Abstract: - The process of clinical practice can be represented as observations leading to diagnoses that trigger the appropriate actions on the patient’s status. In endoscopic surgeries for example, the physician’s observations play an essential role in the patient’s diagnosis. In such context, paper documents are frequently used to record and depict the medical endoscopic findings. They primarily rely on hand-drawn sketches to map the observations made of each endoscopic case. The Hysteroscopy procedure follows such practice of degrees of information, which generates heterogeneity and lack of interoperability in the process of recording digital medical information. In order to solve above mentioned problems and achieve semantic interoperability of Hysteroscopy findings between heterogeneous healthcare systems, the present study integrates an automatic mechanism for recording such findings with clinical archetypes, according to the two-level modeling. A combination of clinical image processing techniques with openEHR archetypes is used. In addition to formalizing the representation of Hysteroscopy, clinical archetypes include bindings to terminologies, such as SNOMED-CT, that semantically describe the findings. The workflow of the proposed approach can also be followed in clinical areas involving similar sketching process and images analysis.

Keywords: semantic interoperability, clinical archetypes, Hysteroscopy, SNOMED-CT, electronic health record, image processing, clinical observation sketching
1. Introduction

The term “semantic interoperability” (SIoP) has been used for more than a decade in computer science as the ability to exchange services and data between components of large-scale, distributed systems in a way that ensures the requesters and providers have a common understanding of the meanings of the requested services and data [1]. With regard to the medical domain, SIoP is mainly associated to the speed and consistency of accessing meaningful health related data. The term has rapidly gained importance since the research by Kalra, Blobel, and Regensburg [2] and by Aspden [3] pointed out that SIoP is essential if we aim to enable electronic health record information to be shared seamlessly and meaningfully, and computational services are to be able to interpret safely clinical data that has been collected from different sources. Then a milestone was reached when the “Semantic interoperability for better health and safer healthcare” report was published by the Semantic HEALTH project [4]. It is a comprehensive report including the goals, trends and technical challenges to achieve SIoP between heterogeneous healthcare systems. In the field of EHR systems, the report acknowledges the model of clinical archetypes, supported by the CEN/ISO EN13606 standard and the openEHR foundation, as a mechanism to express data structures in a shared and semantically interoperable way. The model involves three layers of artifact to represent meaning: (i) generic reference models (e.g. the openEHR RM [5]) for representing clinical EHR data; (ii) agreed clinical data structure definitions, expressed in terms of constraints on the generic reference model (e.g. openEHR archetypes [6]) and; (iii) clinical terminology systems to enable the faithful exchange of meaning between machines and between machines and people (e.g. SNOMED-CT1). Therefore, the development of new techniques, procedures and tools for the integration of these three layers in the clinical practice is among the high priority tasks in the road towards full semantic interoperability.

While in some domains such integration may be addressed as a straightforward process, there are clinical areas such as the ones involving image analysis and visual medical examinations where a significant gap between the physician’s conclusion and the data actually stored in EHR systems must be bridged. For example, in the endoscopic surgery area, where the physician’s observation plays an important role in the final diagnosis, there are procedures such as the endoscopic Hysteroscopy where observations are not established as objective diagnostic elements [7]. Thus, a common technique between physicians is that of mapping their observations of each case by sketching their medical endoscopic findings on hard-copy documents [8–10]. Given the heterogeneity of practices, such technique is highly vulnerable to SIoP problems and may subsequently lead to inefficient healthcare.

The present research can be divided in two main parts, the objectives of which is to bring each end of the above mentioned gap closer to the other, i.e., getting the image assessment in the Hysteroscopy procedure closer to EHR systems based on clinical archetypes and vice-versa. On one hand, and from the EHR perspective, a clinical archetype formally describing the structure of Hysteroscopy findings and their SNOMED based semantics is defined. On the other hand, image analysis, as well as data grouping and clearing techniques, are implemented in a medical application that extend the parent application introduced in [11] to aid semantic annotation and management of Hysteroscopy findings using the clinical archetype from the other part. The research has received the support from the office of hysteroscopy in the Gynecological Department of the Medical School of the University of Ioannina2.

The proposed workflow of: (i) defining a clinical archetype that constraints the structure and contents of the image observation, (ii) mapping the archetype elements to a clinical terminology, and (iii) develop visual
applications and image analysis techniques to instantiate the clinical archetype with concrete patient data; can be further applied to clinical areas which are similar to Hysteroscopy in the sense that they are also suffering from the abovementioned gap. The rest of the paper is structured as follows. Section 2 covers the related work. Then Section 3 introduces the clinical standards and formal languages on which the proposed approach is based, continuing with Section 4 where the implementation of the two-level paradigm for the semantic interoperability of clinical observations sketches is described. Further on, Section 5 describes the sketch edition and management application that has been developed to harmonize physician’s traditional practice with the interoperability requirements. Finally, conclusions and future work are provided in Section 6.

2. Related work
The current state of the art in the exchange of annotated biomedical images is the result of many years of research and development of standards, tools and workflow methodologies. The Digital Imaging and Communications in Medicine (DICOM) Standard [12], which specifies a non-proprietary data interchange protocol, digital image format, and file structure for biomedical images and image-related information, is nowadays widely adopted. As a complement to DICOM, the goal of the Annotation and Image Markup (AIM) project is to develop a standard for serializing image annotation and markup data to solve the lack of pixel data meaning in DICOM’s metadata [13]. Then Picture Archiving and Communication Systems (PACS) and tools such as the Biomedical Image Metadata Manager (BIMM) [14] use these standards to support computer-aided detection and diagnosis and content-based image retrieval, among other applications.

Thus, one of the major challenges for ICT solutions is the integration of binary imaging standards with the text-based EHR standards that model everything else in the clinical workflow. Significant efforts have been carried out for this purpose, including the research by Blazona [15] that addresses HL7 and DICOM integration. These approaches typically involve translation scripts and mapping mechanism between standards to achieve the integration. In contrast with automatically generated and highly detailed images such as the ones in radiology, the observation sketches used to describe certain visual examinations, such as Hysteroscopy, are human-drawn and relatively simple images that graphically represent the findings. Therefore, the present research is based on a different approach to the abovementioned integration. By exploiting the flexibility of the two-level paradigm and the fine-grained SNOMED-CT, the proposed method directly captures all the semantics and interpretations of Hysteroscopy sketches into the widely used, text-based, openEHR standards. As a result, no integration problems must be solved beyond the Reference Model to local and heterogeneous healthcare systems, which is common to any clinical area.

3. Background
For the purpose of achieving interoperability in the healthcare domain, several knowledge artifacts and formal languages addressing the structural, syntactic and semantic aspects of interoperability must be integrated. This section briefly introduces the ones involved in the current research.

3.1 Two-Level paradigm and clinical archetypes
The two-level modeling is a promising solution for the semantic interoperability of healthcare information [16–18]. According to the openEHR approach, the two-level model contains a generic Reference Model (RM) that defines a logical information architecture for the interoperability of EHR systems, constituting a base representation framework. The RM represents the general features of the components of the EHR, how they are organized and the context information needed to satisfy both the
ethical and legal requirements of the record. It includes a flexible syntax and some generic types of clinical information as observations, evaluations, instructions and actions. Then, instances or specializations of that RM are devised in the form of constraints expressed through more concrete “archetypes”, which serve as a shared language for common and specialized clinical concepts. In other words, the RM encloses the stable features like the set of classes that make up the blocks constituting an EHR and the basic syntax of statements, while archetypes allow for sharing a wide variety of combinations of those classes corresponding to EHR fragments created for specific clinical situations. According to Beale, Heard, Kalra, and Lloyd [5] there are four basic types of clinical archetypes: OBSERVATIONs are distinguished from ACTIONs in that ACTIONs record interventions whereas OBSERVATIONs record only information relating to the situation of the patient, not what is done to him/her. Then, an EVALUATION may say for example that “oral cortico-steroids are indicated at a peak flow of 200 l/m”. A corresponding INSTRUCTION would be issued to indicate the actual drug, route, dose, frequency, and so on. The Hysteroscopy archetype, whose definition is described in Section 4, is a specialization of the OBSERVATION entity.

3.2 The ArchetypeDefinition Language
The openEHR Archetype Definition Language, or ADL [19], is a formal language for expressing archetypes that can be categorized as a knowledge description language. It provides a formal, abstract syntax for describing constraints on any domain entity whose data is described by the RM. The syntax is congruent with Frame Logic queries [20]. Further details about ADL are explained in Section 4 for the particular case of the Hysteroscopy archetype.

3.3 SNOMED Clinical Terminology
SNOMED-CT3 is a comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and, as of April 2007, owned, maintained, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a non-for-profit association in Denmark. It provides the core general terminology for the EHR and contains more than 311,000 active concepts with unique meanings and formal logic-based definitions organized into hierarchies [21]. It was selected for this research because of the following reasons:

- As it is widely used in Europe and all over the world, the proposed approach could be significant for several researches in the EHR community.
- It has a large coverage of the clinical concepts required for the Hysteroscopy case study that has been carried out.
- It is simple to use and query, in spite of its large size.
- At least part of the terminology has been classified by DL reasoners, making it more reliable when compared to completely unclassified terminologies. The reliability is based on the assumption that logical errors would have been resolved during classification. The semantic completeness of SNOMED-CT has been audited by Jiang and Chute [22].

4. Archetyping Hysteroscopy findings
This section covers the first two steps of the workflow presented in the Introduction section, i.e., (i) defining a clinical archetype that constraints the structure and contents of the image observation and (ii) mapping the archetype elements to a clinical terminology. The Hysteroscopy archetype will standardize the communications of the clinical data related to hysteroscopies, regardless of the heterogeneity of different healthcare systems, as it is based on the generic RM introduced in the Background section. The ADL file representing the
Hysteroscopy findings starts with a header segment followed by a definition segment and an ontology segment. ADL includes two syntaxes: the data definition syntax or dADL, and the constraints definition syntax or cADL. The main definition segment is written in cADL, and will be described in Section 4.1, that corresponds to step (i). The ontology segment is written in dADL and will be described in Section 4.2, which corresponds to step (ii). For the sake of simplicity, only the most significant fragments of the ADL file are discussed in this manuscript. The full ADL file, including the header segment uniquely identifying the clinical concept and containing metadata about the archetype (e.g. its purpose and use), can be downloaded from http://www.ieru.org/archetypes/openEHR/.

4.1 Structural Interoperability of Hysteroscopy findings

In order to avoid the ambiguity and heterogeneity of the sketches currently used by physicians to describe Hysteroscopy findings, as well as to support further semantic tasks over image repositories, such as classification and semantic search, the instances of the Hysteroscopy archetype must be able to represent the information in Figure 1. It should be noted that physicians’ know-how and experience in the Hysteroscopy sketching process is not affected as the application described in Section 5 automatically transforms the visual annotations into the coded values required by the archetype.

4.2 Semantic Interoperability of Hysteroscopy findings

In collaboration with the office of hysteroscopy of the Gynecological Department of the Medical School of the University of Ioannina, a SNOMED-CT analysis has been performed to investigate the “spatial” division of the uterus supported by the terminology and selected the division terms to be used in the archetype instances and graphical applications. It should be
noted that some SNOMED-CT terms correspond to overlapping regions of the uterus. Therefore, specific terms must be selected from the terminology so that no problems arise during the processing. Then, the colored representation of the uterus in Figure 3 has been created, where every point is characterized by the SNOMED-CT code of the region to which it belongs. The application in Section 5 allows the physician to use this enriched diagram as the point of departure for sketching the Hysteroscopy findings by means of an image editing interface.

The subset of SNOMED-CT in Figure 3 is captured by the Hysteroscopy archetype, specifically within an ontology segment.

5. Integration in healthcare systems and case study

The present section introduces the Graphical User Interface that has been developed according to the previously described Hysteroscopy clinical archetype. The application is designed to preserve the physician’s know-how during the sketching of Hysteroscopy findings while generating a formal and standardized record of the findings. Also semantic searches, classification and management of the Hysteroscopy sketches are supported by the application. A concrete case study is presented throughout the following Figure 4.

5.1 Sketching Hysteroscopy findings

The workflow of the sketching application can be broken down into the following stages:

1) Sketch creation and edition
2) Processing
   a. Linear Image Analysis (LIA)
   b. Data grouping and clearing (GC)
3) Metadata edition and archetype instantiation
5.1.1 Sketch creation and edition

For the image edition stage, the GUI illustrated in Figure 4 has been created according to the patent in [23]. The image is a pre-drawn sketch of a typical uterus where the user can characterize regions by means of the included toolbox and depending on his/her findings. The colours used in the toolbox to define the findings are matched with the corresponding SNOMED-CT terms. Figure 5. Sketch drawn by a physician after performing an hysteroscopy examination along with the patient status data required for the Hysteroscopy Archetype instantiation so when the user clicks on one of the coloured buttons, the SNOMED-CT code for that finding is automatically assigned. The implementation of this interface has been carried out using open-source components for image edition. Furthermore, the Hysteroscopy sketching application is connected to a previously Figure 4. GUI supporting the edition of uterus sketches and the Hysteroscopy archetype instantiation. developed application for the management of the Hysteroscopy procedure [11]. Such “parent application” supplies the patient state data, which is stated in the Hysteroscopy archetype to be required for the proper assessment of the findings. As the content in the parent application is already annotated with SNOMED-CT codes, retrieving the patient state data is a straight forward and automatic process. As an example a case study is described for which the data and sketch in Figure hysteroscopy procedure was undergone, and a finding sketch drawn.

5.1.2 Image Processing

At this step, the following data structure is defined containing two fields for pixel description. The first field stores the SNOMED-CT code linked to the colour code used in the physician’s remarks, and the area_code stores the spatial position of the pixel.

```
Pixel_Property=Record
finding_code: longint;
area_code: longint;
end;
```

Next, a linear analysis of every pixel of the sketch is performed. The logical of Linear analysis are followed for this procedure, whose resulting data is then grouped and cleared As part of the clearing and grouping step, the array created in the linear analysis is sorted by the type of area and then the
data is grouped by the type of findings to achieve totals. A statistical analysis of the sorted and grouped data is performed in order to clear the extreme values and to reject the clearly incorrect values.

Linear image analysis (LIA)

Results from the linear analysis of the image in Figure 5:
1. A large number of pixels in the area of the left side of the image (corresponding to the bottom and left horn of uterus) are green.
2. The pixels corresponding to the right fallopian tube region are all green.
3. The pixels corresponding to the left fallopian tube region have no colour.
4. A small amount of pixels corresponding to the left wall of uterus are green.

Clearing and Grouping (CG)

Results from the automatic clearing and grouping of the linear analysis of the image in Figure 5:
1. 87.2% of the uterus left horn has green colour
2. 44% of the bottom of the uterus have green colour
3. 7% of the left uterus wall have green colour
   The part of the results corresponding to the uterus wall was too small so it was considered to be a drawing error, and therefore cleared by the application.

5.1.3 Metadata edition and archetype instantiation

The final results of the image processing stage, including SNOMED-CT codes for

<table>
<thead>
<tr>
<th>Hysteroscopy archetype</th>
<th>SNOMED-CT code</th>
<th>ELEMENT value</th>
<th>SNOMED-CT code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding site</td>
<td>361690007</td>
<td>Bottom of uterus</td>
<td>26135004</td>
</tr>
<tr>
<td>Percentage of total</td>
<td>23875000</td>
<td>87.20%</td>
<td></td>
</tr>
<tr>
<td>Finding site</td>
<td>361690007</td>
<td>Left horn of uterus</td>
<td>26148004</td>
</tr>
<tr>
<td>Percentage of total</td>
<td>23875000</td>
<td>44%</td>
<td></td>
</tr>
<tr>
<td>Finding site</td>
<td>361690007</td>
<td>Right fallopian tube</td>
<td>26148004</td>
</tr>
<tr>
<td>Percentage of total</td>
<td>23875000</td>
<td>7.78%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Case study results required to instantiate the data segment of the Hysteroscopy archetype.

uterus regions and Hysteroscopy findings, as well as the percentage of the image covered by such findings, are submitted as metadata along with the sketch (JPG image) in a clinical repository. Table 1 contains the results from the case study we have been following in the last subsections. Such
information is then combined with the patient’s state data and ultimately used for instantiating the Hysteroscopy archetype.

Then the sketch repository is enriched with index files to support fast searching of key terms without the need of extensive image analysis. A screenshot of the search interface is shown in Figure 6.

Figure 6. Interface for searching Hysteroscopy findings

6. Conclusion

In accordance with the informatics healthcare priorities established by the European Commission in [4], a method has been proposed to bridge the gap between endoscopic observations requiring sketching and the generation of semantically interoperable descriptions of the findings. In particular, the case for the annotation of Hysteroscopy sketches with SNOMED-CT codes and the definition and automatic instantiation of the Hysteroscopy clinical archetype has been described. The contribution on bridging the gap is twofold. On one hand, the definition of the Hysteroscopy archetype, in collaboration with the Gynaecology Department of the Medical School of Ioannina, formally models the structure and semantics of Hysteroscopy findings. On the other hand, the development of the sketch editing application encourages both the reuse of the archetype definition and provides a tool for the standardized training and education of gynaecology students.

The vast amount of medical images produced daily in medical procedures, creates the need for management techniques not based on the technical characteristics of the images, but on standardized medical information associated to them. The presented approach utilizes a simplified and digitized mapping of the uterus, combined with the SNOMED Clinical Terminology, to solve the problem. A key innovation feature is the imprinting of the SNOMED-CT coding system in the recorded information, without requiring the physician and/or operator to have a specialized knowledge of the terminology.

In addition to semantic interoperability issues between healthcare systems, such as the accuracy of the mappings to SNOMED-CT, providing a seamless and fast sketching and data edition mechanism was also addressed as a key objective. In spite of the fact that this application cannot replace a specialized recording tool when high levels of clinical detail are required, the offered simplification and automation makes the proposed solution ideal for quick creation of large amounts of annotated sketches


