

# Ontology-enabled Breast Cancer Characterization

Oshani Seneviratne, Sabbir M. Rashid, Shruthi Chari, James P. McCusker, Kristin P. Bennett, James A. Hendler, and Deborah L. McGuinness

Rensselaer Polytechnic Institute, Troy NY 12180, USA

**Abstract.** We address the problem of characterizing breast cancer, which today is done using staging guidelines. Our demo will show different breast cancer staging results that leverage the Whyis semantic nanopublication knowledge graph framework [8]. The system we developed is able to ingest breast cancer characterization guidelines in a semi-automated manner and then use our deductive inferencer to generate new information based on those guidelines as described in our ISWC resource track paper ‘Knowledge Integration for Disease Characterization: A Breast Cancer Example’ [11]. In this paper we demonstrate the versatility of our framework using a synthetic patient profile.

## 1 Introduction

The most recent authoritative cancer staging guideline, i.e. the American Joint Committee on Cancer (AJCC) staging 8<sup>th</sup> edition [1], includes additional data streams not previously used, thus characterizing cancer differently from the previous editions. We constructed the cancer staging ontology for the AJCC 7<sup>th</sup> and 8<sup>th</sup> editions to model the guidelines. In order to address data heterogeneity and to improve interoperability of the data sources used in cancer staging and treatment, we used nanopublications in RDF [9] as the underlying knowledge representation model. These nanopublications are generated utilizing the Semantic Data Dictionary (SDD) methodology [10], thus enabling the nanopublications in the knowledge graph to be updated in a semi-automated manner. Our deductive inferencer is used to characterize (or re-characterize) the stage of a tumor in a given patient and present treatment options extracted from a portion of the National Comprehensive Cancer Network’s biomarker compendium [6]. Crowd sourced resources such as Clinical Interpretations of Variants in Cancer (CIViC) [4] are used to identify drugs that may interact with the biomarkers, as well as related articles and their trust ratings are captured as provenance.

## 2 Related Work

There are existing ontological representations for cancer characterization based on previous AJCC cancer staging editions [7,2]. The cancer characterization in

these ontologies is different from ours due to the inclusion of additional biomarkers as per the AJCC 8<sup>th</sup> edition staging criteria. Unlike previous ontologies, our ontologies also include mappings from breast cancer terms to community-accepted terms from the National Cancer Institute thesaurus (NCIt) [3], and incorporate recommended tests and treatment plans from the openly reusable CIViC database [4]. Furthermore, we include terms that are not included in NCIt or AJCC, such as more specific subclasses of tumor characteristics (*T1*, *T1.as*, *T1.am*, *T1NOS*, etc.) that are available in the Surveillance, Epidemiology, and End Results (SEER) dataset [5]. Additionally, we provide an end-to-end system that demonstrates our ontology’s utility for breast cancer characterization.

### 3 Prototype System Demonstration

Some patients who do not respond as expected to traditional treatment paths may require the physician to consider additional testing measures, new sources, and/or alternative treatment plans. We will exemplify the interplay of disease characterization and personalized medicine being offered by our cancer characterization tool, powered by semantic web tools, in our demonstration.

Let’s consider the patient profile with the biomarkers<sup>1</sup> given in Table 1. Suppose there is a physician considering this multitude of parameters related to tumor biology as well as standard pathology to inform a diagnosis, treatment, and monitoring plan for this patient. With the utilization of the SDD methodology, the data in Table 1 is converted to the nanopublication format in the knowledge graph. Specific to this use case, our cancer staging ontologies include the axioms in Listings 3.1 and 3.2 that were extracted from the AJCC 7<sup>th</sup> and 8<sup>th</sup> editions respectively.

Tumor Size	In lymph nodes?	Metastasized?	Tumor Grade	HER2	ER	PR
5cm (T3)	Yes (N1)	No (M0)	Aggressive (Grade3)	Positive	Positive	Positive

**Table 1.** Patient Profile for Triple Positive Large Tumor

```
@prefix cst: <http://purl.bioontology.org/ontology/CST> .
[] a owl:Class; rdfs:subClassOf cst:R7_Stage_IIIA;
   owl:intersectionOf ( cst:T3 cst:N1 cst:M0 ).
```

**Listing 3.1.** OWL axiom for a tumor to be classified as *Stage IIIA* in the 7<sup>th</sup> guideline

```
[] a owl:Class; rdfs:subClassOf cst:R8_Stage_IIB;
   owl:intersectionOf ( cst:T3 cst:N1 cst:M0
   cst:Grade3 cst:HER2_Pos cst:ER_Pos cst:PR_Pos ).
```

**Listing 3.2.** OWL axiom for a tumor to be classified as *Stage IIB* in the 8<sup>th</sup> guideline

<sup>1</sup> The abbreviations used include: HER2 (Human Epidermal Growth factor receptor 2 ), ER (Estrogen Receptor), PR (Progesterone Receptor).

Once these axioms, along with the other 400+ staging axioms, are applied on the patient nanopublication data, the stages as per the 7<sup>th</sup> and 8<sup>th</sup> editions are determined (IIIA in 7<sup>th</sup> and IIB in 8<sup>th</sup>). Because of the additional data streams considered in the 8<sup>th</sup> guideline (i.e. Tumor Grade, HER2, ER and PR), the patient now has improved prognosis. The visualization tool in Fig. 1 shows the changes to the treatment and monitoring options based on the new stage.

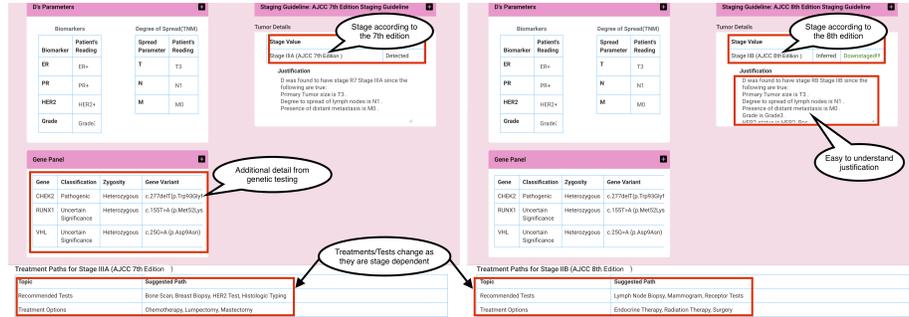


Fig. 1. Example of Downstaged Breast Cancer Characterization

More information on our system, including a video of the system demonstration, is available at <https://cancer-staging-ontology.github.io>.

## 4 Conclusion

With new understanding of cancer biology, guidelines are expected to increase in complexity and personalized care is more sought after as a component of treatment. Similarly, as the data streams for diagnosing and treating cancer patients becomes complicated, physicians may have to consult many different trusted sources and use knowledge from clinical trials and literature to decide on alternative treatment options which can take a great deal of the doctors' and the patients' precious time. Using our cancer characterization tool, physicians have access to patient records in RDF nanopublication format, and can infer the stage, as per the cancer staging ontology, that models the new staging guidelines, and then can investigate alternative, evidence-based treatment options. Our visualizations present evidence-based, updated staging determinations and treatment options, along with provenance. Physician-facing software applications can use our cancer characterization tool to provide physicians with an efficient way to investigate alternative treatment options based on different staging guidelines.

In the future, new guidelines for cancer staging are expected to incorporate genomic test results analyzed in the context of the patient's history. Many predict a rapid influx of information related to cancer from clinical trials, as well as from basic science research. Leveraging all these heterogeneous data sources and making connections to understand the data is of utmost importance. Our system

demonstration shows how semantics are being used to support this fast changing landscape.

## Acknowledgements

This work is partially supported by IBM Research AI through the AI Horizons Network. We thank our colleagues from IBM (Amar Das, Ching-Hua Chen) and RPI (John Erickson, Alexander New, Rebecca Cowan) who provided insight and expertise that greatly assisted the research.

## References

1. Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., Meyer, L., Gress, D.M., Byrd, D.R., Winchester, D.P.: The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more personalized approach to cancer staging. *CA: a cancer journal for clinicians* **67**(2), 93–99 (2017)
2. Boeker, M., França, F., Bronsert, P., Schulz, S.: TNM-O: ontology support for staging of malignant tumours. *Journal of biomedical semantics* **7**(1), 64 (2016)
3. Golbeck, J., Frago, G., Hartel, F., Hendler, J., Oberthaler, J., Parsia, B.: The National Cancer Institute’s thesaurus and ontology. *Web Semantics: Science, Services and Agents on the World Wide Web* **1**(1) (2011)
4. Griffith, M., Spies, N.C., Krysiak, K., McMichael, J.F., Coffman, A.C., Danos, A.M., Ainscough, B.J., Ramirez, C.A., Rieke, D.T., Kujan, L., et al.: CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nature genetics* **49**(2), 170 (2017)
5. Hayat, M.J., Howlader, N., Reichman, M.E., Edwards, B.K.: Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *The oncologist* **12**(1), 20–37 (2007)
6. Kim, H.L., Puymon, M.R., Qin, M., Guru, K., Mohler, J.L.: NCCN clinical practice guidelines in oncology (2013)
7. Massicano, F., Sasso, A., Tomaz, H., Oleynik, M., Nobrega, C., Patrao, D.F.: An Ontology for TNM Clinical Stage Inference. In: *ONTOBRAS* (2015)
8. McCusker, J.P., Rashid, S.M., Agu, N., Bennett, K.P., McGuinness, D.L.: The Whyis Knowledge Graph Framework in Action. *International Semantic Web Conference* (2018)
9. Mons, B., Velterop, J.: Nano-Publication in the e-science era. In: *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*. pp. 14–15 (2009)
10. Rashid, S.M., Chastain, K., Stingone, J.A., McGuinness, D.L., McCusker, J.P.: The Semantic Data Dictionary Approach to Data Annotation & Integration. *Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci)* pp. 47–54 (2017)
11. Seneviratne, O., Rashid, S.M., Chari, S., McCusker, J.P., Bennett, K.P., Hendler, J.A., McGuinness, D.L.: Knowledge Integration for Disease Characterization: A Breast Cancer Example. *International Semantic Web Conference* (2018)