

# Reasoning on Cancer

## Automatic Staging using Ontology-based Knowledge Representation

G Napolitano<sup>1,2</sup>, A Marshall<sup>2</sup>, P Hamilton<sup>3</sup>, A Gavin<sup>1</sup>

<sup>1</sup>Northern Ireland Cancer Registry, Centre for Public Health, Queen's University Belfast; <sup>2</sup>Centre for Statistical Science and Operational Research, School of Mathematics and Physics, Queen's University Belfast; <sup>3</sup>Bioimaging Core Technology Unit, Queen's University Belfast

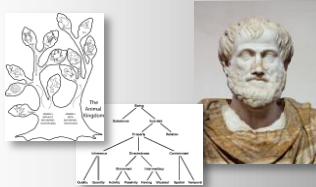


### Background

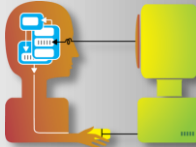
In computer science, a formal **ontology** is a model of a domain of the world.

This domain is described by means of *classes*, *individuals* and *properties*.

The classification is hierarchical, usually following Aristotelian principles.



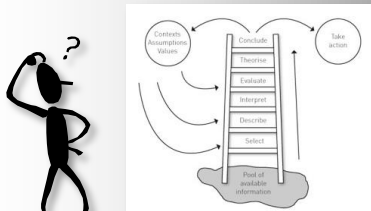
Importantly, it is **machine-readable** but also **human-friendly**, normally specified by means of a logic language.



**Biomedical** ontologies are becoming more and more ubiquitous for the natural classification of biomedical concepts.

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY	
Statistics	Value
Ontologies	506
Terms	6,394,583
Resources Indexed	23
Indexed Records	4,315,417
Direct Annotations	2,103,175,130
Direct Plus Expanded Annotations	15,116,696,652

In the simplest scenarios they can be used as terminologies or controlled vocabularies. But their real strength resides in their capacity to be used as a basis for **inferring information** which is not explicitly mentioned in documents and databases.



Very few experiments have been made to explore the use of ontologies for cancer staging.

## TNM

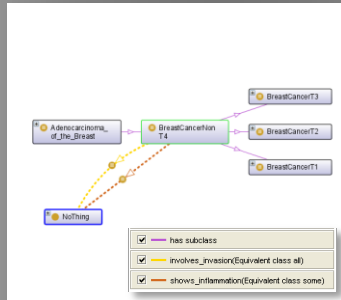
Here we show a successful example of automatic pT classification of breast cancer.

### How it works

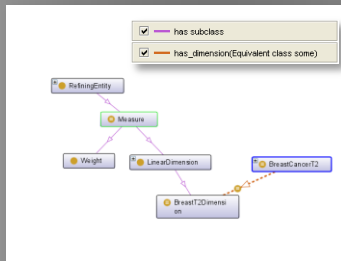
OWL2 is the standard logic language chosen for the Semantic Web.



Protégé 4 is a free tool to design ontologies. Here, screenshots of essential knowledge representation for TNM (ver.6) classification of breast cancer are shown graphically.

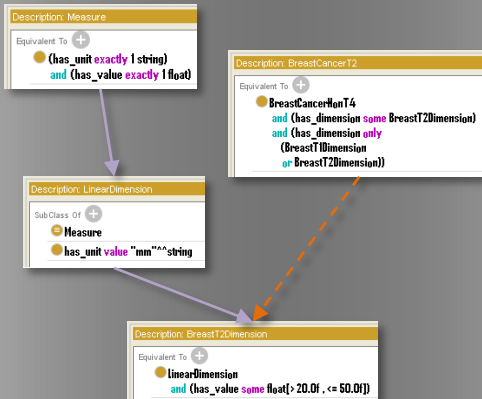


T1, T2 and T3 breast cancers do not involve skin or chest wall invasion or inflammation.



T2 breast cancers are defined only within a certain range of dimensions.

Logical definitions – axioms – are behind these graphs. For instance, the following axioms define the allowed dimensions for T2 breast cancers:



### Results

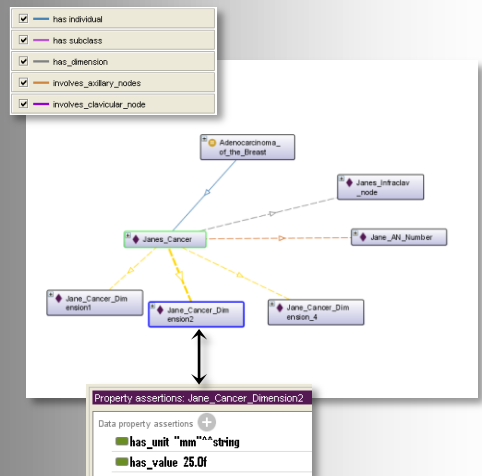
Sample assertions were used to simulate the existence of patients affected by tumours with three linear dimensions.

All three linear dimensions had to be specified, due to the Open World Assumption (OWA) typical of most logic languages.

**OWA: if something is not asserted to be true, it cannot be assumed to be false!**

In our case, if only two dimensions are specified we cannot assume a third (bigger) dimension does not exist and automatic inference cannot occur...

By varying these dimensions to include all possible scenarios, the ontology representation proved to be correct.



Here, for instance, the cancer of a patient called Jane has, among other characteristics, three linear dimensions (the second of which has its properties detailed in the smaller box).

All dimensions will be classified as dimensions of type *BreastT2Dimension*.

So *Jane\_Cancer* will be classified as a cancer of type *BreastCancerT2*.

### Conclusion

It is possible to use the description logic language OWL2 to represent and use knowledge about staging classification to actually stage individual tumours. The Open World Assumption of the language, which causes not expressed information to be assumed always possible, was addressed by a workaround. This solution was suitable in this case but more studies are needed to confirm other classification items can be successfully represented in OWL2.

**Acknowledgements**  
The project was carried out at the Northern Ireland Cancer Registry, which is funded by the Public Health Agency, Northern Ireland.

Correspondence: g.napolitano@qub.ac.uk



...and don't forget to visit our Staging Tool stand!