

Article

An Ontology to Model the International Rules for Multiple Primary Malignant Tumours in Cancer Registration

Nicholas Charles Nicholson *, Francesco Giusti , Manola Bettio, Raquel Negroo Carvalho , Nadya Dimitrova, Tadeusz Dyba, Manuela Flego, Luciana Neamtiiu, Giorgia Randi and Carmen Martos

European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy;

Francesco.Giusti@ec.europa.eu (F.G.); Manola.Bettio@ec.europa.eu (M.B.);

Raquel.Carvalho@ec.europa.eu (R.N.C.); Nadya.Dimitrova@ec.europa.eu (N.D.);

Tadek.Dyba@ec.europa.eu (T.D.); Manuela.Flego@ext.ec.europa.eu (M.F.); Luciana.Neamtiiu@ec.europa.eu (L.N.);

Giorgia.Randi@ec.europa.eu (G.R.); Carmen.Martos@ec.europa.eu (C.M.)

* Correspondence: Nicholas.Nicholson@ec.europa.eu

Featured Application: The ontology provides a standalone application for validating cases of multiple primary tumours against the international rules. The ontology can also be incorporated into the suite of ontologies planned for devolving to the local cancer-registry level the validation process of the variables contributing to the common European-harmonised data set and thereby help accelerate the availability of data for pan-European and inter-regional comparison.



Citation: Nicholson, N.C.; Giusti, F.; Bettio, M.; Negroo Carvalho, R.; Dimitrova, N.; Dyba, T.; Flego, M.; Neamtiiu, L.; Randi, G.; Martos, C. An Ontology to Model the International Rules for Multiple Primary Malignant Tumours in Cancer Registration. *Appl. Sci.* **2021**, *11*, 7233. <https://doi.org/10.3390/app11167233>

Academic Editors:
Catalina Martínez-Costa and Jesualdo Tomás Fernández Breis

Received: 27 May 2021
Accepted: 3 August 2021
Published: 5 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Population-based cancer registry data provide a key epidemiological resource for monitoring cancer in defined populations. Validation of the data variables contributing to a common data set is necessary to remove statistical bias; the process is currently performed centrally. An ontology-based approach promises advantages in devolving the validation process to the registry level but the checks regarding multiple primary tumours have presented a hurdle. This work presents a solution by modelling the international rules for multiple primary cancers in description logic. Topography groupings described in the rules had to be further categorised in order to simplify the axioms. Description logic expressivity was constrained as far as possible for reasons of automatic reasoning performance. The axioms were consistently able to trap all the different types of scenarios signalling violation of the rules. Batch processing of many records were performed using the Web Ontology Language application programme interface. Performance issues were circumvented for large data sets using the software interface to perform the reasoning operations on the basis of the axioms encoded in the ontology. These results remove one remaining hurdle in developing a purely ontology-based solution for performing the European harmonised data-quality checks, with a number of inherent advantages including the formalisation and integration of the validation rules within the domain data model itself.

Keywords: cancer registry; data validation; multiple primary tumours; ontology; machine-based reasoning

1. Introduction

Population-based cancer registries (CRs) play a pivotal role in the surveillance of cancer at population level [1]. They provide the data from which pan-regional and pan-national information systems such as the North American Surveillance, Epidemiology, and End Results Program (SEER [2]) and the European Cancer Information System (ECIS [3]) are populated, as well as for international epidemiological studies on cancer incidence and mortality (Cancer Incidence in Five Continents, CI5 [4]) and survival (Concord [5] and Eurocare [6]).

CRs collect summary case records of all cancer patients from hospitals and clinics within the associated population area. CRs spend considerable effort in ensuring the consistency and accuracy of the data, which are then used for monitoring and surveillance

purposes in the population, as well as for research. CRs often work together on joint research projects targeted to specific research questions on the basis of given trends in the population data. In Europe, most CRs are affiliated to the European Network of Cancer Registries (ENCR).

Tackling cancer remains a political priority in the European Union and the European Commission services have a close working relationship with the ENCR to harmonise data collected from the CRs and derive the indicators (incidence, mortality, and survival) for monitoring and comparing the cancer burden across the EU Member States. The inter-comparison of indicators for different cancer sites and their changing trends over time allows the EU to coordinate activities aimed at reducing disparities and also to identify and promote good practices.

The ENCR comprises over 150 individual cancer registries operating within markedly different healthcare infrastructures and funding mechanisms. Some registries are nationally based whereas others are regional and some even metropolitan. In order to derive a harmonised set of indicators, a common European CR data set has been defined consisting of some 50 variables that have to conform to explicitly defined formats and rules [7]. Currently, the data is collected from the CRs and processed centrally to ensure adherence to the rules prior to being aggregated and made publicly available in the ECIS web application [3]. The centralised data cleaning, which validates the conformance of the individual CR data sets to the common data set rules, adds further delays to the availability of the data and the associated administrative overheads have been compounded by the introduction of the EU's general data protection regulation. Many difficulties could be circumvented by devolving the validation and aggregation tasks to the local CR level.

The question however arises as how to ensure the necessary level of synchronisation across over 150 different entities in a process that is inherently dynamic and often leads to changes in the common data set and the rule dependencies. The development and free distribution of European CR validation software [8] has been a first major milestone towards this objective. However, the data validation rules are complex and moreover are periodically updated, leading to a substantial amount of recoding effort and correspondingly high maintenance costs with a continued reliance on a specific centralised entity for controlling the whole task. Ontologies offer one possible and elegant solution owing to their ability to integrate the entities defined by a data model with their associated semantic relationships. This means that any change to the data model can be kept in synchronisation with the validation rules that derive from the semantic relationships between the variables. Moreover, basing validation software on ontologies developed in the Web Ontology Language (OWL [9]) helps ensure the veracity of the link to the most updated version release of the data model/semantic relations and further serves to loosen the dependence on any given central entity. Other advantages concern the fact that the validation rules can be formalised in description logic (DL) and that the metadata can be linked semantically to the Uniform Resource Identifiers (URIs) of the related OWL classes.

The benefits accruing to CRs from ontologies and semantically linked data have been highlighted by others [10], and ontology-based tools have been deployed for specific CR activities relating to: automatic rule extraction from data sets [11]; information extraction and reasoning from unstructured pathology reports [12]; analysis of disease courses [13]; and inference of clinical stage [14].

Whereas an initial attempt to develop an ontology-based CR-data validation tool has shown promising first results [15], a critical test lies in its ability to handle the whole set of CR validation checks [7]. Although it was demonstrated that the majority of the rules can be modelled with terminological component (TBox) axioms, the checks to validate multiple primary tumours (MPTs) in case histories of patients with more than a single primary tumour are arguably the most challenging to perform from an ontological point of view. MPTs that do not fulfill the explicit conditions described by the international rules on the validity of multiple primary tumours [7,16] must be considered for reporting purposes as a single primary tumour. The consequence of not performing these specific checks can

lead to skewed cancer statistics and provide erroneous comparisons at population level. This paper presents a solution for modelling the MPT checks in an OWL ontology.

1.1. Validation of Health Data

It has been observed that it is not standard practice to validate data following a systematic process prior to using electronic health record (EHR) data for research purposes [17]. Given the potential of research results to influence clinical practice, having the means of validating data or at least of gauging their quality is a critical prerequisite, especially where data has to be linked or compared across multiple sources. Formal validation approaches based on shapes constraint language (SHACL), description logic, or predicate logic with machine learning have been proposed for validation purposes relating to EHRs, knowledge-acquisition systems, and clinical-support systems. Martinez-Costa and Schulz [18] used SHACL ontology design patterns to validate EHR information models and through this approach were able to trap modelling and terminology-binding errors in CIMI (Clinical Information Modelling Initiative) models. Maqbool et al. [19] derived a predictive model using automated knowledge-acquisition techniques from patient records and used it to validate existing guidelines and create a refined clinical knowledge model from which they generated rules for treatment plans taking the oral cavity as an example. Maqbool et al. [20] built on the earlier work of [19] and incorporated Z notation to provide inferenceable mathematical models and formally verify the knowledge-acquisition model for consistency. Afzal et al. [21] applied machine-learning techniques for correcting staging values in EHRs and deriving rules for treatment plans based on decision paths drawn from the acquired knowledge base. Boeker et al. [22] used DL for the basis of a decision-aid tool for staging of mammary gland and colorectal tumours (serving as a blueprint for to other types of tumour) to help overcome the high deviation rates found in tumour-staging practice.

The validation of CR data sets contributing to the European CR common data set is in contrast a more well-defined task due firstly to the structured nature of the data; and secondly, to the already harmonised nature of the variables comprising the data set. The validation then proceeds against a set of rules that have previously been agreed either at international level (such as the rules for multiple primary tumours) or at pan-European level. The work cited in the previous paragraph would be particularly pertinent to the up-stream processes relating to the primary data providers and to the derivation and verification of the rule base itself. At the level of European harmonisation, the rules are the touchstone against which all the data sets are validated. The task is therefore one of modelling the rules reliably in software and verifying that the software applies the rules correctly using a comprehensive set of pre-defined test cases.

1.2. Data Validation Using OWL

With regard to data validation, it could be argued that OWL is not necessarily the most suitable tool; it was never intended as such and moreover is based on the open-world assumption (OWA) that essentially restricts the inferences that can be made on a given data set (even when the data set is known to be complete). The OWA assumes additional information is always potentially available but not yet disclosed. The CR common data set in contrast corresponds to a closed-world system in which all the information regarding a data set is known (at least as far as the validation rules are concerned).

Whereas the OWA does present challenges, they are generally not insurmountable although they do have certain implications on the design of an ontology used for inferring purposes. The advantages of OWL (particularly with reference to its position in the semantic web stack) more than outweigh the disadvantages and provide a convenient framework to facilitate the devolution of the centralised cleaning process to the local CR level. One of the ways of overcoming OWA-related issues is by making use of the pattern of general concept inclusions [23], whereby a complex class is subclassed from an atomic class rather than the more usual contrary mechanism. The pattern is followed extensively in the axioms described in Section 2.4. An ontology designed along these lines marks a de-

parture from the majority of axioms in OWL ontologies that have atomic classes subclassed from complex classes [23] with the consequence that axioms of standard ontologies cannot easily be reutilised for inferencing purposes. Within the Open Biological and Biomedical Ontologies (OBO) Foundry [24], the NCIt has a number of terms relevant to the MPT ontology, but the ontology is structured in such a way that the inferences needed for the multiple-primary-tumour rules cannot be made. In addition, the standard ontologies tend to be quite large to cater for all the wide variety of needs (NCIt has over 2 million axioms) and would impose performance constraints on automatic reasoning.

In contrast to OWL, the RDF data shapes languages SHACL and ShEx (Shape Expressions) do allow the closed world assumption (CWA, in which something not stated to be true is considered false) and are particularly suited to data validation. SHACL has been used with good effect to validate EHRs [18] and ShEx provides a useful tool for a preprocessing step to trap syntactic errors, range errors, and the more basic inter-variable semantic checks within the common CR data sets. However, both ShEx and SHACL suffer limitations when dealing with the more complex types of inter-variable checks that are better suited to OWL.

1.3. Modelling of Multiple Primary Tumour (MPT) Rules

Validation tools for CR common data sets are generally based on traditional programming techniques [8,25,26]. The model works for a centralised approach to data collection and harmonisation but is less appropriate for a federated model in which a more open approach is encouraged to maintaining and developing the data model and validation rules.

The authors are unaware of any other initiatives to model the international rules for multiple primary malignant tumours or indeed any of the other CR validation rules using an ontological approach.

2. Materials and Methods

2.1. Multiple Primary Tumour Validation Rules

A cancer patient may have multiple cancers but the condition for an MPT is that the cancers are independent of each other. For the purpose of reporting cancer incidence rates, tumours that are not independent are only counted once.

The rules determining MPTs depend on topography (location of the tumour) and morphology (form/structure of the tumour), where topographies and morphologies are encoded according to the third edition of the International Classification of Diseases for Oncology (ICD-O-3 [27]).

The process for determining a violation of the rules for MPTs is shown in the flow chart of Figure 1. All tumour permutations for a MPT patient need to be pairwise compared. First of all, the morphology codes of any one tumour pair are compared; if the codes belong to an equivalent haematological morphology group the two tumours are in breach of the rules. If otherwise they belong to an equivalent (non-haematological) morphology group, the topography codes then need to be considered. Violation of the rules occurs in these cases if the two topography codes also belong to an equivalent group. The conditions defining which morphology codes constitute equivalent groups are defined in Table 10 of [7] and partly elaborated in Section 2.4.1.

For the purposes of the MPT validation checks, a tumour can be defined by an entity having exactly one topography code and exactly one morphology code. This is modelled in DL by the equivalence axiom:

$$\text{ICDO3Tumour} \equiv =1 \text{ hasMorphology.ICDO3 Morphology } \sqcap =1 \text{ hasTopography.ICDO3Topography}$$

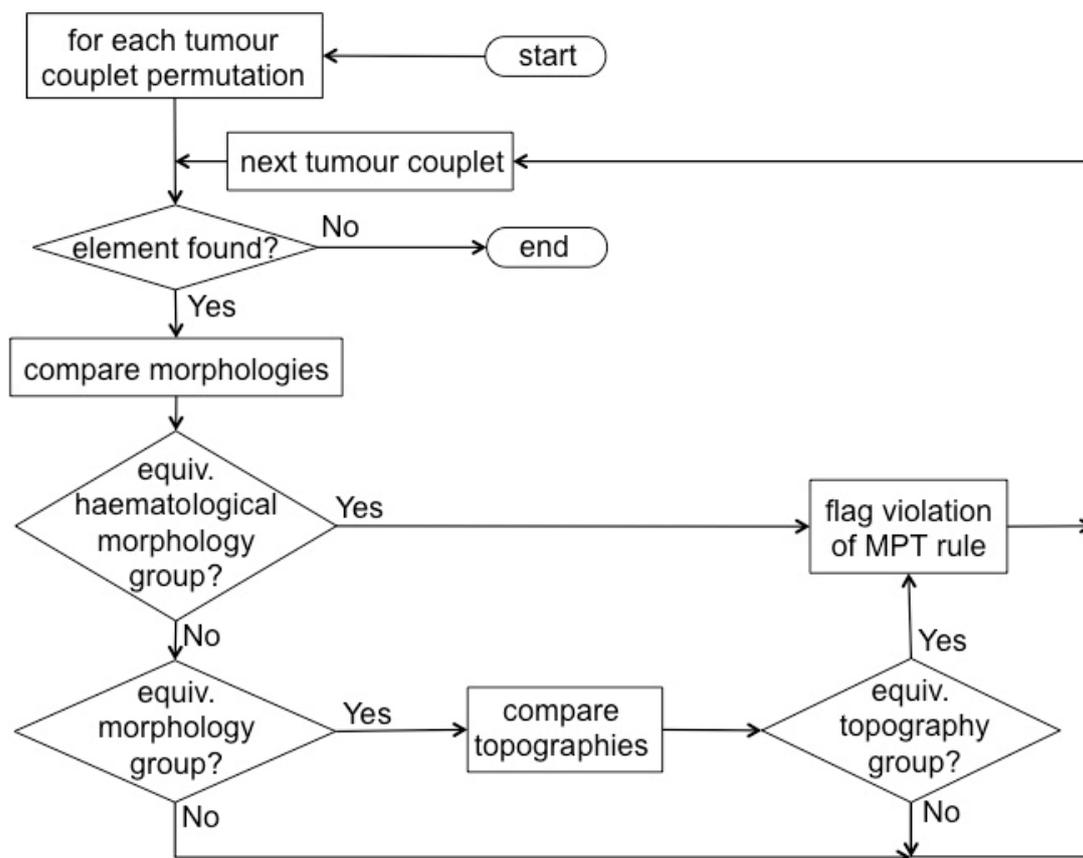


Figure 1. Condensed process for flagging violation of the multiple primary tumour (MPT) rules.

2.2. Morphology Pairing

Axioms to determine whether two morphologies are contained within a single-entity morphological group are modelled on the basis that they are uniquely mapped to only one category (c.f. Table 10 in [7]). There are however three particular categories to be considered common to other morphology categories: one associated with carcinomas; one associated with haematological-type morphologies; and one associated with all other morphology categories. Categories are considered equivalent if either of the morphologies is in a group common to the category of the other morphology.

In order to determine equivalence of morphology groups, axioms need to perform the following three tests:

- Test 1: Determine the single-entity morphology groups following conjunction of the two morphology categories. In case of a single group, infer the morphologies are equivalent by default;
- Test 2: Verify whether the category associated with one of the morphologies is a common one and infer the groups are equivalent if it can be ascertained that the other morphology is in a category associated with that common category. Furthermore, infer that the morphologies are in an equivalent group if either of the morphologies is in the category that is common to all other morphology categories;
- Test 3: For all positive equivalences, ascertain if the two morphologies are of haematological type and then infer that the tumours violate the MPT rules.

2.3. Topography Pairing

The axioms for determining whether any two topographies belong to a single site are more complicated since some codes are duplicated across different groups (c.f. Table 9 in [7]). However, by introducing a further level of categorisation, it may be verified that

codes can be uniquely assigned to a main grouping, and duplicated codes re-assigned to a set of common groups associated with the collective categories of the main groupings. Reformulating the groups in this way allows any duplication of codes to be restricted solely to the common groups, and makes it easier to express the rules in axiomatic form. The original 48 topography groups translate to 49 new groups (1–49) distributed over 19 categories (A–S), where each category has an associated common group of topography codes. The complete redistribution of the groups is provided in Table A1 of Appendix A.

The creation of an extra group (group 49) arises from the need to split the original group 32 containing topography codes C54 and C55. The new group 32 retaining only the code C54 is categorised along with group 31 (containing the code C53) into category N with common codes C55, C578, C579, C763 and the new group 49 (containing code C55 from the old group 32) is categorised with groups 29, 30, and 33 with common codes C578, C579, C763. In this way, code C55 does not need to be duplicated across more than one main group. The tests for ascertaining the equivalence of two topography codes can then proceed along similar lines to Tests 1–2 (Section 2.2) for the morphology codes.

2.4. Ontology Structure

The MPT-validation ontology is based on a modular design and imports several sub-ontologies that are also used for other purposes outside the MPT checks.

The ontology tree is illustrated in Figure 2. At the base are two ontologies comprising the ICD-O-3 classes and updates. Essentially, an ICD-O-3 code includes a topography code; a morphology code; and a behaviour code (describing whether a tumour is benign/uncertain/in situ/unknown).

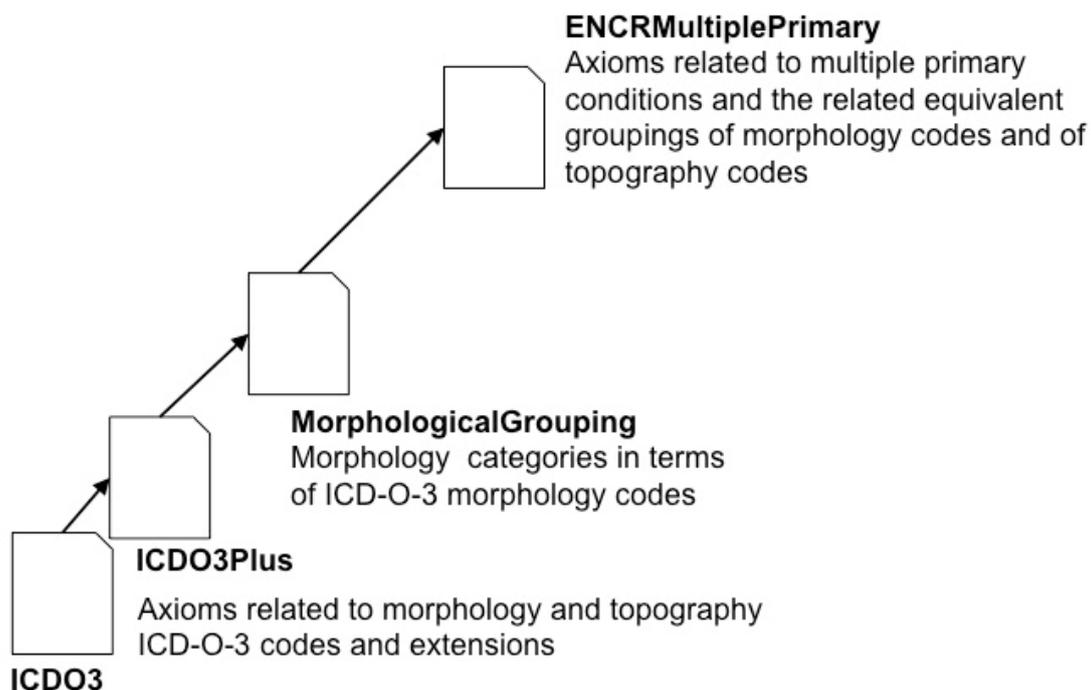


Figure 2. Ontology tree comprising the MPT-validation check.

Within the ICD-O-3 sub-ontologies, topography, morphology, and behaviour are represented by their own OWL classes. Three-digit topographies (e.g., C028) are subclassed from their two-digit counterparts (e.g., C02). Two-digit topographies may be used in the axiom rule-bases but individuals are only created from the three-digit topographies. Morphology classes take the form of the example M_8084 and are subclassed from their three-digit super classes (e.g., M_808) which are in turn subclassed either from a haematological or non-haematological super class. In addition, morphology codes explicitly referring to behaviour take the form M_8084_3 where the trailing digit signifies the behaviour code. The

morphologies with appended behaviour codes are subclassed both from the conjunction of the associated morphology class and behaviour class, e.g.,

$$M_8084_3 \sqsubseteq M_8084 \sqcap \text{Behaviour Code3}$$

The intermediate ontology (MorphologicalGrouping) in Figure 2 provides a categorisation of the morphology codes according largely to those defined in (c.f. Table 10 in [7]), but with a more granular breakdown resulting in separate groups for neuroendocrine tumours, melanomas, thymomas, and sarcoma subgroups (needed for purposes outside the MPT validation checks).

The top-most ontology, ENCRMultipPrimary, contains the axioms for testing the equivalence of the single-entity groups described in Sections 2.2 and 2.3.

2.4.1. Terminological Part (TBox Axioms)

An example of an axiom modelling Test 1 (Section 2.2) using the basal cell carcinoma subclass of the morphological carcinoma category is:

$$\text{Tumour Couplet} \sqcap \geq 2 \text{ hasMorphology.BasalCellCarcinoma} \sqsubseteq \text{DuplicateMorphologyGroup} \quad (1)$$

Axiom (1) essentially states that anything that is both a TumourCouplet and has at least two basal cell carcinoma type morphologies is classified as belonging to a duplicate morphology group. A TumourCouplet is itself expressed as the conjunction of two existential restrictions relating to a permutation of morphology codes and topography codes:

$$\text{TumourCouplet} \equiv \exists \text{hasTumourPermutationMorphology.TumourPermutationMorphology} \sqcap \\ \exists \text{hasTumourPermutationTopography.TumourPermutationTopography}$$

Axiom (1) is replicated for all the different morphology categories, including the common categories.

An example of an axiom for Test 2 is:

$$\exists \text{hasMorphology.UnspecifiedCarcinoma} \sqcap \exists \text{hasMorphology.Adenocarcinoma} \sqsubseteq \text{DuplicateMorphologyGroup}$$

where the concept name UnspecifiedCarcinoma forms a common category with each of the other specific carcinoma morphology categories. This axiom is replicated for all the other specific morphology categories and their common categories. However, the morphology category that is common to all other morphology categories (UnspecifiedCancer) is directly subclassed from DuplicateMorphologyGroup, since the morphologies will be duplicate by default:

$$\exists \text{hasMorphology.UnspecifiedCancer} \sqsubseteq \text{DuplicateMorphologyGroup} \quad (2)$$

Test 3 is handled by the one single axiom:

$$\text{DuplicateMorphologyGroup} \sqcap \exists \text{hasMorphology.ICDO3HaematologicalMorphology} \sqsubseteq \text{DUPLICATE_PRIMARY_TUMOUR} \quad (3)$$

Axiom (3) results in the subsumption under DUPLICATE_PRIMARY_TUMOUR of the conjunction of a previously determined DuplicateMorphologyGroup with a role hasMorphology having at least one filler of ICDO3HaematologicalMorphology.

The axioms for the topography groups are created in a similar manner; however, for Test 2 type conditions (common categories) they generally involve the conjunction of a common category with the set of disjunctions of the main groups associated with the common category, e.g., for the common topography group D:

$$\text{SingleEntityGroupTopographyCommonD} \sqcap$$

$$(\text{SingleEntityGroupTopography09} \sqcup \text{SingleEntityGroupTopography10} \sqcup$$

SingleEntityGroupTopography11) \sqsubseteq DuplicateTopographyGroup

The topography category common across all other categories is directly subclassed from the DuplicateTopographyGroup, in a similar fashion to Axiom (2):

$$\exists \text{hasTopography.TopographyGroupCommonS} \sqsubseteq \text{SingleEntityGroupTopographyCommonForAll} \sqsubseteq \text{DuplicateTopographyGroup} \quad (4)$$

where

$$\exists \text{hasTopography.(C768} \sqcup \text{C80)} \sqsubseteq \exists \text{hasTopography.TopographyGroupCommonS}$$

For the other common topography groups, it can be verified that the following two axioms catch all the cases in which two topography codes fall within a single arbitrary common group:

$$\exists \text{hasTopography.C76} \sqcap \exists \text{hasTopography.(C26} \sqcup \text{C68)} \sqsubseteq \text{DuplicateTopographyGroup}$$

i.e., one tumour with a topography of C76 (other and ill-defined sites) and the other with a topography either of C26 (other and ill-defined digestive organs) or C68 (other and unspecified urinary organs); and

$$\text{TumourCouplet} \sqcap \geq 2 \text{ hasTopography.(C26} \sqcup \text{C68} \sqcup \text{C76)} \sqsubseteq \text{DuplicateTopographyGroup} \quad (5)$$

i.e., both tumours with more than one topography in any of the groups (or subgroups) of C26 or C68 of C76.

Trapping Duplicate Primary Tumours

The cases in which a multiple primary tumour can be ruled out for any given tumour couplet are defined by Axiom (3) for haematological morphologies and Axiom (6) below for non-haematological morphologies:

$$\text{DuplicateMorphologyGroup} \sqcap \text{DuplicateTopographyGroup} \sqsubseteq \text{DUPLICATE_PRIMARY_TUMOUR} \quad (6)$$

Axiom (6) models the rule that tumours are considered dependent if they have topographies and morphologies in equivalent categories.

Tumours with identical morphology or topography codes have to be handled in a slightly different way owing to the fact that the respective tumour permutation results in only one code rather than two. These cases may be trapped using qualified number restrictions. For example, for a tumour couplet with an identical morphology:

$$\text{TumourCouplet} \sqcap =1 \text{ hasMorphology.ICDO3Morphology} \sqsubseteq \text{DuplicateMorphologyGroup} \quad (7)$$

Axiom (7) states that tumour couplets with only one morphology code belong to the duplicate morphology class. However for this axiom to work as intended, the corresponding tumour couplet has to be instanced using a value-restriction expression:

$$\text{TumourCouplet} \sqcap \exists \text{hasMorphology.ICDO3Morphology} \sqcap \forall \text{hasMorphology.}(\exists \text{hasMorphology}^{-}.\{\text{tumourID_X}\})$$

In which the superscript ‘⁻’ inside the value-restriction expression indicates an inverse role. The latter is used to ascertain the morphology code of either one or other of the specific tumour individuals (tumourID_X) forming the tumour couplet. A practical example is provided in Axiom (A1) of Appendix B.

Finally, a patient with multiple tumours can be modelled as an entity with one or more tumour couplets:

$$\text{MultipleTumourPatient} \equiv \geq 1 \text{ hasTumourCouplet}$$

2.4.2. Assertional Part (ABox Axioms)

For each multiple-tumour case, assertional component (ABox) axioms need to be created for all the associated topography and morphology codes; each tumour consisting of a single topography and morphology code; and each tumour pair for all the unique permutations (of which there are $n(n - 1)/2$ for n different tumours). The TBox axioms provide the rule base determining whether the ABox axioms contain duplicate primary cases.

For a hypothetical patient (p_1) with three individual tumours (p_{1_t1} , p_{1_t2} , p_{1_t3}), two of which (p_{1_t2} and p_{1_t3}) have the same morphology code, the three possible unique permutations of tumour couplets are (p_{1_t1} , p_{1_t2}), (p_{1_t1} , p_{1_t3}), and (p_{1_t2} , p_{1_t3}) denoted respectively by p_{1_tc1} , p_{1_tc2} , p_{1_tc3} . The associated ABox statements are provided in Appendix B.

3. Results

Results are presented using the Protégé user interface [28] with automatic reasoning for the different possible scenarios described in Section 2.4 and in reference to the flow diagram of Figure 1 representing the MPT validation-check process. Text with highlighted yellow background in the following figures is information inferred by the reasoner.

Figures 3–6 show the inferences made by the reasoner on the basis of the axioms described in Section 2.4 for the three tumour-couplet permutations p_{1_tc1} , p_{1_tc2} , and p_{1_tc3} .

Figure 3 reveals that one of the associated morphologies is a haematological one whereas the other is not and since the non-haematological morphology is not in the group that is common for all, the morphology groups are considered independent. The two topographies (C150 and C269) are however determined to be equivalent due to topography code C269 being in a group (group C) that is common to the group containing the topography code C150 (group 8). Since the morphology groups are independent, tumours t_1 and t_2 are accepted as multiple primaries.

The screenshot displays the Protégé interface for the tumour couplet p_{1_tc1} . The left pane, titled 'Description: p1_tc1', lists various types. The right pane, titled 'Property assertions: p1_tc1', shows object property assertions for 'hasMorphology' and 'hasTopography'.

Types (Left Pane):

- hasMorphology some (inverse (hasMorphology) value p1_t1)
- hasMorphology some (inverse (hasMorphology) value p1_t2)
- hasTopography some (inverse (hasTopography) value p1_t1)
- hasTopography some (inverse (hasTopography) value p1_t2)
- TumourCouplet
- DuplicateTopographyGroup
- ICDO3HaematologicalMorphology
- ICDO3NonHaematologicalMorphology
- SingleEntityGroupMorphologyCommonCarcinoma
- SingleEntityGroupMorphologyHaemTcellINKcell
- SingleEntityGroupTopography07
- SingleEntityGroupTopographyC26
- SingleEntityGroupTopographyCommonB
- SingleEntityGroupTopographyCommonC
- SingleEntityGroupTopographyCommonD
- SingleEntityGroupTopographyCommonE

Property assertions (Right Pane):

- Object property assertions:
 - hasMorphology m8020_3
 - hasMorphology m9705_3
 - hasTopography c150
 - hasTopography c269
- Data property assertions: (None listed)
- Negative object property assertions: (None listed)
- Negative data property assertions: (None listed)

Figure 3. Results of the reasoning operation for the tumour couplet p_{1_tc1} . Only the topography codes are equivalent (classified under DuplicateTopographyGroup) and the couplet is a valid MPT.

Figure 4 shows that the two morphologies are not in the same group. Even though the topographies are equivalent (since C768 is in the group S that is common to all other topography groups), the two tumours t1 and t3 are accepted as multiple primaries.

Figure 4. Results of the reasoning operation for the tumour couplet p1_tc2. Although the topography codes are equivalent via Axiom (4), the morphology codes are independent and the couplet is a valid MPT.

In Figure 5, the two morphologies are the same and therefore automatically in the same group, and one of the topographies (C768) is in group S that is common to all other topography groups). Thus, both morphologies and both topographies are in duplicate groups and the tumours are considered as a single entity.

Figure 6 shows the result of Axiom (A2), which flags the fact that patient p1 contains at least one invalid MPT.

Figures 7–10 illustrate the results for four other hypothetical patients with multiple tumours showing some of the other possible scenarios. Figure 7 is a valid MPT case (p2_tc1) in which: (a) the two topography codes are both subclasses of the topography code C26 that is contained only within one or more common topography groups and therefore considered equivalent, c.f. Axiom (5); (b) the two morphology codes are both subclasses of the carcinoma group of morphologies but are also in different subgroups of carcinomas (and therefore not equivalent).

Figure 8 is a valid MPT case (p3_tc1) in which: (a) the two topography codes are subclassed from C39, which from Table A1 is considered a duplicate in both a main group (group 22), and a common group (group H); (b) the two morphology codes are independent since one is of haematological type but the other not.

Figure 9 is an invalid MPT case (p4_tc1) in which: (a) the two topography codes are C398 and C761 which are considered single entities from the perspective of a main group (group 22) and its common group (group I), or through a duplication in a common group (group H); (b) the two morphology codes are contained within the same non-haematological main group (adenocarcinoma) and therefore considered a single entity.

The screenshot displays a reasoning tool interface for the tumour couplet `p1_tc3`. It is divided into two main sections: 'Description: p1_tc3' on the left and 'Property assertions: p1_tc3' on the right.

Description: p1_tc3

- Types:**
 - `hasMorphology only (inverse (hasMorphology) value p1_t2)`
 - `hasMorphology some ICDO3Morphology`
 - `hasTopography some (inverse (hasTopography) value p1_t2)`
 - `hasTopography some (inverse (hasTopography) value p1_t3)`
 - `TumourCouplet`
 - `DUPLICATE_PRIMARY_TUMOUR` (highlighted)
 - `DuplicateMorphologyGroup`
 - `ICDO3NonHaematologicalMorphology`
 - `SingleEntityGroupMorphologyCommonCarcinoma`
 - `SingleEntityGroupTopographyC26`
 - `SingleEntityGroupTopographyCommonB`
 - `SingleEntityGroupTopographyCommonC`
 - `SingleEntityGroupTopographyCommonD`
 - `SingleEntityGroupTopographyCommonE`
 - `SingleEntityGroupTopographyCommonForAll`
 - `SingleEntityGroupTopographyCommonL`
 - `SingleEntityGroupTopographyCommonQ`
 - `SingleEntityGroupTopographyCommonR`

Property assertions: p1_tc3

- Object property assertions:**
 - `hasMorphology m8020_3`
 - `hasTopography c768`
 - `hasTopography c269`
- Data property assertions:** (empty)
- Negative object property assertions:** (empty)
- Negative data property assertions:** (empty)

Figure 5. Results of the reasoning operation for the tumour couplet `p1_tc3`. Since the topography codes are equivalent and the morphology codes (of non-haematological type) are identical, the couplet is an invalid MPT and is classified as a duplicate primary tumour.

The screenshot displays a reasoning tool interface for patient `p1`. It is divided into two main sections: 'Description: p1' on the left and 'Property assertions: p1' on the right.

Description: p1

- Types:**
 - `ENCRPatient`
 - `DUPLICATE_PRIMARY_CASE` (highlighted)
 - `MultipleTumourPatient`

Property assertions: p1

- Object property assertions:**
 - `hasTumourCouplet p1_tc3`
 - `hasTumourCouplet p1_tc1`
 - `hasTumourCouplet p1_tc2`

Figure 6. Application of Axiom (A2) alerts the user (via the classification of the patient under `DUPLICATE_PRIMARY_CASE`) to the fact that patient `p1` contains at least one tumour couplet in violation of the MPT rules.

Finally, Figure 10 is an invalid MPT case (`p5_tc1`) in which: (a) the two topography codes are in different main groups and therefore independent; (b) the two morphology groups are contained within the same main haematological group (T-cell and NK-cell neoplasms) which will result in a duplicate primary tumour regardless of the independence between the topography groups.

Description: p2_tc1 ? || ≡ □ ×

Property assertions: p2_tc1 ? || ≡ □ ×

Types +

- **hasMorphology some (inverse (hasMorphology) value p2_t1)** ? @ × ○
- **hasMorphology some (inverse (hasMorphology) value p2_t2)** ? @ × ○
- **hasTopography some (inverse (hasTopography) value p2_t1)** ? @ × ○
- **hasTopography some (inverse (hasTopography) value p2_t2)** ? @ × ○
- **TumourCouplet** ? @ × ○
- DuplicateTopographyGroup ? @
- ICDO3NonHaematologicalMorphology ? @
- SingleEntityGroupMorphologyAdenocarcinoma ? @
- SingleEntityGroupMorphologySquamous ? @
- SingleEntityGroupTopographyC26 ? @
- SingleEntityGroupTopographyCommonB ? @
- SingleEntityGroupTopographyCommonC ? @
- SingleEntityGroupTopographyCommonD ? @
- SingleEntityGroupTopographyCommonE ? @

Object property assertions +

- hasMorphology m8140_3 ? @
- hasMorphology m8051_3 ? @
- hasTopography c260 ? @
- hasTopography c269 ? @

Data property assertions +

Negative object property assertions +

Negative data property assertions +

Figure 7. Results of the reasoning operation for the tumour couplet p2_tc1. Only the topography codes are equivalent and the couplet is a valid MPT.

Description: p3_tc1 ? || ≡ □ ×

Property assertions: p3_tc1 ? || ≡ □ ×

Types +

- **hasMorphology some (inverse (hasMorphology) value p3_t1)** ? @ × ○
- **hasMorphology some (inverse (hasMorphology) value p3_t2)** ? @ × ○
- **hasTopography some (inverse (hasTopography) value p3_t1)** ? @ × ○
- **hasTopography some (inverse (hasTopography) value p3_t2)** ? @ × ○
- **TumourCouplet** ? @ × ○
- DuplicateTopographyGroup ? @
- ICDO3HaematologicalMorphology ? @
- ICDO3NonHaematologicalMorphology ? @
- SingleEntityGroupMorphologyHaemBcell ? @
- SingleEntityGroupMorphologyOtherSpecifiedCancer ? @
- SingleEntityGroupTopography22 ? @
- SingleEntityGroupTopographyCommonG ? @
- SingleEntityGroupTopographyCommonH ? @
- SingleEntityGroupTopographyCommonK ? @

Object property assertions +

- hasMorphology m9365_3 ? @
- hasMorphology m9823_3 ? @
- hasTopography c390 ? @
- hasTopography c398 ? @

Data property assertions +

Negative object property assertions +

Negative data property assertions +

Figure 8. Results of the reasoning operation for the tumour couplet p3_tc1. Only topography codes are equivalent and therefore the couplet is a valid MPT.

Description: p4_tc1 ? ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏

Property assertions: p4_t ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏

Types +

- **hasMorphology** some (**inverse** (hasMorphology) value p4_t1) ? @ × ○
- **hasMorphology** some (**inverse** (hasMorphology) value p4_t2) ? @ × ○
- **hasTopography** some (**inverse** (hasTopography) value p4_t1) ? @ × ○
- **hasTopography** some (**inverse** (hasTopography) value p4_t2) ? @ × ○
- **TumourCouplet** ? @ × ○
- **DUPLICATE_PRIMARY_TUMOUR** ? @
- **DuplicateMorphologyGroup** ? @
- **DuplicateTopographyGroup** ? @
- **ICDO3NonHaematologicalMorphology** ? @
- **SingleEntityGroupMorphologyAdenocarcinoma** ? @
- **SingleEntityGroupTopography22** ? @
- **SingleEntityGroupTopographyCommonB** ? @
- **SingleEntityGroupTopographyCommonE** ? @
- **SingleEntityGroupTopographyCommonG** ? @
- **SingleEntityGroupTopographyCommonH** ? @
- **SingleEntityGroupTopographyCommonI** ? @
- **SingleEntityGroupTopographyCommonJ** ? @
- **SingleEntityGroupTopographyCommonK** ? @
- **SingleEntityGroupTopographyCommonL** ? @
- **SingleEntityGroupTopographyCommonQ** ? @
- **SingleEntityGroupTopographyCommonR** ? @

Object property assertions +

- **hasMorphology** m8260_3 ? @
- **hasMorphology** m8140_3 ? @
- **hasTopography** c398 ? @
- **hasTopography** c761 ? @

Data property assertions +

Negative object property assertions +

Negative data property assertions +

Figure 9. Results of the reasoning operation for the tumour couplet p4_tc1. Both topography codes and the non-haematological morphology codes are equivalent and the couplet is an invalid MPT.

Description: p5_tc1 ? ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏

Property assertions: p5_t ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏ ⏏

Types +

- **hasMorphology** some (**inverse** (hasMorphology) value p5_t1) ? @ × ○
- **hasMorphology** some (**inverse** (hasMorphology) value p5_t2) ? @ × ○
- **hasTopography** some (**inverse** (hasTopography) value p5_t1) ? @ × ○
- **hasTopography** some (**inverse** (hasTopography) value p5_t2) ? @ × ○
- **TumourCouplet** ? @ × ○
- **DUPLICATE_PRIMARY_TUMOUR** ? @
- **DuplicateMorphologyGroup** ? @
- **ICDO3HaematologicalMorphology** ? @
- **SingleEntityGroupMorphologyHaemTcellNKcell** ? @
- **SingleEntityGroupTopography11** ? @
- **SingleEntityGroupTopography13** ? @

Object property assertions +

- **hasMorphology** m9724_3 ? @
- **hasMorphology** m9714_3 ? @
- **hasTopography** c209 ? @
- **hasTopography** c220 ? @

Data property assertions +

Negative object property assertions +

Negative data property assertions +

Figure 10. Results of the reasoning operation for the tumour couplet p5_tc1. Only the morphology codes are equivalent, but since they are haematological-type morphologies the tumour couplet is an invalid MPT independent of the topography codes.

3.1. Performance

In terms of performance, the tests took 7 s to complete on a 3 GHz Intel Core i7 processor using the FaCT++ reasoner and 45 s with the HermiT reasoner. The FaCT reasoner however sometimes exited with inconsistent-ontology diagnostic messages that could only be resolved by removing all disjoint axiom statements from the ontology (without compromise of performance/functionality). The entire ontology consisted of approximately 10,000 axioms (including 2800 classes, 5000 subclasses, 660 general concept inclusions, and 110 individuals).

3.2. DL Query Interface

The results shown in Figures 3–10 may not be so intuitive to users not directly involved in cancer registration, but Protégé also includes a DL query interface for framing specific queries on the reasoned knowledge base. In addition, with the OWL application programme interface (OWL-API [29]), it is possible to tailor the output of the reasoner (or even swap out the reasoner with programme logic) along the lines shown in Figure A7 of Appendix C and discussed further in Section 4.

Figures 11–17 show four examples of DL queries with the reasoner-provided explanations for the inferences. They may be crosschecked with the outputs illustrated in Figure A7. The DL query in Figure 11 is to find all entities having at least one duplicate primary tumour, c.f. Axioms (3) and (6). Three patients were found corresponding to p1, p4, and p5.

The screenshot shows a web-based interface for a DL query. At the top, there is a yellow header with the text 'DL query:'. Below this, a grey box contains the query 'hasTumourCouplet some DUPLICATE_PRIMARY_TUMOUR'. There are two buttons: 'Execute' and 'Add to ontology'. Below the query box, there is a section titled 'Query results'. It shows a list of instances: 'p1', 'p4', and 'p5'. Each instance is preceded by a purple diamond icon. To the right of the list, there is a 'Query for' section with several checkboxes: 'Direct superclasses', 'Superclasses', 'Equivalent classes', 'Direct subclasses', 'Subclasses', and 'Instances'. The 'Instances' checkbox is checked.

Figure 11. Results of a DL query for entities that have duplicate primary tumours.

Figure 12 shows the explanation provided by the reasoner for the case of patient p1 (which is in effect the situation depicted in Figure 5), where it has been found that:

- one of the topography codes C768 in the tumour couplet is in the group common to all other topography groups and therefore a duplicate topography group condition, c.f. Axiom (4) and lines 5, 15, and 1 in Figure 12;
- the morphology codes in the tumour couplet are identical since there is only one morphology code and therefore a duplicate morphology condition, c.f. Axiom (7) and lines 2, 14, and 13 in Figure 12;
- the tumour couplet p1_tc3 is a duplicate tumour since it has both a duplicate topography condition as well as a duplicate morphology condition, c.f. Axiom (6) and line 7 in Figure 12;
- Finally, since p1 has a tumour couplet p1_tc3 (line 9 in Figure 12) and the tumour couplet is a duplicate primary tumour, p1 is one of the entities found by the DL query specified in Figure 11.

Explanation for: p1 Type hasTumourCouplet some DUPLICATE_PRIMARY_TUMOUR

- 1) SingleEntityGroupTopographyCommonForAll **SubClassOf** DuplicateTopographyGroup
- 2) p1_tc3 **Type** TumourCouplet
- 3) p1_tc3 **Type** hasTopography **some** (**inverse** (hasTopography) **value** p1_t3)
- 4) c768 **Type** C768
- 5) hasTopography **some** C768 **SubClassOf** hasTopography **some** TopographyGroupCommonS
- 6) ICDO3Tumour **EquivalentTo** (hasMorphology **exactly** 1 ICDO3Morphology) **and** (hasTopography **exactly** 1 ICDO3Topography)
- 7) DuplicateMorphologyGroup **and** DuplicateTopographyGroup **SubClassOf** DUPLICATE_PRIMARY_TUMOUR
- 8) p1_tc3 hasTopography c768
- 9) p1 hasTumourCouplet p1_tc3
- 10) p1_t2 **Type** ICDO3Tumour
- 11) p1_tc3 **Type** ICDO3Tumour
- 12) hasTopography **Range** ICDO3Topography
- 13) TumourCouplet **and** (hasMorphology **max** 1 ICDO3Morphology) **SubClassOf** DuplicateMorphologyGroup
- 14) p1_tc3 **Type** hasMorphology **only** (**inverse** (hasMorphology) **value** p1_t2)
- 15) hasTopography **some** TopographyGroupCommonS **SubClassOf** SingleEntityGroupTopographyCommonForAll

Figure 12. Explanation provided by the reasoner concerning why patient p1 has a tumour couplet that is a duplicate primary tumour.

The DL query in Figure 13 is to ascertain entities with haematological morphologies in a common group (which should automatically flag a duplicate primary tumour, c.f. Figure 1). The reasoner-provided explanation for returning the result of the tumour couplet p5_tc1 is shown in Figure 14.

DL query:

Query (class expression)

hasMorphology **some** ICDO3HaematologicalMorphology **and** DuplicateMorphologyGroup

Execute
Add to ontology

Query results

Instances (1 of 1)

◆ p5_tc1

Query for

Direct superclasses

Superclasses

Equivalent classes

Direct subclasses

Subclasses

Instances

Figure 13. DL query for tumour couplets with haematological morphologies in a common group. The reasoner has found one such case—p5_tc1, with the explanation provided in Figure 14.

Explanation for: p5_tc1 Type DuplicateMorphologyGroup and (hasMorphology some ICDO3HaematologicalMorphology)

- 1) p5_t2 hasMorphology m9714_3
- 2) M_9714 SubClassOf M_971
- 3) hasMorphology Range ICDO3Morphology
- 4) M_9724_3 SubClassOf M_9724
- 5) m9724_3 Type M_9724_3
- 6) TumourCouplet and (hasMorphology min 2 TcellNKcellNeoplasm) SubClassOf DuplicateMorphologyGroup
- 7) ICDO3Tumour EquivalentTo (hasMorphology exactly 1 ICDO3Morphology) and (hasTopography exactly 1 ICDO3Topography)
- 8) hasMorphology some M_9714 SubClassOf hasMorphology some TcellNKcellNeoplasm
- 9) M959–M972 SubClassOf ICDO3HaematologicalMorphology
- 11) M_971 SubClassOf M967–M972
- 12) M967–M972 SubClassOf M959–M972
- 13) p5_tc1 Type TumourCouplet
- 14) M_9714_3 SubClassOf M_9714
- 15) p5_t1 hasMorphology m9724_3
- 16) hasMorphology some M_9724 SubClassOf hasMorphology some TcellNKcellNeoplasm
- 17) m9714_3 Type M_9714_3
- 18) p5_tc1 Type hasMorphology some (inverse (hasMorphology) value p5_t2)
- 19) p5_tc1 Type hasMorphology some (inverse (hasMorphology) value p5_t1)
- 20) p5_t2 Type ICDO3Tumour
- 21) p5_t1 Type ICDO3Tumour

Figure 14. Explanation provided by the reasoner for the result of the DL query in Figure 13. Lines 8 and 16 reveal that the two morphologies in the couplet belong to the TcellNKcellNeoplasm category and therefore trigger the duplicate morphology group condition (line 6). Line 9 reveals that both morphologies are classified as haematological morphologies. This set of conditions serves to flag the duplicate primary tumour condition of p5_tc1 in the related Figure 10, c.f. Axiom (3).

Figure 15 is DL query for all entities having both a duplicate morphology group condition and a duplicate topography group condition (one of the other triggers for a primary multiple tumour, c.f. Figure 1). The reasoner found the two tumour couplets: p1_tc3 and p4_tc1 (c.f. Figures 5 and 9, and Figure A7). The explanation behind p1_tc3 is provided in Figure 16.

DL query: []

Query (class expression)

DuplicateMorphologyGroup and DuplicateTopographyGroup

Execute
Add to ontology

Query results

Instances (2 of 2)

- p1_tc3 ?
- p4_tc1 ?

Query for

- Direct superclasses
- Superclasses
- Equivalent classes
- Direct subclasses
- Subclasses
- Instances

Figure 15. DL query for entities with both a duplicate topography and a duplicate morphology condition (one of the triggers for a multiple primary tumour).

Explanation for: p1_tc3 Type DuplicateMorphologyGroup and DuplicateTopographyGroup

- 1) c269 **Type** C269
- 2) p1_tc3 **Type** TumourCouplet
- 3) (hasTopography **some** C76) **and** (hasTopography **some** (C26 **or** C68)) **SubClassOf** DuplicateTopographyGroup
- 4) p1_tc3 **Type** hasTopography **some** (**inverse** (hasTopography) **value** p1_t2)
- 5) p1_tc3 **Type** hasTopography **some** (**inverse** (hasTopography) **value** p1_t3)
- 6) c768 **Type** C768
- 7) C768 **SubClassOf** C76
- 8) ICDO3Tumour **EquivalentTo** (hasMorphology **exactly** 1 ICDO3Morphology) **and** (hasTopography **exactly** 1 ICDO3Topography)
- 9) p1_t3 hasTopography c768
- 10) p1_t2 hasTopography c269
- 11) p1_t2 **Type** ICDO3Tumour
- 12) p1_t3 **Type** ICDO3Tumour
- 13) C269 **SubClassOf** C26
- 14) hasTopography **Range** ICDO3Topography
- 15) TumourCouplet **and** (hasMorphology **max** 1 ICDO3Morphology) **SubClassOf** DuplicateMorphologyGroup
- 16) p1_tc3 **Type** hasMorphology **only** (**inverse** (hasMorphology) **value** p1_t2)

Figure 16. Explanation behind one of the results (p1_tc3) in Figure 15.

In Figure 16, the reasoner has ascertained in lines 16 and 15 that there is only one morphology (i.e., the morphology codes are the same in both tumours of the tumour couplet). In lines 1, 13, 6, 7, and 3 the reasoner has ascertained that the conjunction of topographies C269 and C768 automatically triggers a duplicate topography group condition, c.f. Axiom (5).

The final example in Figure 17 shows the results of a DL query on all entities having exactly one morphology. The reasoner has returned all the individual tumour cases (p1_t1–p5_t2) since these by definition have only one morphology and it has also returned the tumour couplet p1_tc3. This tumour couplet has two identical morphologies, c.f. Axiom (A1) in Appendix B and Figure 5.

DL query: ⏏ ⏏ ⏏

Query (class expression)

hasMorphology **exactly** 1 ICDO3Morphology

Execute Add to ontology

Query results

Instances (12 of 12)

- ◆ p1_t1 ?
- ◆ p1_t2 ?
- ◆ p1_t3 ?
- ◆ **p1_tc3** ?
- ◆ p2_t1 ?
- ◆ p2_t2 ?
- ◆ p3_t1 ?
- ◆ p3_t2 ?
- ◆ p4_t1 ?
- ◆ p4_t2 ?
- ◆ p5_t1 ?
- ◆ p5_t2 ?

Query for

Direct superclasses

Superclasses

Equivalent classes

Direct subclasses

Subclasses

Instances

Result filters

Name contains

Figure 17. DL query for entities with exactly one morphology code.

4. Discussion

The results shown in Section 3 are intended only as a proof of concept to illustrate how the axioms of the MPT ontology are able to trap the different scenarios that lead to violation of the international rules for multiple primary malignant tumours. Whereas they are not meant as an exhaustive set of checks, they do illustrate all the mechanisms whereby a multiple primary tumour condition is ascertained. The work was undertaken primarily to understand whether the MPT validation checks could be handled in an ontology. The next step will be to perform a full field-test evaluation using standard ENCR test case records and thereafter to appraise its performance against the full set of European CR common data sets. As noted earlier, OWL was never intended as a validation tool and a language such as Datalog could undoubtedly offer greater performance. The driving motivation behind an ontology is the ability to integrate the validation rules (which form part of the semantic context) into the data model description and to have the means of maintaining the data model whilst keeping the data-validation rules (and all the down-stream processes) in synchronisation. The ultimate goal is to federate the whole data-validation process of the European CR common data set to the local CR level.

OWL provides a key to accomplishing such a goal. It not only integrates the semantic context as axioms into the data model but also provides automatic reasoning based on those axioms. In addition, it provides the possibility of building up a larger more comprehensive ontology from a set of smaller ones via the ontology import mechanism and consequently allows the reuse of base ontologies, which in turn simplifies maintenance and further development. Moreover, it forms parts of the semantic web stack and builds on RDF and linked open data principles thereby facilitating federation of data processes. Users themselves can access the OWL class hierarchy and axioms via editors such as Protégé, which provide a friendly user interface with which to explore and understand the data model from various perspectives. Via the DL query interface, users can verify the accuracy of the inferences made by the reasoner and use it interactively to test modifications or additional axioms.

Whereas it has been demonstrated that the MPT rules can in principle be modelled in description logic and consequently benefit from automated reasoning techniques, the expressivity required exacts a cost in performance. Although this is not so much an issue for the verification of several test cases it becomes prohibitive for the validation of many tens of thousands of records typical in a historical CR data set.

Formulating the axioms in the ways described in Sections 2.4.1 and 2.4.2 equates to a DL expressivity described by ALCOIQ (attributive language with complex concept negation, nominals, inverse properties, and qualified cardinality restrictions). It has been shown for concept satisfiability that the complexity of ALCOIQ is both NEXPTIME-complete and NEXPTIME-hard [30]. This compares with PSPACE-completeness (for acyclic TBoxes) for expressivities: ALC, ALCI, ALCQ, and ALCIQ. Owing to the need of having to deal with tumour permutations, it is not possible to remove the dependence on nominals/individuals. Whereas suitably optimised Tableau-algorithm based reasoners have facilitated increasingly expressive DLs, there are limitations for real-time applications and care has to be exercised on the types and numbers of axiom constructs added to the ontology if the automatic reasoning function is to be invoked.

Performance issues however can be circumvented without having to sacrifice all the other advantages offered by ontologies outside their automatic reasoning capacities. OWL, for example, has the application programme interface (OWL-API [29]) allowing the development of dedicated application software to interface to an ontology. Through such an interface it is possible to have direct access to the ontology axioms facilitating the developing of application software in an entirely different way than that used in the traditional sense since the ontology remains the formal basis of all the underlying axioms and entity relationships. The task of the interface programme is solely to glean the necessary information from the ontology relevant for a given computational task. The programming

effort is thereby considerably reduced and the knowledge resides exclusively in the axioms of the ontology.

Appendix C describes the programme operations that can be implemented to replace the automatic reasoning. The same test cases described in Section 3 were instantaneous using the OWL-API programme without invoking the reasoner, and a thousand such tests took 15 s (non-optimised code and writing to standard output).

Expressivity of OWL DLs can also be increased using the Semantic Web Rule Language (SWRL). SWRL provides a set of further logic functions but comes at a potential cost of decidability and interoperability [31,32]. A particular benefit of SWRL is the functionality it provides with the date, time, and duration built-ins. In the development of the MPT ontology, we did not come up against the need to incorporate SWRL rules and one of the design considerations for maintenance purposes was to keep the ontology as simple as possible.

In view of these results and the advantages afforded by semantic web technologies, there are few practical reasons against migrating the existing European CR quality-check software to an ontology-based tool-kit. A front-end tailored to users' needs would hide the internal mechanisms for running the checks, but the ontology would provide the direct interface for polling the knowledge base and for determining the inter-variable dependencies. It would thereby constitute a machine-readable rendering of the validation rules and remove the need for maintaining a separate set of validation tables as is currently the case.

5. Conclusions

It has been demonstrated that the international rules for MPTs can be modelled comprehensively in an OWL ontology with DL expressivity ALCOIQ.

Since the MPT rules are the most complex of the data-quality checks from an ontological point of view these results confirm the applicability of OWL for a basis of European CR data-validation checks.

OWL provides many advantages to help support the devolution of the CR data-validation process and motivate the eventual federation of CR data sources. Firstly and foremost, it provides the means of formally representing the validation rules using description logic, thereby removing much of the ambiguity in natural language representation. Secondly, it removes the need for a dedicated data-validation software application, which in turn reduces development and maintenance costs. Thirdly, it unifies the validation rules with the underlying data model and provides automatic synchronisation when modifications are made. Fourthly, it provides a unique address to the ontology via the ontology URI that avoids the sorts of versioning issues experienced with distributed software. Finally, it provides an automatic reasoning functionality for DL querying of the knowledge base.

In cases where the use of machine reasoning leads to performance overheads for high DL expressivities, the OWL application programme interface can be used to develop application software. Software developed in this way has direct access to the ontology axioms and can be used to emulate aspects of the reasoner at a fraction of the effort of developing dedicated software.

The results remove one of the remaining hurdles to developing CR validation software based entirely on an ontology and should therefore facilitate the devolution of the data-validation process to the local level. This in turn will serve to federate harmonised CR data sets and accelerate their availability for research and healthcare-policy purposes.

The next step will be to integrate the MPT ontology with those addressing other aspects of the entire set of validation rules [7] and to benchmark the performance of the ontology suite with the existing ENCR quality-check software. The aim will be to replace the latter with an ontology-based tool that can be more easily and openly maintained and developed by the user base and that can support the devolution of the data quality-checking process.

Author Contributions: Conceptualisation, N.C.N.; Data curation, M.F. and G.R.; Formal analysis, N.C.N.; Investigation, T.D. and L.N.; Methodology, N.C.N., F.G. and C.M.; Project administration, M.B.; Resources, N.D.; Software, N.C.N.; Supervision, C.M.; Validation, R.N.C.; Writing—original draft, N.C.N.; Writing—review & editing, F.G., M.B., R.N.C., N.D., T.D., M.F., L.N., G.R. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: All authors are employed by governmental or supranational entities and report no additional funding for the development of this manuscript.

Data Availability Statement: The ontologies are available at: <http://data.europa.eu/89h/2a110a2e-d1e2-439d-9b9f-7e2e5436cc49> (accessed on 4 August 2020).

Acknowledgments: This work was partly conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ABox	Assertional part of a knowledge base; a set of concepts and role assertions.
ALCOIQ	A specific DL expressivity incorporating: attributive language with complex concept negation, nominals, inverse properties, and qualified cardinality restrictions.
CR	Cancer Registry; an entity that collects and analyses cancer cases.
CWA	Closed World Assumption; according to which something not known to be true is false.
DL	Description Logic; a structured knowledge-representation language.
HER	Electronic Health Record.
ENCR	European Network of Cancer Registries.
EU	European Union.
ICD-O-3	International Classification of Diseases for Oncology, 3rd edition; provides a set of codes for describing tumours.
MPT	Multiple Primary Tumour; describes CR patients having more than one independent primary tumour.
NCIt	NCIthesaurus.
NEXPTIME	Non-deterministic Exponential Time; computational complexity class for solving a particular set of problems.
PSPACE	Polynomial Space; computational complexity class for solving a particular set of problems.
OBO	Open Biological and Biomedical Ontologies Foundry.
OWA	Open World Assumption; according to which something cannot be assumed to be false unless it has explicitly been stated as false.
OWL	Web Ontology Language; a language specification for developing ontologies for the semantic web.
OWL-API	OWL Application Programming Interface; a set of methods for interfacing a computer programme with an OWL ontology.
RDF	Resource Description Framework.
SHACL	Shapes Constraint Language.
ShEx	Shape Expressions.
SWRL	Semantic Web Rule Language.
TBox	Terminology part of a knowledge base; a set of concepts and roles.
URI	Uniform Resource Identifier; a unique identifier of a web resource.

Appendix A

Table A1. Further categorisation of topography codes to circumvent duplication of codes in the main groups. The original 48 groups (indicated in bold font) correspond with the rows taken consecutively in Table 9 of [7].

Category	Original Topography Groups Considered as Single Entities	Common Group of Topography Codes for the Associated Category
A	1 (C00, C03, C04, C05, C06); 2 (C01, C02); 3 (C07); 4 (C08); 5 (C09, C10, C12, C13, C19); 6 (C11); 42 (C69); 43 (C70); 44 (C71); 45 (C72); 46 (C73)	C760
B	7 (C15)	C268, C269, C761
C	8 (C16); 14 (C23, C24)	C268, C269, C762
D	9 (C17); 10 (C18), 11 (C19, C20)	C26, C762
E	12 (C21)	C26, C76
F	13 (C22); 15 (C25)	C268, C762
G	16 (C30); 17 (C31); 18 (C32)	C39, C760
H	19 (C33, C34)	C39, C760, C761
I	22 (C39)	C760, C761
J	20 (C37); 28 (C50)	C761
K	21 (C38)	C398, C761
L	23 (C40, C41); 24 (C44); 25 (C47); 26 (C48); 27 (C49); 47 (C74); 48 (C75)	C76
M	29 (C51); 30 (C52); 33 (C56); 49 (C55)	C578, C579, C763
N	31 (C53); 32 (C54)	C55, C578, C579, C763
O	34 (C57); 35 (C58), 39 (C63)	C763
P	36 (C60); 37 (C61); 38 (C62)	C638, C639, C763
Q	40 (C64)	C688 C689, C76
R	41 (C65, C66, C67)	C68, C76
S	<i>Common to all topography groups</i>	C80, C768

Appendix B

For a hypothetical patient (p1) with three individual tumours (p1_t1, p1_t2, p1_t3), two of which (p1_t2 and p1_t3) have the same morphology code, the three possible unique permutations of tumour couplets are (p1_t1, p1_t2), (p1_t1, p1_t3), and (p1_t2, p1_t3) denoted respectively by p1_tc1, p1_tc2, p1_tc3. The associated ABox statements are:

Topography codes:
 c150:C150
 c269:C269
 c768:C768
 Morphology codes:
 m9705_3:M_9705_3
 m8020_3:M_8020_3
 Tumours:
 p1_t1:ICDO3Tumour
 p1_t2:ICDO3Tumour
 p1_t3:ICDO3Tumour
 (p1_t1,m9705_3):hasMorphology \sqcap
 (p1_t1,c150):hasTopography
 (p1_t2,m8020_3):hasMorphology \sqcap
 (p1_t2,c269):hasTopography
 (p1_t3,m8020_3):hasMorphology \sqcap
 (p1_t3,c768):hasTopography
 Tumour permutations:
 p1_tpM1: TumourPermutationMorphology \sqcap
 \exists hasMorphology.(\exists hasMorphology $^-$.{p1_t1}) \sqcap
 \exists hasMorphology.(\exists hasMorphology $^-$.{p1_t2})
 p1_tpT1: TumourPermutationTopography \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t1}) \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t2})
 p1_tpM2: TumourPermutationMorphology \sqcap
 \exists hasMorphology.(\exists hasMorphology $^-$.{p1_t1}) \sqcap
 \exists hasMorphology.(\exists hasMorphology $^-$.{p1_t3}) (A1)
 p1_tpT2: TumourPermutationTopography \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t1}) \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t3})
 p1_tpM3: TumourPermutationMorphology \sqcap
 \exists hasMorphology.ICDO3Morphology \sqcap
 \forall hasMorphology(\exists hasMorphology $^-$.{p1_t2})
 p1_tpT3: TumourPermutationTopography \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t2}) \sqcap
 \exists hasTopography.(\exists hasTopography $^-$.{p1_t3})
 Tumour couplets:
 p1_tc1:TumourCouplet
 p1_tc2:TumourCouplet
 p1_tc3:TumourCouplet
 (p1_tc1,p1_tpM1):hasTumourPermutationMorphology \sqcap
 (p1_tc1,p1_tpT1):hasTumourPermutationTopography
 (p1_tc2,p1_tpM2):hasTumourPermutationMorphology \sqcap
 (p1_tc2,p1_tpT2):hasTumourPermutationTopography
 (p1_tc3,p1_tpM3):hasTumourPermutationMorphology \sqcap
 (p1_tc3,p1_tpT3):hasTumourPermutationTopography
 Patient:
 p1:ENCRPatient
 (p1, p1_tc1):hasTumourCouplet \sqcap (p1,
 p1_tc2):hasTumourCouplet \sqcap
 (p1, p1_tc3):hasTumourCouplet

The case in which a patient has at least one duplicate tumour can be flagged using the TBox equivalent statement:

$$\text{DUPLICATE_PRIMARY_CASE} \equiv \exists \text{hasTumourCouplet.DUPLICATE_PRIMARY_TUMOUR} \quad (\text{A2})$$

Axiom (A2) provides a convenient means for screening violations of the multiple primary rules at the patient level without having to look individually at all tumour-couplet instances.

Appendix C

The programme operations to replace the automatic reasoning in software for the MPT validation checks are depicted in the flowcharts of Figures A1–A6.

Programme Flowcharts

Figure A1 shows the steps for ascertaining the morphology category of each tumour in a multiple-tumour case. Morphology codes are either uniquely assigned to a particular category or not at all. Thus, it is a matter of cycling through the morphology category axioms to determine the category for a particular code, which is then recorded before the step is repeated for the next tumour. Since the axioms for the morphology categories are expressed in terms of 3-digit or 4-digit morphology codes, if no group has been found from an initial search on the 4-digit codes, the search has to be repeated for the 3 digit codes.

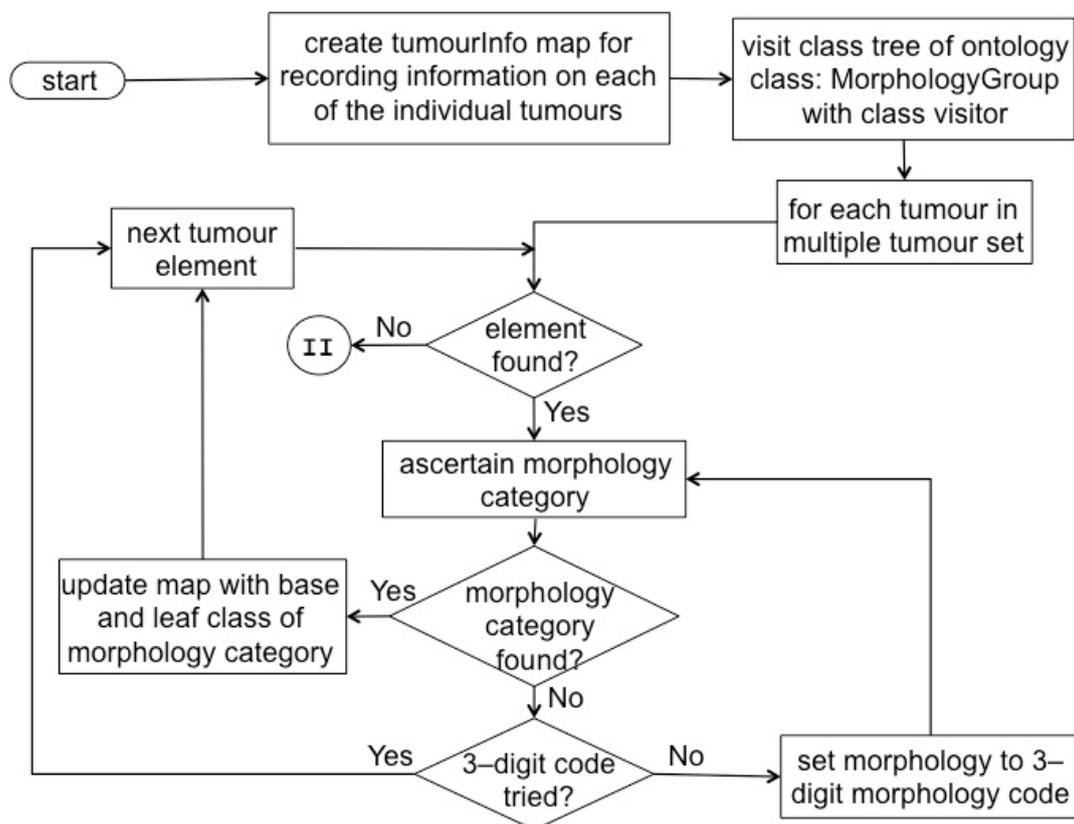


Figure A1. Process for determining the morphology group associated with a morphology code from the ontology axioms.

After processing all the tumours in this way, the next stage (elaborated in Figure A2) is to ascertain in which single-entity morphology group each tumour's morphology category lies. Since the single-entity morphology groups can be described also in terms of the base classes of the morphology categories, both the leaf and root classes of the morphology category need to be searched.

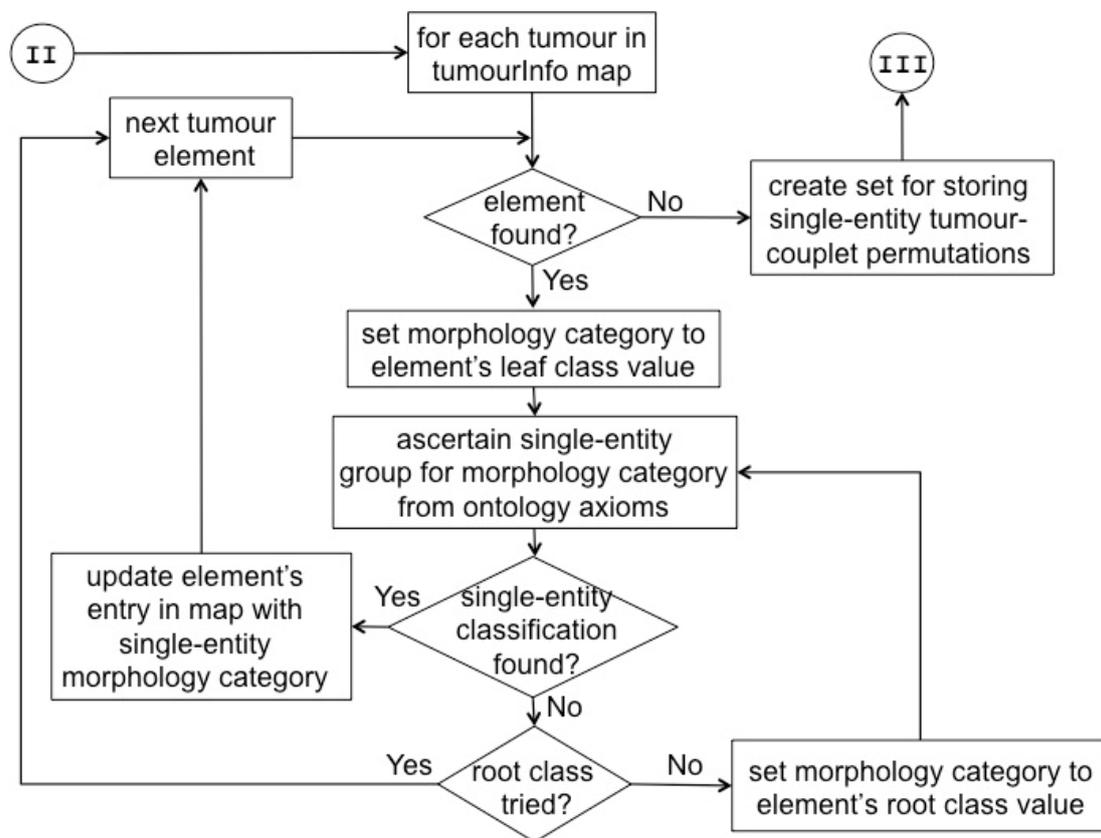


Figure A2. Process for determining from the ontology axioms the single-entity morphology group associated with the morphology category found in Figure A1 from a tumour's morphology code.

The next stage (Figures A3 and A4) is to determine whether the morphology groups are equivalent for each of the tumour-pair permutations. Figure A3 shows the steps for cycling through the tumour-couplet permutations. Figure A4, which is contained in Figure A3 via the reference connector "IV", is a switch-type construct on the names of the two single-entity morphology groups of a tumour pair. The first test determines whether either of the groups is the common one across all groups. The other tests determine whether the two groups are either identical or equivalent via a possible common-group association. Thereafter, it has to be ascertained whether the pair is of haematological morphology type in order to infer whether the permutation can immediately be considered to be a single tumour independent of the value of the topography codes. The information derived from this stage is added to a set holding all tumour pairs characterised by single-entity morphologies.

The final part of the process (Figures A5 and A6) checks the equivalence of the single-entity topography groups of the tumour pairs that were added to the single-entity tumour-couplet set from the previous stage.

Since the ontology axioms for single-entity topography groups may be described in terms of 2-digit or 3-digit topography codes, both sets of codes need to be considered Figure A5.

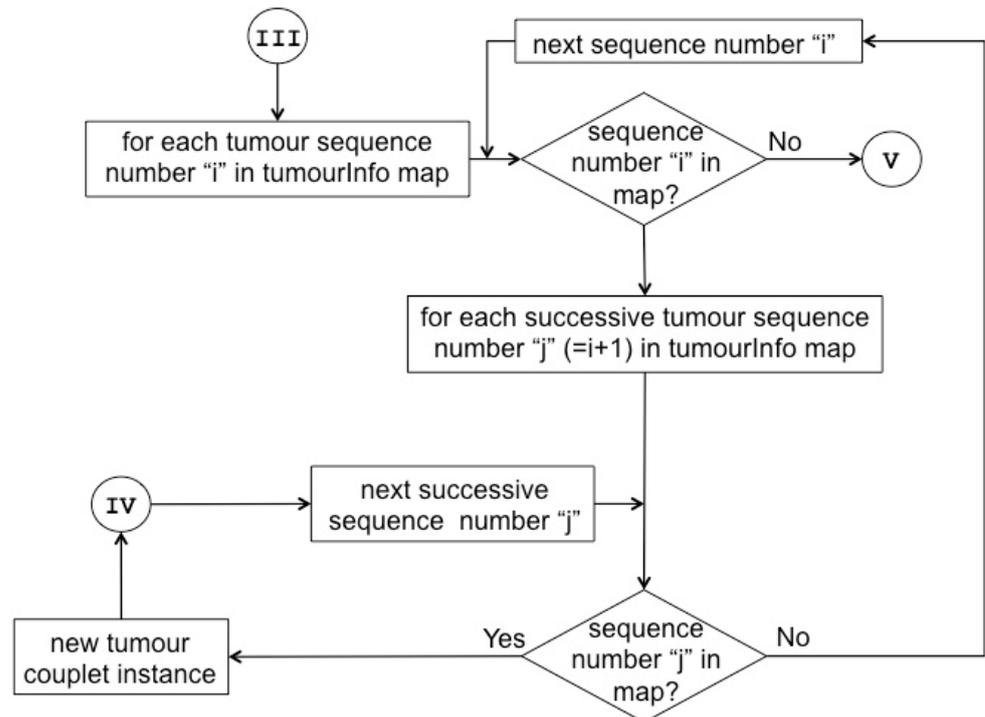


Figure A3. Flowchart for cycling through the tumour-couplet permutations in a case having multiple tumours. The connector labelled “IV” refers to the process described in Figure A4.

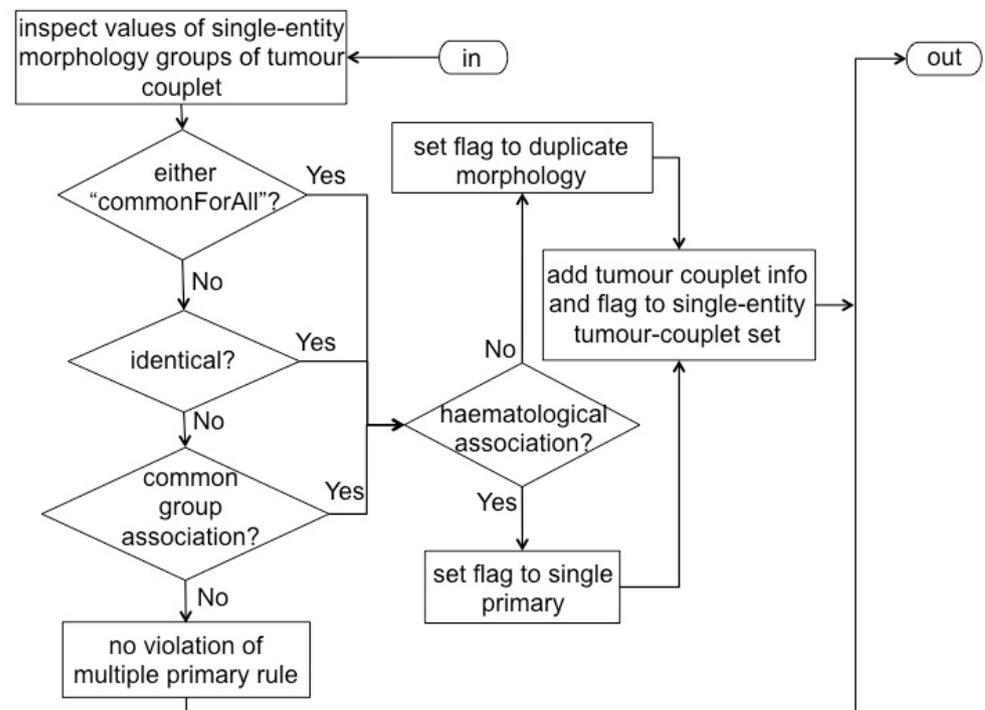


Figure A4. Process (encapsulated in connector “IV” in Figure A3) For determining whether the two single-entity morphology groups are equivalent for a given tumour-pair permutation.

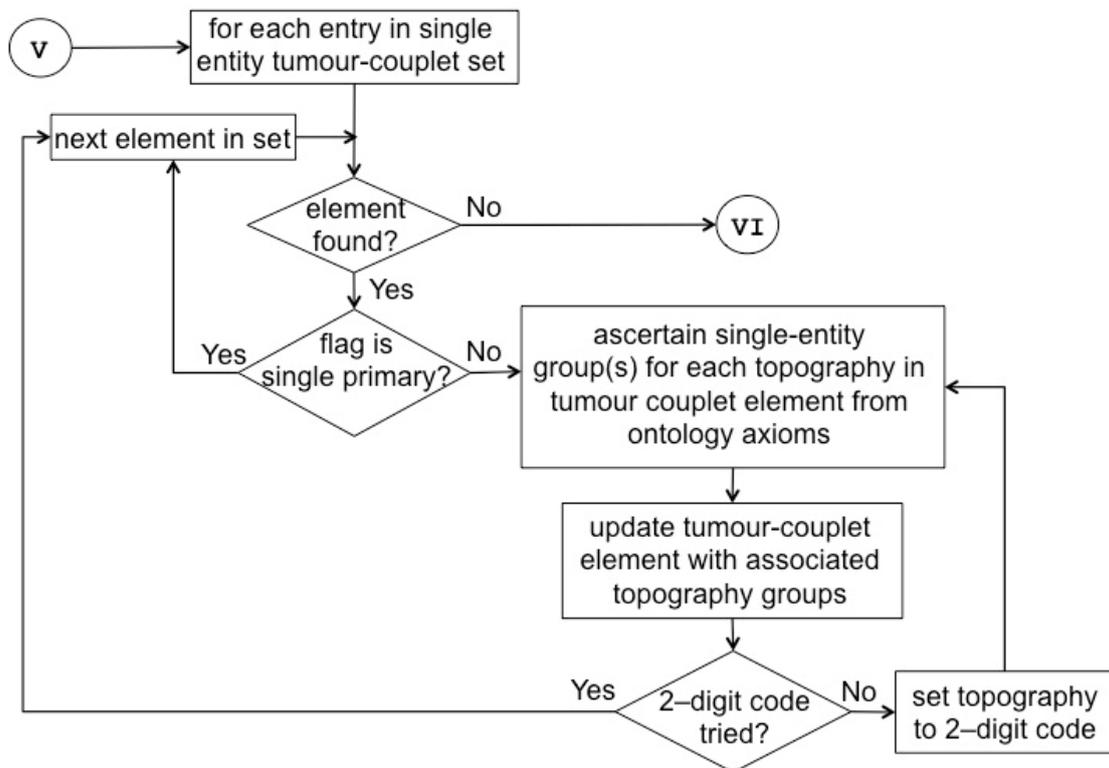


Figure A5. Process to determine single-entity topography groups for each pair of tumours previously added to a set.

The remaining stage of the process (Figure A6) is to cycle through the set of tumour pairs and output a warning to the user for any pair inferred to be a single tumour. The first condition is to verify if a single tumour was determined purely from the morphology-code checks (i.e., conjunction of haematological morphologies with an equivalent single-entity morphology group) and if not, to verify whether the topography single-entity groups are equivalent. The tests for the latter are similar to those for the morphologies but instead of comparing single variables, the tests need to look for the patterns in lists of variables (owing to the fact that each topography code may be associated with more than one single-entity topography group).

Figure A7 illustrates the results using a programme interface developed along these lines for the same five patient test cases described in Section 3. The topography and morphology codes are shown for each patient. Violation of the MPT rules is indicated with a warning followed by the tumour permutation causing the violation and the reason. Valid MPT test cases are those without warnings.

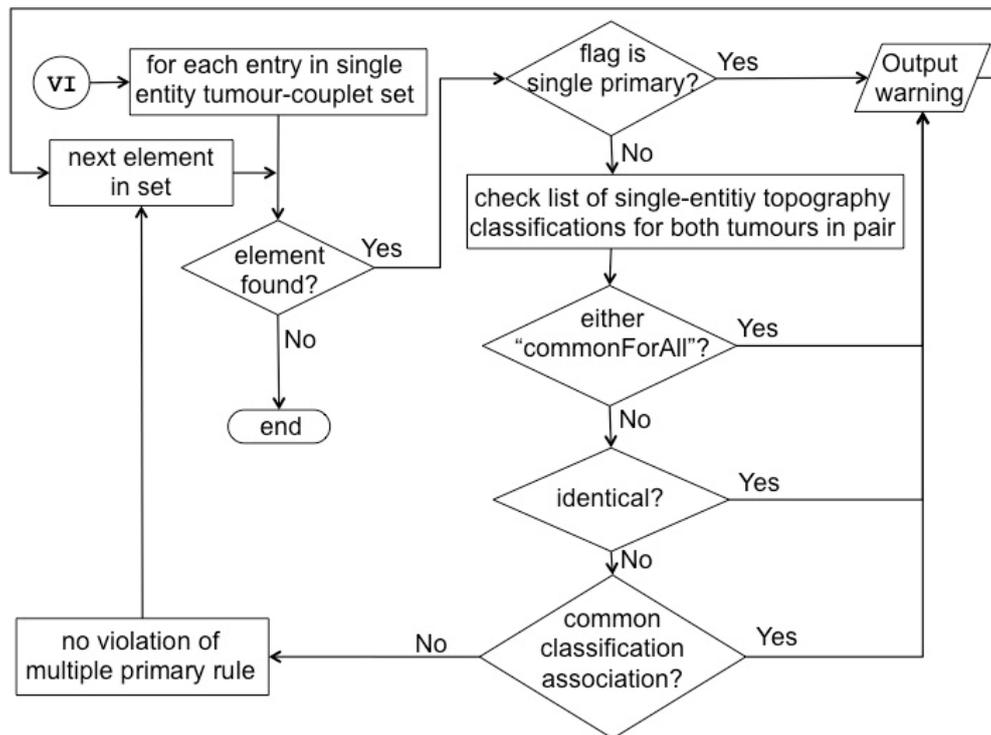


Figure A6. Final stage of the process that ascertains whether a tumour-pair permutation should be considered as a single entity and outputs a warning accordingly.

```

p1:
0001_p1 Topog/Morph: (1): C150 M_9705
0001_p1 Topog/Morph: (2): C269 M_8020
0001_p1 Topog/Morph: (3): C768 M_8020
Warning: 0001_p1. Single tumour site detected:
Tumour seq nums: 2,3
(C269 C768) Topog grp:[TopographyGroupCommonE, TopographyGroupCommonB,
TopographyGroupCommonD, TopographyGroupCommonC] & TopographyGroupCommonS,
(M_8020 M_8020) Morph grp:SingleEntityGroupMorphologyCommonCarcinoma

p2:
0002_p2 Topog/Morph: (1): C260 M_8051
0002_p2 Topog/Morph: (2): C269 M_8140

p3:
0003_p3 Topog/Morph: (1): C390 M_9823
0003_p3 Topog/Morph: (2): C398 M_9365

p4:
0004_p4 Topog/Morph: (1): C398 M_8260
0004_p4 Topog/Morph: (2): C761 M_8140
Warning: 0004_p4. Single tumour site detected:
Tumour seq nums: 1,2
(C398 C761) Topog grp:TopographyGroupCommonH,
(M_8260 M_8140) Morph grp:SingleEntityGroupMorphologyAdenocarcinoma

p5:
0005_p5 Topog/Morph: (1): C220 M_9724
0005_p5 Topog/Morph: (2): C209 M_9714
Warning: 0005_p5. Single tumour site detected:
  
```

Figure A7. Output of the OWL-API programme interface for the same multiple-tumour test cases of Section 3. The highlighted “Warning” text indicates violation of the MPT rules for the tumour permutations and reasons provided underneath it.

References

1. Muir, C.S.; Démaret, E.; Boyle, P. The cancer registry in cancer control: An overview. In *The Role of the Registry in Cancer Control*; IARC Scientific Publications No., 66, Parkin, D.M., Wagner, G., Muir, C.S., Eds.; IARC Publications, World Health Organization: Geneva, Switzerland, 1985; pp. 13–26.
2. National Cancer Institute. Surveillance, Epidemiology, and End Results Program (SEER). Available online: <https://seer.cancer.gov/icd-o-3/> (accessed on 9 September 2019).
3. European Commission. European Cancer Information System (ECIS). 2018. Available online: <https://ecis.jrc.ec.europa.eu> (accessed on 4 July 2019).
4. World Health Organization, CI5 Cancer Incidence in Five Continents. Available online: <https://ci5.iarc.fr/Default.aspx> (accessed on 8 October 2020).
5. London School of Hygiene and Tropical Medicine. CONCORD Programme. Available online: <https://csg.lshtm.ac.uk/research/themes/concord-programme/> (accessed on 8 October 2020).
6. EUROCARE. Survival of Cancer Patients in Europe. Available online: <https://www.eurocare.it> (accessed on 8 October 2020).
7. Martos, C.; Crocetti, E.; Visser, O.; Rous, B.; Giusti, F.; Cancer Data Quality Checks Working Group. A Proposal on Cancer Data Quality Checks: One Common Procedure for European Cancer Registries. JRC Technical Report, Version 1.1. 2018. Available online: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC105078/cancer_data_quality_checks_procedure_report_1.1_print_5_1.pdf (accessed on 15 January 2020).
8. Giusti, F.; Martos, C.; Adriani, S.; Dyba, T.; von Voithenberg, L.V.; Neamtiu, L.; Carvalho, R.N.; Randi, G.; Dimitrova, N.; Nicholson, N. The JRC-ENCR Quality Check Software (QCS) for the Validation of Cancer Registry Data: User Compendium. JRC Technical Report version 1.8.1. 2019. Available online: https://encr.eu/sites/default/files/User_compendium_v1_8_1.pdf (accessed on 15 January 2020).
9. W3C. Web Ontology Language (OWL). 2013. Available online: <https://www.w3.org/OWL/> (accessed on 4 July 2019).
10. Napolitano, G.; González-Beltrán, A.; Fox, C.; Marshall, A.; Finkelstein, A.; Mccarron, P. Biomedical Ontologies and Grid Computing—As New Resources for Cancer Registries. In Proceedings of the Second International Conference on Health Informatics, Porto, Portugal, 14–17 January 2009; pp. 512–517.
11. Brüggemann, S. Rule Mining for Automatic Ontology Based Data Cleaning. In *Progress in WWW Research and Development, Proceedings of the 10th Asia-Pacific Web Conference, Shenyang, China, 26–28 April 2008*; Zhang, Y., Yu, G., Bertino, E., Xu, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 522–527.
12. Napolitano, G.; Marshall, A.; Hamilton, P.; Fox, C.; Gavin, A. An Ontology-Based System for Information Extraction, Reasoning and Cancer Registration from Pathology Reports. In Abstracts from the NCRI Cancer Conference. 2012. Available online: <https://abstracts.ncri.org.uk/abstract/an-ontology-based-system-for-information-extraction-reasoning-and-cancer-registration-from-pathology-reports-2/> (accessed on 6 July 2021).
13. Esteban-Gil, A.; Fernandez-Breis, J.T.; Boeker, M. Analysis and visualization of disease course in a semantically-enabled cancer registry. *J. Biomed. Semant.* **2017**, *8*, 46. [[CrossRef](#)] [[PubMed](#)]
14. Massicano, F.; Sasso, A.; Amaral-Silva, H.; Oleynik, M.; Nobrega, C.; Patrao, D.F. An Ontology for TNM Clinical Stage Inference. In Proceedings of the Brazilian Seminar on Ontologies (ONTOBRAS), São Paulo, Brazil, 8–11 September 2015; Freitas, F., Baiao, F., Eds.; CEUR-WS: Aachen, Germany, 2015.
15. Nicholson, N.C.; Giusti, F.; Bettio, M.; Negro Carvalho, R.; Dimitrova, N.; Dyba, T.; Flego, M.; Neamtiu, L.; Randi, G.; Martos, C. An ontology-based approach for developing a harmonised data-validation tool for European cancer registration. *J. Biomed. Semant.* **2021**, *12*, 1. [[CrossRef](#)] [[PubMed](#)]
16. International Association of Cancer Registries. International rules for multiple primary cancers. *Asian Pac. J. Cancer Prev.* **2005**, *6*, 104–106.
17. Van Hoesen, L.R.; De Bruijne, M.C.; Kemper, P.F.; Koopman, M.M.; Rondeel, J.M.; Leyte, A.; Koffijberg, H.; Janssen, M.P.; Roes, K.C. Validation of multisource electronic health record data: An application to blood transfusion data. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 107. [[CrossRef](#)] [[PubMed](#)]
18. Martínez-Costa, C.; Schulz, S. Validating EHR clinical models using ontology patterns. *J. Biomed. Inform.* **2017**, *76*, 124–137. [[CrossRef](#)] [[PubMed](#)]
19. Maqbool, H.; Muhammad, A.; Taqdir, A.; Rahman, A.; Wajahat, A.K.; Arif, J.; Sungyoung, L.; Byeong, H.K.; Khalid, L. Data-driven knowledge acquisition, validation, and transformation into HL7 Arden Syntax. *Artif. Intell. Med.* **2018**, *92*, 51–70. [[CrossRef](#)]
20. Maqbool, H.; Muhammad, A.; Khalid, M.M.; Taqdir, A.; Wajahat, A.K.; Muhammad, I.; Arif, J.; Sungyoung, L. Acquiring guideline-enabled data driven clinical knowledge model using formally verified refined knowledge acquisition method. *Comput. Methods Programs Biomed.* **2020**, *197*, 105701. [[CrossRef](#)]
21. Muhammad, A.; Maqbool, H.; Wajahat, A.K.; Taqdir, A.; Sungyoung, L.; Eui-Nam, H.; Hafiz, F.A.; Arif, J.; Hassan, I.; Muhammad, I.; et al. Comprehensible knowledge model creation for cancer treatment decision making. *Comput. Biol. Med.* **2017**, *82*, 119–129. [[CrossRef](#)]
22. Boeker, M.; França, F.; Bronsert, P.; Schulz, S. TNM-O: Ontology support for staging of malignant tumours. *J. Biomed. Semant.* **2016**, *7*, 64. [[CrossRef](#)] [[PubMed](#)]
23. Sattler, U.; Stevens, R. Being Complex on the Left-Hand Side: General Concept Inclusions. *Ontogenesis* **2012**. Available online: <http://ontogenesis.knowledgeblog.org/1288> (accessed on 4 August 2021).

24. Open Biological and Biomedical Ontology (OBO) Foundry. Available online: <http://www.obofoundry.org/> (accessed on 6 July 2021).
25. International Association of Cancer Registries (IACR). Check and Conversion Program. Available online: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iaccrgtools&catid=68&Itemid=445 (accessed on 6 July 2021).
26. North American Association of Central Cancer Registries (NAACCR). Recommended Data Edits and Software Coordination of Standards. Available online: <http://datadictionary.naacrr.org/default.aspx?c=4&Version=21> (accessed on 6 July 2021).
27. World Health Organization. International Classification of Diseases for Oncology (ICD-O)—3rd Edition, 1st Revision. 2013. Available online: <https://apps.who.int/iris/handle/10665/96612> (accessed on 6 July 2021).
28. Musen, M.A.; Protégé Team. The Protégé Project: A Look Back and a Look Forward. *AI Matters* **2015**, *1*, 4–12. [[CrossRef](#)] [[PubMed](#)]
29. Horridge, M.; Bechhofer, S. The OWL API: A Java API for OWL Ontologies. *Semant. Web* **2011**, *2*, 11–21. [[CrossRef](#)]
30. Baader, F.; Horrocks, I.; Lutz, C.; Sattler, U. *An Introduction to Description Logic*; Chapter 5; Cambridge University Press: Cambridge, UK, 2017.
31. Pan, J.Z. Requirements for a Semantic Web Rule Language. Available online: <https://www.w3.org/2004/12/rules-ws/paper/51/> (accessed on 2 July 2021).
32. Horrocks, I.; Patel-Schneider, P.F.; Boley, H.; Tabet, S.; Grosz, B.; Dean, M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission 21 May 2004. Available online: <https://www.w3.org/Submission/2004/SUBM-SWRL-20040521/> (accessed on 2 July 2021).