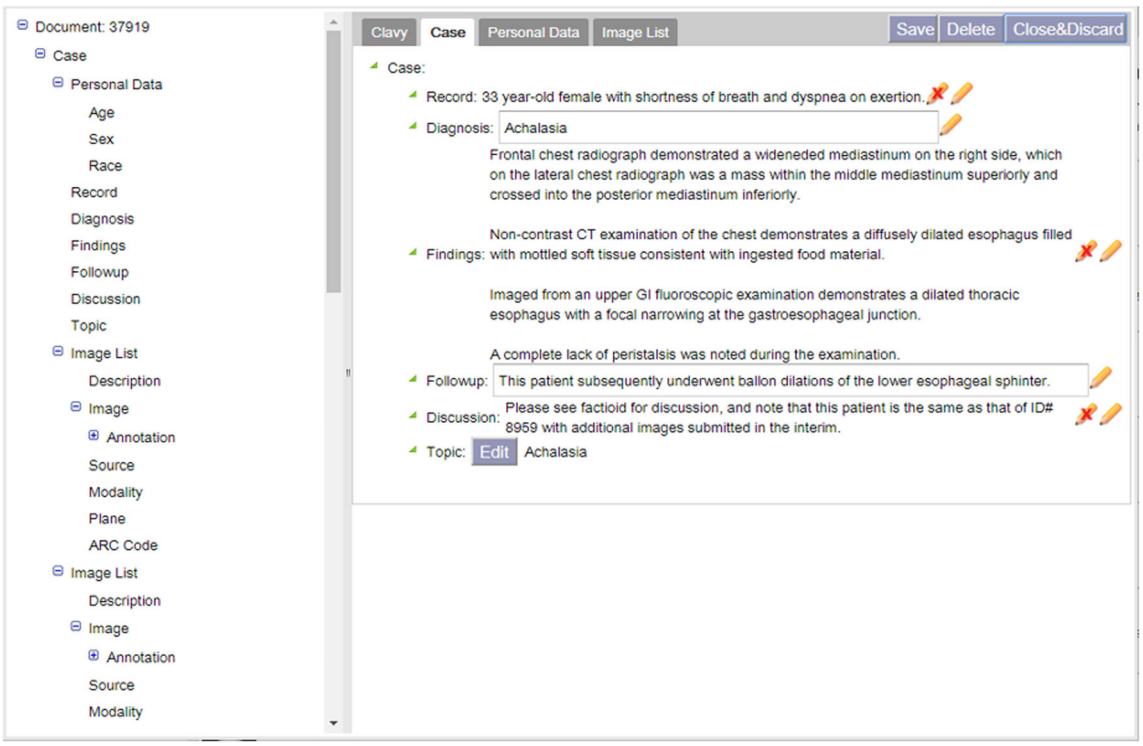


Fig. 3 Screenshot of the document editor in the *Clavy* platform

Fig. 4 An edition plug-in in action

displayed on Fig. 3 and it enables users to edit information items such as the *Findings* of the case, or the associated *Discussion*. *Clavy* makes it possible to customize the edition flow by introducing another plug-in architecture: the *edition plug-in architecture*. For instance, Fig. 4 outlines an *edition plug-in* for annotating medical images with explanations about relevant findings in the images.

Exportation

The last step in the *Clavy* generation process is addressed to exporting the previously introduced and reconfigured medical information. To support such exportation, *Clavy* provides a third plug-in architecture: the *exportation plug-in architecture*. Using this architecture, developers can add *exportation plug-ins* into *Clavy*, and medical instructors can access these plug-ins to generate the final e-Learning packages in different standard formats [13].

Assessment of the approach: From *MedPix* to IMS content packages

In order to assess our approach, we undertook the extraction of IMS Content Packages (CP hereafter) from the *MedPix* medical collection by using the *Clavy* platform. The role of *developers* was played by our programming team at the Complutense University of Madrid (Spain). The role of *medical instructors* was played by medical instructors and residents at *La Fe* Hospital (Valencia, Spain).

Importing the *MedPix* Collection

MedPix organizes all the information concerning clinical cases in a collection-specific way. In particular, in *MedPix*, *clinical cases* include clinical images and additional descriptive information and cover different *clinical topics*, since the two types of elements are cross-referenced. Therefore, a *MedPix* collection-specific *plug-in* was provided by developers. This *plug-in* used the *MedPix* REST API to retrieve all this information and reorganized it in terms of *Clavy*'s canonical representation.

Once the plug-in was available, *MedPix* materials were successfully imported into *Clavy*. Figure 5 shows an excerpt of the schema for the resulting *Clavy* collection. This schema contains 72 elements, and mirrors the basic *MedPix* structure in *Clavy* terms:

- The *Case* element groups together all the information that concerns clinical cases. Among other aspects, this element contains data related to the patient associated with the clinical case (*sex*, *age*, *diagnosis*, etc.), the *encounters* held between physicians and patient, links to topics related to the case, and clinical images associated with such a case (for each image, besides the caption and the URL, technical data similar to that supported by formats such as DICOM is also provided).

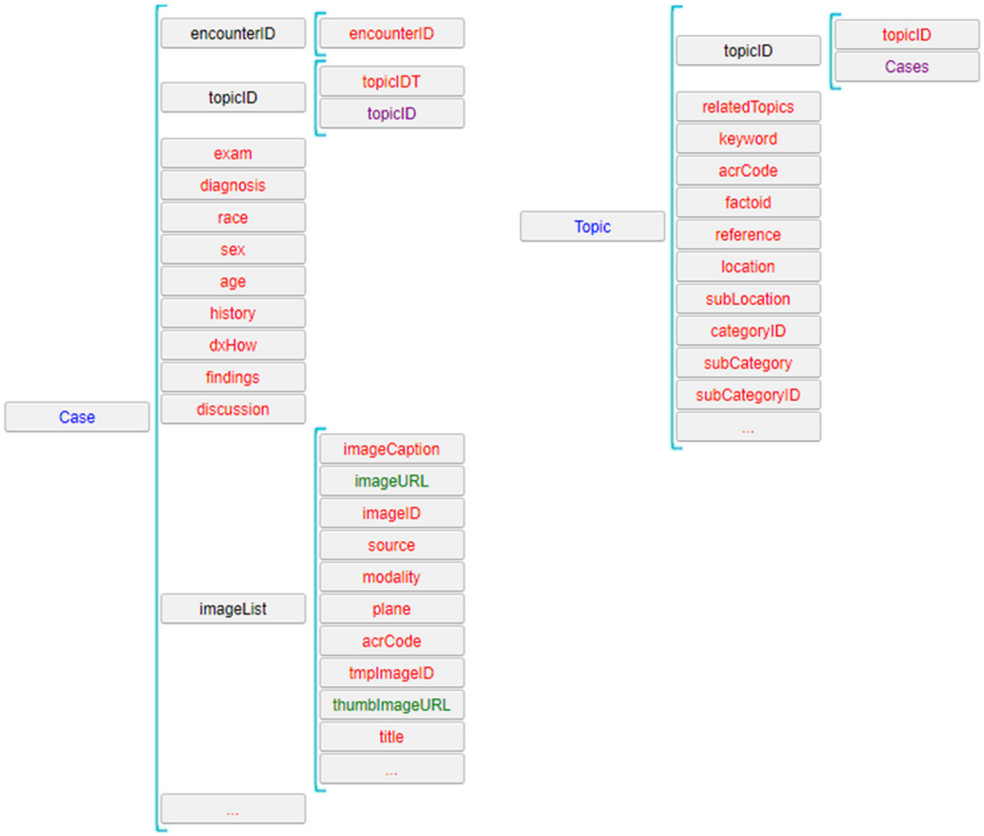


Fig. 5 Excerpt of the initial *Clavy* schema for the content imported from *MedPix*

- The *Topic* element contains all the information concerning a medical topic. Among other aspects, it supports links to related cases, description of related topics, classification in terms of keywords, the ACR Code [14], categories and subcategories, etc.

The importation plug-in was created by a single developer at the UCM within a week of full-time dedication. The resulting *Clavy* collection included a considerable volume of information: 73219 clinical images, 6758 documents associated with clinical cases, with an average of 176.37 information items per document (95% CI [172.60,180.15]), and 3969 documents for topics, with an average of 169.87 information items per document (95% CI [164.42,175.32]).

Reconfiguring the original *Medpix* structure

Once the canonical collection for *MedPix* was made available in *Clavy*, medical instructors reconfigured this collection by using the schema and document editors.

Most of the reconfiguration process was carried out by reorganizing the collection schema. The schema was tweaked by instructors to:

- Suppress information which they considered unnecessary for educational purposes; for instance, elements like internal *MedPix* IDs used for cross-reference and internal organization, or detailed properties of the images that had hardly any serious educational value, and therefore were erased by instructors from the schema.
- Promote a better structure for the documents. For instance, all the information concerning the patient in a clinical case was grouped behind a *Personal Data* structural element, elements were given more descriptive names, etc.

Figure 6 shows the resulting *Clavy* schema after this reconfiguration process. The 72 initial elements were reduced to 28, those that instructors considered the most useful from an educational point of view, plus some additional elements oriented towards enhancing the collection structure (e.g., the aforementioned *Personal Data* structural element). Medical instructors found *Clavy*'s canonical representations intuitive, and the schema and document editors easy to use. Finally, the document editor let them inspect how changes in the schema propagated and curate contents without significant difficulties.

Exporting *Medpix* cases

We developed an exportation plug-in to export *Clavy* content as standardized learning packages (*exportation engine development* activity). Such plug-in components used additional attributes attachable with information items to decide where to represent

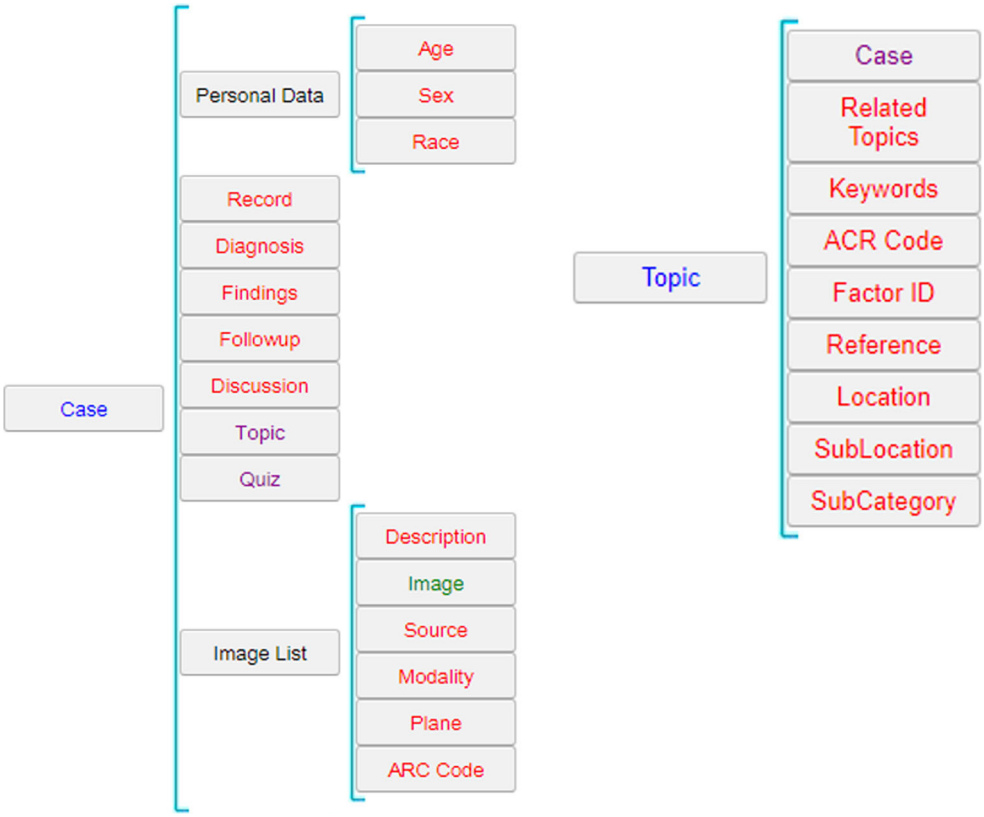
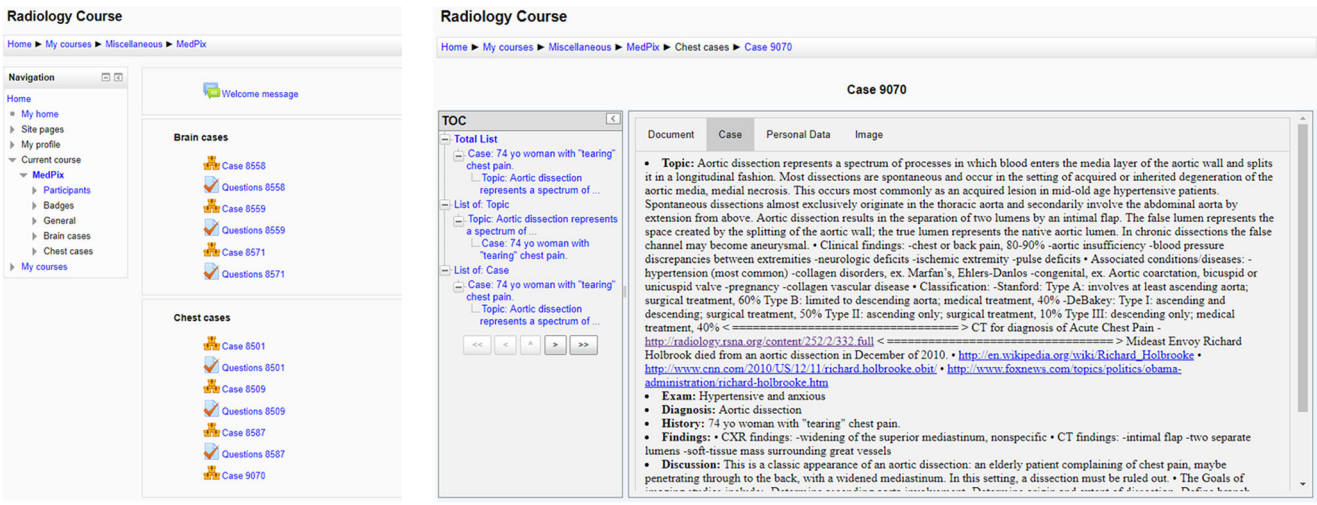


Fig. 6 Reconfigured *Clavy* schema for the information items imported from *MedPix*

descriptive elements as organization items or as additional IMS CP resources. It also monitored and broke potential cycles in the link structure, since the IMS CP information model is primarily tree-shaped. This plug-in was also developed by a single programmer during an additional full-time week of dedication.

By using this *plug-in*, content packages, ready to be incorporated into any learning management system that supports the IMS CP standard, were exported from the reconfigured *MedPix* content. For instance, Fig. 7a shows a screenshot of a sample radiology course deployed in Moodle, which interleaved the description of clinical cases incorporated as IMS content packages generated from *Clavy* with quizzes on the cases also gathered from *MedPix*. Figure 7b illustrates the case information associated with one of these clinical cases such as the topic



(a) Course Structure. (b) An IMS CP

Fig. 7 An example of Moodle course that integrates *MedPix* clinical cases bundled as IMS content packages

covered, its findings or their discussion. As reported in [9], medical instructors provided very positive feedback concerning the instructional usefulness of this kind of course.

Conclusions and future work

In this paper, we have described a novel approach to the generation of standardized learning content from existing medical collections. The approach presented relies on the importation of data and content from this kind of collections according to specifically designed plug-in components. Therefore, the imported collection can be adequately reconfigured and adapted to the specific instructional needs of particular training scenarios. Finally, the assets in the reconfigured collection can be exported and packaged using widely used e-Learning standards. We have also shown how this approach was implemented by using an information management tool called *Clavy*, and we have assessed its feasibility with a case study concerning the importation, reconfiguration and exportation as IMS content packages of clinical cases coming from *MedPix*, a well-known digital medical collection in the field of radiology.

Currently, we are working on supporting the exportation of *Clavy* documents to other standard e-Learning formats that support interaction (e.g., in particular, SCORM packages) [15]. We are also studying the importation of interactive quizzes coming from *MedPix* and the deployment of these quizzes in SCORM packages. Further research plans to support the exploitation of *Clavy* aggregation features to define learning paths, the exportation of content in terms of the IMS *Common Cartridge* recommendation [15], and the assessment of courses integrating learning content generated through our approach in hospital residency programs.

Funding This study was funded by the Ministry of Science, Innovation and Universities, Spain (grants numbers TIN2014–52010-R and TIN2017–88092-R).

Compliance with ethical standards

Conflict of interest Félix Buendía declares that he has no conflict of interest. Joaquín Gayoso-Cabada declares that he has no conflict of interest. José-Luis Sierra declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., and Prior, F., The Cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging* 26:1045–1057, 2013. https://doi.org/10.1007/s10278-013-9622-7.
2. Ross, J. S., Ritchie, J. D., Finn, E., Desai, N. R., Lehman, R. L., Krumholz, H. M., and Gross, C. P., Data sharing through an NIH central database repository: A cross-sectional survey of BioLINCC users. *BMJ Open* 6, 2016. https://doi.org/10.1136/bmjopen-2016-012769.
3. Altman, M., The clinical data repository: A challenge to medical student education. *Journal of the American Medical Informatics Association : JAMIA* 14:697–699, 2007. https://doi.org/10.1197/jamia.M2483.
4. Schwartz, A., Pappas, C., and Sandlow, L., Data repositories for medical education research: Issues and recommendations. *Medical Education Research* 85:837–843, 2010.
5. MedPix. https://MedPix.nlm.nih.gov/. Last Accessed 12/05/2018.
6. Gros, B., and García-Peñalvo, F. J., Future trends in the design strategies and technological affordances of E-learning. In: Spector, M. J., Lockee, B. B., Childress, M. D. (Eds), *Learning, design, and technology: An international compendium of theory, research, practice, and policy*. Cham: Springer International Publishing, 2016, 1–23.
7. Gayoso-Cabada, J., Rodríguez-Cerezo, D., and Sierra, J.-L., Browsing digital collections with reconfigurable faceted thesauri. In: Goluchowski, J., Pankowska, M., Linger, H., Barry, C., Lang, M., Schneider, C. (Eds), *Complexity in information systems development*. Springer International Publishing, 2017, 69–86.
8. IMS Content Packaging Specification Primer Version 1.2 Public Draft V2.0.
9. Buendía, F., Gayoso-Cabada, J., Sierra, J-L, From Digital Medical Collections to Radiology Training E-Learning Courses. In: Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality. ACM, New York, NY, USA, pp 488–494, 2018.
10. Buendía, F., Gayoso-Cabada, J., Sierra, J-L, Using Digital Medical Collections to Support Radiology Training in E-learning Platforms. In: Pammer-Schindler V, Pérez-Sanagustín M, Drachsler H, Elferink R, Scheffel M (eds) Lifelong Technology-Enhanced Learning. Springer International Publishing, pp 566–569, 2018
11. Hsiao, C.-H., Hsu, T.-C., Chang, J. N., Yang, S. J. H., Young, S.-T., and Chu, W. C., Developing a medical image content repository for E-learning. *Journal of Digital Imaging* 19:207–215, 2006. https://doi.org/10.1007/s10278-006-0588-6.
12. Bamidis, P. D., Konstantinidis, S. T., Kaldoudi, E., Bratsas, C., Nikolaidou, M. M., Koufogiannis, D., Maglaveras, N., Pappas, C., New Approaches in Teaching Medical Informatics to Medical Students. In: 2008 21st IEEE International Symposium on Computer-Based Medical Systems. pp 385–390, 2008
13. Kurilovas, E., Interoperability, standards and metadata for E-learning. In: Papadopoulos, G. A., Badica, C. (Eds), *Intelligent distributed computing III*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, 121–130.
14. American College of Radiology. Index for Radiological Diagnoses. 4th ed. Reston, Va: American College of Radiology, 1992.
15. Gonzalez-Barbone, V., and Anido-Rifon, L., From SCORM to common cartridge: A step forward. *Computers & Education* 54: 88–102, 2010. https://doi.org/10.1016/j.compedu.2009.07.009.