

# TNM-O: A Modular Ontological Approach for the Representation of Tumour Entities across TNM Versions

Susanne Zabka<sup>a,d</sup>, Stefan Schulz<sup>b</sup>, Oliver Brunner<sup>c,d</sup>, Martin Boeker<sup>d1</sup>

<sup>a</sup>*Institute for Distance Learning (Medical Informatics),  
Beuth University of Applied Sciences Berlin, Germany*

<sup>b</sup>*Institute of Medical Informatics, Statistics and Documentation,  
Medical University of Graz, Austria*

<sup>c</sup>*Department of Computer Science, University of Freiburg, Germany*

<sup>d</sup>*Institute for Medical Biometry and Medical Informatics & Comprehensive Cancer  
Center, Faculty of Medicine and Medical Center, University of Freiburg, Germany*

**Abstract.** The TNM classification (Tumour-Node-Metastasis) is the most important coding scheme used to stage tumours based on size or location. Its coding rules often change with different TNM versions, such that the same tumour may be represented by different codes in different TNM versions. We present an ontology-based modular architecture for the management of the TNM coding system. Separate OWL files representing the coding rules for pancreas tumours in the considerably different TNM versions 7 and 8 were created to demonstrate how mappings between TNM versions can be supported. A modular approach with BioTopLite2 as domain top-level ontology, a “hub”-ontology TNM-O containing general TNM and tumour criteria and an ontology for the anatomical entities based on the Foundational Model of Anatomy (FMA) was used as a common basis. For each tumour location and TNM version additional OWL files are created, following strictly defined design patterns. An important feature of the architecture is that for each tumour location and TNM version mappings are encoded in bridging ontologies, which enable re-classification of tumour instances. This work describes a bridging approach using SWRL rules to represent the mapping criteria between the TNM versions, which were tested with instance data. We could show that a tumour with defined characteristics was correctly classified in different versions of the TNM classification.

**Keywords.** Ontology, TNM classification, pancreas tumour, SWRL rules, TNM-O

## 1. Introduction

The TNM classification has become the accepted basis of cancer staging [1]. The system has undergone several revisions, with the 7<sup>th</sup> edition released in 2009 and the 8<sup>th</sup> one in 2017. TNM supports treatment planning, prognosis, evaluation of treatment results, exchange of information between different participants in the treatment process

---

<sup>1</sup> Corresponding Author: Martin Boeker, Institute for Medical Biometry and Statistics, Medical Center – University of Freiburg, Faculty of Medicine, Stefan-Meier-Str. 26, 79104 Freiburg i. Br., Germany; E-mail: martin.boeker@imbi.uni-freiburg.de

as well as cancer research and control. The TNM coding is a “shorthand notation” with the following codes for the three main components, together with numeric modifiers to describe the extent of the disease [2]:

- T (tumour): primary tumour, codes: Tx, T0, Tis, T1-T4
- N (node): metastatic regional lymph nodes, codes: NX, N0, N1-3
- M (metastasis): distant metastasis, codes: M0 or M1

The meaning of the modifiers depends on the respective tumour entity. TX and NX: no assessment possible, T0, N0, M0: no evidence of tumour, Tis: carcinoma in-situ, numbers 1 – 4: presence of tumour with increasing size or local extent. Depending on the type of tumour, further subdivisions are possible indicated by lower case characters (e.g. N2a and N2b). With a prefix, the pre-treatment cTNM (c = clinical) and post-surgical pTNM (p = pathological) classification are distinguished. A series of additional symbols exists, of which this work will only address the descriptors T, N, and M. TNM is different for each anatomical region, which yields more than sixty different sets of rules.

Table 1 describes the differences between the TNM versions for pancreas tumours. While there was just one set of rules for all pancreas tumours in TNM7, TMN8 distinguishes between tumours of the exocrine pancreas and well-differentiated tumours of the insulin-producing (neuroendocrine) pancreas (grades 1 and 2). The classification of neuroendocrine tumours of higher grades uses the rules for the exocrine pancreas.

**Table 1.** Coding rules for pancreas tumours in TNM7 and TNM8 (slightly abbreviated). Rules for the codes TX, T0, NX, N0 and M0 are not listed here, as they are identical in both TNM versions. [2], [3]

Code	Pancreas TNM7	Exocrine Pancreas TNM8, (incl. neuroendocrine pancreas tumours of grades not 1 and 2)	Neuroendocrine Pancreas TNM8 (grade 1 and 2)
Tis	Carcinoma in-situ	Carcinoma in-situ	--
T1	Confined to pancreas, size max 2 cm	size max 2 cm*	size max 2 cm*
T1a	--	size max 0.5 cm*	--
T1b	--	size 0.5-1 cm*	--
T1c	--	size 1-2 cm*	--
T2	Confined to pancreas, size > 2 cm	size 2-4 cm*	size 2-4 cm*
T3	Invades structures beyond pancreas, but not coeliac axis or superior mesenteric artery	size > 4 cm*	size > 4 cm* or invades duodenum or bile duct
T4	Invades coeliac axis or superior mesenteric artery	Invades coeliac axis or superior mesenteric artery or common hepatic artery	Perforates visceral peritoneum (serosa) or invades other organs/neighbouring structures
N1	Metastatic regional lymph nodes present	1-3 metastatic regional lymph nodes	Metastatic regional lymph nodes present
N2	--	>=4 metastatic regional lymph nodes	--
M1	Distant metastasis present	Distant metastasis present	Distant metastasis present
M1a	--	--	- in liver
M1b	--	--	- in other organ
M1c	--	--	- in liver and in other organ

\*: includes invasion of peripancreatic soft tissue

The table demonstrates that the classification of pancreas tumours is mainly based on size and extension into the adjacent tissue and that the view on the tumour charac-

teristics and the rating of their contribution to the tumour malignancy has changed between TNM7 and TNM8. As a consequence, a tumour of the same extent can be coded very differently in these versions.

In the previous work on TNM for breast [4] and colon cancer [5], we proposed a description-logic [6] – based representation of TNM and argued that its rooting in formal ontologies have advantages over its release as text, because axiomatic descriptions are more precise than textual ones. Thus, descriptions are formally decomposed into all their defining criteria. Further, an overarching TNM ontology can be used for automatic classification of clinical data [5].

In this paper we describe the next step towards a fully implemented TNM ontology, with the new feature of mapping between TNM versions. This allows us to re-classify tumours in different TNM versions.

## 2. Methods

Ontologies were created using Protégé 5.2 [7] in a modular approach. Organ and version-specific ontologies are imported into the “hub”-ontology TNM-O [5] under BioTopLite2 [8], [9]. The ontology TNM-O-BodyParts contains codes for anatomical entities with expressions borrowed from the Foundational Model of Anatomy FMA [10] whenever possible. All ontologies were imported into TNM-O. SWRL rules were set up in the human readable syntax as described in [11]. The ontologies were tested using the HermiT DL reasoner version 1.3.8 [12].

## 3. Results

Three pancreas ontologies were created (Pancreas TNM7, Exocrine Pancreas TNM8 and Neuroendocrine Pancreas TNM8) following a similar, albeit slightly improved structure as already described for breast cancer [4] and colorectal cancer [5]. The basic structure in each of these ontologies is the following: A tumour located in an anatomical region and with specific characteristics, e.g. defined by a quality and its value, is represented by a TNM code. A tumour can be a primary tumour or a tumour aggregate with metastatic regional lymph nodes and/or distant metastases. Classes for tumour qualities, value regions and the representational units were defined in the TNM-O “hub”-ontology and can thus be re-used to create ontologies representing TNM coding rules for various organs. TNM-O serves as the central ontology and imports all other OWL files, thus creating a modular structure as shown in Figure 1. All classes for anatomy were defined in the ontology TNM-O-Bodyparts, using expressions from FMA [10] whenever possible. A bridge ontology contains mappings between the TNM versions.

Mapping rules between TNM7 and TNM8 or vice versa were defined as SWRL rules for every possible combination of conditions for TNM rules. The mapping rules follow the general structure:

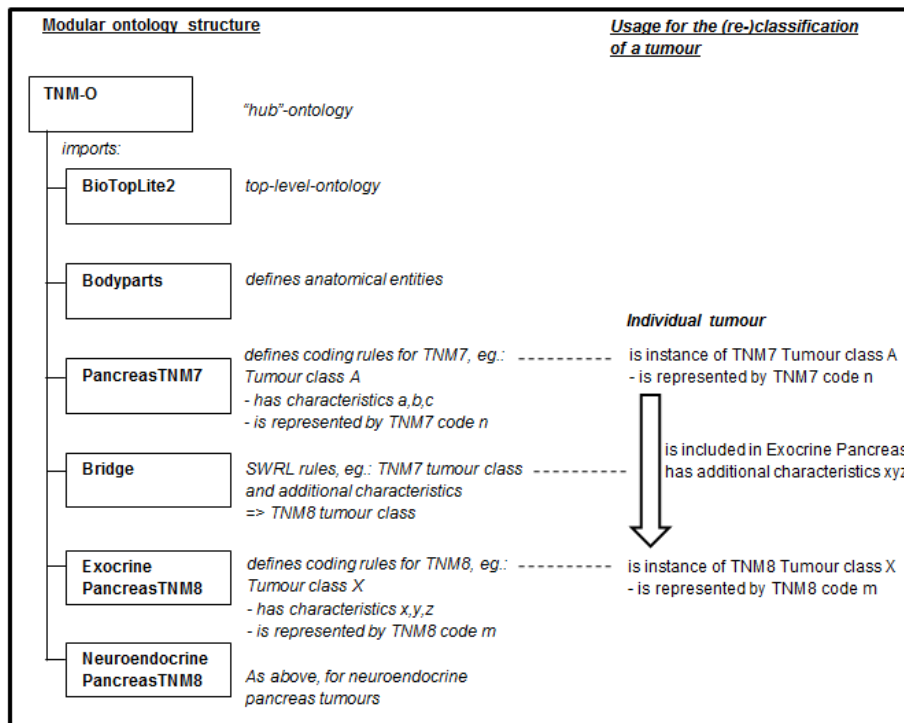
$$TNM7 \text{ tumour} \wedge \text{additional criteria} \Rightarrow TNM8 \text{ tumour}$$

An example for such a rule in human readable syntax is listed below. This shows the rules for the re-classification of a tumour class from the TNM7 ontology, “*InvasivePancreasTumorNotBeyondCeliacTrunkOrSuperiorMesentericArtery*” which

is represented by TNM7 code T3 (compare table 1). As TNM8 does not describe a tumour with exactly the same conditions, this tumour can only be transformed into a tumour class in one of the pancreas TNM8 ontologies by providing further information. In the example below the tumour invades the Common Hepatic Artery and thus can be transformed into the TNM8 tumour class “*InvasiveExocrinePancreasTumorInfiltratingDefinedBloodVessels*“ represented by TNM8 code T4 (compare table 1).

$$\begin{aligned}
 & \text{InvasivePancreasTumorNotBeyondCeliacTrunkOrSuperiorMesentericArtery} (?x) \\
 \wedge & \text{isIncludedIn} (?x, ?exopancreas) \wedge \text{ExocrinePancreas} (?exopancreas) \wedge \text{hasPart} (?x, ?tumorpart) \\
 \wedge & \text{isIncludedIn} (?tumorpart, ?loc) \wedge \text{CommonHepaticArtery} (?loc) \\
 \Rightarrow & \text{InvasiveExocrinePancreasTumorInfiltratingdefinedBloodVessels} (?x)
 \end{aligned}$$

This mapping approach can be used to re-classify a tumour already classified in one TNM version, if the additional criteria needed for the coding in the other TNM version are known. In addition, the approach has the advantage that all “mapping rules” can be easily listed using the SWRL tab in Protégé. The SWRL rules were defined for all possible combinations of tumour characteristics defined in TNM7 and TNM8 and tested with individuals representing all these cases.



**Figure 1.** Modular structure of the TNM ontology as described in the text, explaining how the SWRL rules are used to re-classify an individual tumour.

Another approach without SWRL rules, implementing the mapping rules using a bridge ontology with additional subclass relations or other axioms, is also under evaluation.

#### 4. Conclusion

A modular approach was used to create a set of ontologies for the representation of the TNM coding rules across TNM versions. It could be demonstrated that mapping between different versions of the TNM scheme can be implemented using a “bridge” ontology with SWRL rules. This could be a useful tool for the re-assignment of TNM-codes in different TNM versions and can easily be extended to represent other organs as well.

The ontology will be made available as open source via GitHub.

#### Acknowledgements

This work was 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.

#### References

- [1] Brierley, J., The evolving TNM cancer staging system: an essential component of cancer care. *CMAJ: Canadian Medical Association Journal*, **174**(2) (2006), 155–156.
- [2] Sobin LH, Gospodarowicz MK, Wittekind C., *TNM Classification of Malignant Tumours*, 7edn. John Wiley& Sons, Chichester, West Sussex, Hoboken, 2009.
- [3] Brierley J.D., Gospodarowicz M.K., Wittekind C.: *TNM classification of malignant tumours*, 8edn. Wiley Blackwell, Chichester, West Sussex, UK, 2017.
- [4] Boeker M, Faria R, Schulz S., A Proposal for an Ontology for the Tumor-Node-Metastasis Classification of Malignant Tumors: a Study on Breast Tumors In: Jansen L, Boeker M, Herre H, Loebe F, editors. *Ontologies and Data in Life Sciences (ODLS 2014). Proceedings of the 6th Workshop of the GI Workgroup Ontologies in Biomedicine and Life Sciences (OBML). Volume 1/2014*. Universität Leipzig, Leipzig, 2014.
- [5] Boeker, M., França, F., Bronsert, P., Schulz, S., TNM-O: ontology support for staging of malignant tumours. *Journal of Biomedical Semantics*, **7** (2016), 64. doi:10.1186/s13326-016-0106-9.
- [6] Baader, F. et al., *The Description Logic Handbook*, Cambridge University Press, Cambridge, 2007.
- [7] Musen, M.A., The Protégé project: A look back and a look forward. *AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, **1**(4). (2015) doi: 10.1145/2557001.25757003
- [8] Schulz S. and M. Boeker M., Biotoplite: An upper level ontology for the life sciences evolution, design and application. in *GI-Jahrestagung* (2013) pp. 1889–1899.
- [9] Schulz S, Boeker M, Martinez-Costa C, The BioTop Family of Upper Level Ontological Resources for Biomedicine. *Stud Health Technol Inform*. **235** (2017) 41-445.
- [10] Rosse C, Mejino Jr. JLV., A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. **36**(6) (2003) 478–500. doi:10.1016/j.jbi.2003.11.007
- [11] SWRL: <https://www.w3.org/Submission/SWRL/#2.2> (accessed: 6/5/2017).
- [12] HerMIT OWL reasoner: <http://www.hermit-reasoner.com> (accessed: 6/5/2017)