

Towards the Creation of the Cardiovascular Magnetic Resonance Quality Assessment Ontology (*CMR-QA*)*

Ernesto Jiménez-Ruiz^{1,4}, Valentina Carapella², Elena Lukaschuk²,
Nay Aung³, Kenneth Fung³, Jose Paiva³, Mihir Sanghvi³,
Stefan Neubauer¹, Steffen Petersen³, Ian Horrocks⁴, Stefan Piechnik²

¹ Logic and Intelligent Data Group, Department of Informatics, University of Oslo, Norway

² Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Radcliffe Department of Medicine, University of Oxford, UK

³ William Harvey Research Institute, NIHR Cardiovascular Biomedical Research Unit at Barts, Queen Mary University of London, UK

⁴ Information Systems Group, Department of Computer Science, University of Oxford, UK

1 Background

UK Biobank⁵ is a large scale population study that started in 2006 aimed at improving the understanding, diagnosis and treatment of a wide range of diseases, such as cancer, stroke or cardiac pathologies [2]. The recruitment of volunteers took place at the national level and reached the considerable number of 500,000 volunteers aged 40-69 across UK. All volunteers agreed to go through a series of clinical tests and have their health condition checked on follow-up. Different clinical imaging modalities are included in the protocol applied to the volunteers. In particular, a number of Cardiac Magnetic Resonance Imaging (CMR) sequences are employed to evaluate cardiac function: cine-MRI, tag-MRI, T1-mapping and blood flow imaging [3]. The work presented in this paper is related to a specific pilot study of 5,000 CMR scans limited to cine-MRI, the most common CMR modality in clinical practice. The results of this pilot study are about to be released to the public together with a first part of the associated data analysis,

Data analysis for the 5,000 cine-MRI was carried out by a team of *observers* from two centres, OCMR and Barts Hospital, and consisted of two parts: image analysis and assessment of image quality. Image analysis is essentially manual delineation of contours (also known as segmentation) of the four chambers of the heart, which then results in the computation of fundamental parameters of cardiac function. Quality assessment of the Cine-MRI scans was carried out through a combination of free-text comments and numerical quality scores. Figure 1 highlights the two components of the analysis. Quality assessment and general data analysis progress was managed through a shared spreadsheet by the team. For the purposes of our work, we focus only on the quality assessment data and the combination of numerical quality scores and free-text annotation. The quality scores alone (1 = optimal, 2 = suboptimal, 3 = not-analysable or

* This paper represents a short but more technical version of the paper “Towards the Semantic Enrichment of Freetext Annotation of Image Quality Assessment for UK Biobank Cardiac Cine MRI Scans” [1]

⁵ <http://www.ukbiobank.ac.uk/>

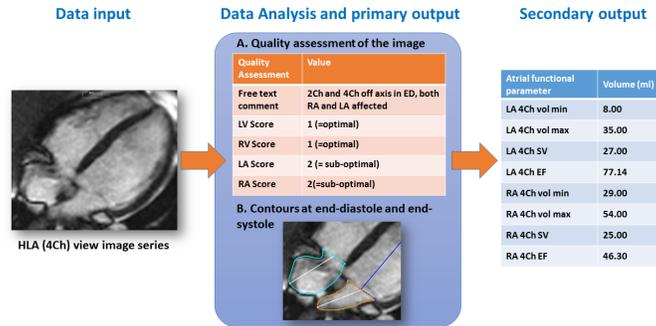


Fig. 1. Example of analysis pipeline combining image analysis and image quality assessment. Quality scores are 1 = optimal, 2 = sub-optimal and 3 = unreliable or non-analysable.

not-reliable) only provide a quick overall classification. The free-text annotation is rich in information but cannot be processed in an easy and efficient manner as the numerical scores.

In this ongoing work we aim at employing tools from the Semantic Web for the efficient structuring of free-text documentation. The semantics of the free-text annotations, which describe the quality of the image analysis, will be defined via a structured vocabulary or ontology, which we are going to call *CMR-QA* (Cardiovascular Magnetic Resonance Quality Assessment). The aimed semantic layer (ontology, rules and data) will provide machine-readable data and will be a powerful tool for (i) fast and efficient processing of the free-text comments; (ii) automatic image quality assessment from such comments and generation of quality scores; (iii) evaluation of the quality of the free-text comments in terms of information completeness, ambiguity and variability; (iv) training purposes (e.g., showing preferred annotation styles for different types of images); (v) efficient semantic access (i.e. database querying) to the images by the UK Biobank target users, such as researchers in the field of automatic segmentation, or clinical researchers who need a specific subset as a control group in their study.

2 Definition of the semantic layer

Free-text annotation are prone to variability. Variability is due to natural human variability (e.g., used nomenclature, differences in point of view), but it can also reflect different ranges of professional expertise. For example, different observers can correctly flag as *LA off axis*, *LA out of plane*, *wrong LA plane*, *2Ch out of plane*, or *LA foreshortened* an image where the plane chosen to acquire a long axis view was not optimally aligned to measure left atrial (LA) volumes.

CMR-QA ontology aims at defining the semantics of the free-text annotations in order to help identify commonalities and disagreement among different annotations. An ontology may reveal, as a source of variability, the use of different synonyms, or the use of narrower, broader or even sibling terms for the same kind of annotations. The use of an ontology (in combination with rules) may also reveal a source of ambiguity or incompleteness if some required information is not provided (e.g., the annotation *LA off axis* should always refer to the cardiac cycle where it was observed).

2.1 The ontology

The *CMR-QA* will include general knowledge about the domain,⁶ for example it will encode that the concepts *CineMRI Scan* and *TI-Mapping Scan* are a kind of *MRI Scan*. It will also encode more concrete knowledge about the quality assessment process, for example, *wrong image plane orientation* is a kind of *technical issue*. Similarly, *RA off axis* is a specific type of *wrong image plane orientation*, while *mistriggering* can be either a kind of *technical issue* (as an artefact) or a *patient-related issue* (as a pathology). The ontology also encodes relationships between concepts, for example the property *has technical issue* relates concepts with types of *technical issue*. Non-logical knowledge in the form of lexical information can also be added to the ontology; for example, the ontology may include that *RA out of plane* is an alternative label for *RA off axis*. Equations (1)-(6) show the formalization of the above knowledge into ontology axioms.⁷

- SubClassOf(*CineMRI_Scan MRI_Scan*) (1)
- SubClassOf(*Wrong_Orientation Technical_Issue*) (2)
- SubClassOf(*RA_offaxis Wrong_Orientation*) (3)
- SubClassOf(*Mistriggering ObjectUnionOf(Technical_Issue Patient_Issue)*) (4)
- ObjectPropertyRange(*has_technical_issue Technical_Issue*) (5)
- AnnotationAssertion(*alt_label RA_offaxis RA out of plane*) (6)

2.2 The rules

The ontology is being extended with (manually created) rules to infer implicit knowledge from the annotations. For example if the free-text annotation includes the comment *RA off-axis* then the comment is necessarily referred to the *Horizontal long axis (HLA) view*. Analogously, the free-text comment *basal slice is missing* implies a lack of coverage associated to the *short axis (SA) view*. The rules will also be used to infer the quality assessment scores. For example, *Lack of coverage* will always lead to a quality score associated to the right and left ventricle equal or greater than 2 (e.g., suboptimal or unreliable). In addition, rules may also be used to reveal potential incompleteness or ambiguity. For example, we can classify the technical issue *LA off axis* as incomplete if the cardiac cycle is not indicated. Equations (7)-(9) show the formalization of some of the aforementioned rules.⁸

⁶ We could not find any ontology in BioPortal [4] meeting all our requirements, which evidences the necessity of a more specific ontology in this particular domain.

⁷ We use OWL functional-style syntax: <https://www.w3.org/TR/owl2-syntax/>.

⁸ We use datalog rules (a subset of Prolog). A rule of the form $A(x) \leftarrow B(x) \wedge R(x, y) \wedge C(y)$ means that the combination of the concepts B and C via the relationship R (for the given individuals 'x' and 'y') implies that the individual 'x' is a member (or a type) of the concept A.

$$\begin{aligned} \text{has_issue_in_HLA_view}(mri, issue) \leftarrow & \text{CineMRI_Scan}(mri) \wedge \\ & \text{has_technical_issue}(mri, issue) \wedge \text{RA_offaxis}(issue) \end{aligned} \quad (7)$$

$$\begin{aligned} \text{has_RV_quality_score}(mri, 2) \leftarrow & \text{CineMRI_Scan}(mri) \wedge \\ & \text{has_technical_issue}(mri, issue) \wedge \text{Lack_coverage}(issue) \end{aligned} \quad (8)$$

$$\begin{aligned} \text{IncompleteIssueDefinition}(issue) \leftarrow & \text{RA_offaxis}(issue) \wedge \\ & \text{not observed_in_cardiac_cycle}(issue, cc) \end{aligned} \quad (9)$$

2.3 The data

The free-text content in the spreadsheet is rich in information but cannot be processed in an easy and efficient manner. In this work we are developing custom named entity recognition (NER) techniques to transform the free-text comments into semantically rich data according to the *CMR-QA*. For example, the free-text comment “*basal slice is missing. wrong planes ra/la*” associated to a *CineMRI Scan* leads to the following seven ontology facts:

$$\begin{array}{ll} \text{CineMRI_Scan}(mri) & \text{has_technical_issue}(mri, issue_1) \\ \text{Lack_coverage}(issue_1) & \text{has_technical_issue}(mri, issue_2) \\ \text{RA_offaxis}(issue_2) & \text{has_technical_issue}(mri, issue_3) \\ \text{LA_offaxis}(issue_3) & \end{array}$$

Where mri , $issue_1$, $issue_2$ and $issue_3$ are the identifiers of the ontology data extracted from the annotations. For example, $issue_1$ is associated to the chunk of text “*basal slice is missing*” and represents a concrete individual of an observed *Lack of coverage* (i.e., $issue_1$ is a member of the concept *Lack_coverage*, that is, $Lack_coverage(issue_1)$).

These facts together with the ontology and rules introduced in the previous section will lead via reasoning to new (implicit) facts. For example, using the ontology axioms in Equations (1)-(3) we can infer the new facts $Technical_issue(issue_1)$ and $MRI_Scan(mri)$. The rules in Equations (7)-(9) will also infer new knowledge (e.g., $has_RV_quality_score(mri, 2)$ via rule (8) and $has_issue_in_HLA_view(mri, issue_2)$ via rule (7)) or raise potential warnings with regard to ambiguity/incompleteness (e.g., $IncompleteIssueDefinition(issue_2)$ and $IncompleteIssueDefinition(issue_3)$ via ontology rule (9)).

3 Discussion

In this preliminary study we have set the first axioms and rules to define a structured vocabulary associated to the quality assessment data. Preliminary experiments to automatically extract ontology data from the free-text annotations have also been conducted.

Related Work. There have also been recent efforts in adding a semantic layer to describe the information within a biobank. Andrade et al. [5] envisaged the benefits of using ontologies for querying and searching the information in a biobank and across

biobanks. Muller et al. [6] presents an updated overview of the state of the art and open challenges for the description and interoperability across biobanks where the use of Semantic Web technologies will play a key role. Examples of concrete Semantic Web-based solutions in biobanks can also be found in [7, 8].

Future Work. As immediate future work, we plan to complete the *CMR-QA*, define the necessary ontology rules and finalize the implementation of the techniques to text mine the comments to extract ontology facts. We also aim at integrating *CMR-QA* with other parts of UK Biobank where different ontologies and controlled vocabularies may be used, and with standard medical vocabularies like SNOMED CT. Furthermore, we will perform an extensive evaluation to analyse the correctness of our approach. For example, (automatic) validation will be carried out by comparing the generated automatic scores using rules with those manually assigned by the observers as part of their quality assessment.

Acknowledgements. SEP, SN and SP acknowledge the British Heart Foundation (BHF) for funding the manual analysis to create a cardiovascular magnetic resonance imaging reference standard for the UK Biobank imaging resource in 5,000 CMR scans (PG/14/89/31194, PI Petersen, 6/2015 to 5/2018). SKP, VC and SN were additionally funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at The Oxford University Hospitals Trust at the University of Oxford. EJ and IH were funded by the EC FP7 project Optique, and the EPSRC projects Score!, ED3 and DBOnto. EJ was also funded by the Centre for Scalable Data Access (SIRIUS) and the RCN project BigMed.

References

1. Carapella, V., et al.: Towards the Semantic Enrichment of Free-text Annotation of Image Quality Assessment for UK Biobank Cardiac Cine MRI Scans. In: MICCAI Workshop on Large-scale Annotation of Biomedical data and Expert Label Synthesis (LABELS). (2016)
2. Petersen, S.E., et al.: Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance* **15**(1) (2013) 1–10
3. Schulz-Menger, J., et al.: Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. *Journal of Cardiovascular Magnetic Resonance* **15**(1) (2013) 1–19
4. Fridman Noy, N., et al.: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* **37**(Web-Server-Issue) (2009)
5. Andrade, A.Q., Kreuzthaler, M., Hastings, J., Krestyaninova, M., Schulz, S.: Requirements for semantic biobanks. *Stud Health Technol Inform.* **180** (2012) 569–573
6. Müller, H., et al.: State-of-the-Art and Future Challenges in the Integration of Biobank Catalogues. In: *Smart Health - Open Problems and Future Challenges.* (2015) 261–273
7. Pathak, J., et al.: Applying semantic web technologies for phenome-wide scan using an electronic health record linked Biobank. *J. Biomedical Semantics* **3** (2012) 10
8. Brochhausen, M., et al.: Developing a semantically rich ontology for the biobank-administration domain. *J. Biomedical Semantics* **4** (2013) 23