Representing biomedical knowledge: A view from the philosophy of science

Author **Aleksandra Sojic**
Supervisors Giovanni Boniolo, Salvatore Pece, Sabina Leonelli
Studies/Stage Ph.D. thesis submission
Affiliation European School of Molecular Medicine;
European Institute of Oncology; University of Milan; Milan, Italy
E-Mail aleksandra.sojic@unimi.it

Co-Author **Oliver Kutz**
Studies/Stage Post-doctoral researcher
Affiliation Research Center on Spatial Cognition (SFB/TR 8),
University of Bremen, Germany
E-Mail okutz@informatik.uni-bremen.de

Aims and Objectives of the Research

We aim at identifying the modelling demands across heterogeneous biomedical fields that intersect in breast cancer research. We describe a methodology that covers various epistemic and pragmatic needs in breast cancer conceptualisation as well as formal modelling, taking into account specifically representational means, expressivity and granularity. Moreover, we propose an ontology model for HER2+ breast cancer phenotypes that integrates molecular and clinical representations.

Justification for the Research Topic

The definition of an *ontology* as a specification of a conceptualisation that *represents* particular aspects of the world for a particular purpose [1] invokes the question of how ontology representations (for short, ORs) relate to other kinds of *scientific representations* (for short, SRs). In answering this question from the philosophy of science perspective [2,3,4], we focus on the functional role of SRs, analysing SRs as a tool that allows competent and informed agents to draw specific inferences regarding the representational target [2]. Thus, we aim at providing insights that support an integrative treatment of heterogeneous kinds of 'objects' represented in ontology models, e.g. 'material entities', 'experimental methods', and 'diagnosis'. We distinguish representational targets of SRs and ORs as: a) SRs aim at capturing phenomena that are 'objects' of knowledge, e.g. cell type, carcinogenesis etc., while representing them by various means (e.g. signs, pictures, graphs); b) ORs aim at representing biomedical knowledge, while capturing *what scientists know* and how they think about their target problem. I.e., the claim that 'HER2 is a protein' in an OR captures not only the 'objects' HER2 and protein, but it intends to explicitly represent scientific knowledge, namely the (refutable) scientific inference that HER2 *is* a protein. This 'corpus of scientific knowledge' is communicated in language (formal or not) and various SRs, which are distributed in publications as scientific claims, theories, and models. Thus, the 'corpus of knowledge' about a domain, captured, *inter alia*, in SRs, is the target of ORs[1]. Hence, ORs are specific SRs that function as epistemic vehicles [4,5], providing information on how scientists conceptualise a domain, while targeting particular 'objects'. Accordingly, the concepts and entities represented in an OR are the 'units' of represented knowledge,

[1] Notice, ORs are a special kind of SRs. A significant difference between ORs and other SRs is in the particular features of the ORs which target not only phenomena, but foremost the reasoning about phenomena (domain knowledge). Just like other SRs, the representational means of ORs can vary, including visualisations of the domain conceptualisation in the form of graphs such as semantic and concept maps, as well as formal logical representations.

which function as the epistemic vehicles for reasoning about the domain's knowledge targets. In other words, heterogeneous kinds of 'objects' represented in an ontology serve the common purpose of representing knowledge. Therefore, a treatment of represented 'objects' as epistemic tools justifies their integration into one OR.

Even so, modelling and representing biomedical knowledge involves numerous challenges [6,7]. The complexity of living systems that are objects of biomedical knowledge emerges as an indispensable problem for the design of biomedical ORs [7]. In order to address some of the challenges that an ontology model needs to face in establishing a shared conceptualisation [1], we identify fundamental issues related to modelling the breast cancer domain, e.g. dealing with complexity, heterogeneity, combination of qualitative and quantitative data, mereological and functional representations, as well as various kinds of reasoning within and across domains [5]. In order to approach this heterogeneity of representations, we invoke the view of SRs as sketched above. As a result, we propose a functional, pragmatic approach as the most productive way to treat these various kinds of 'objects' that are constitutive parts of biomedical knowledge, which will eventually be represented as the units of a biomedical OR. Namely, the *pragmatics* here concerns the specific purpose that represented 'units' play in ORs. Each of the units (objects of knowledge) has a particular pragmatic function within a representation. Discussing the case of molecular and clinical breast cancer classification [8,9,10,11], we further analyse the roles that particular classificatory units play in clinical and molecular practice, e.g. explanatory, predictive etc., satisfying practical needs of the domain. We distinguish the most relevant categories of the classificatory units used to characterise molecular and clinical features of breast cancer, dividing them into normative, quantitative and qualitative categories such as *prognosis, therapy response, biomarkers, clinical grading, molecular grading, carcinogenic mechanisms* etc. We also specify how the subcategories of each of the groups can be related to the OBO foundry ontologies for representation of biological processes, metrological, mereological and functional features.

Research Questions

1. How can biomedical ontology integrate molecular and clinical knowledge?
2. Can various perspectives, epistemic and pragmatic needs (e.g. in representing breast cancer) be merged into a single ontology model?
3. What are the specific functions and roles that heterogeneous kinds of 'objects' play within biomedical knowledge that an ontology aims to represent?
4. How can formal ontology deal with breast cancer knowledge representation?

Research Methodology

Our approach combines methodologies from the philosophy of science with ontology research. We thus analyse scientific methods and practices and discuss their impact on the problem of modelling and representing the breast cancer domain. We moreover attempt to extrapolate from the results of philosophical analysis to the field of applied ontology, specifying a methodology for modelling a breast cancer ontology.

Research Results to Date

We have developed a segment of an ontology for a breast cancer subtype, namely HER2+ breast cancer. We moreover modelled the distinction between 'normal' and 'abnormal' HER2 related phenotypes by combining mereological relations with non-monotonic reasoning based on circumscription. In addition to the formal model, we provided an analysis of the relevant scientific methods, representation and reasoning types, thereby justifying particular choices concerning how to represent classes and relations among them. We host the HER2 ontology at OntoHub (http://ontohub.org/), a new ontology repository with formal semantics and supporting different ontology languages and links between ontologies, enabling various reasoning and interoperability scenarios between ontologies.

References

1. Guarino N, Oberle D, and Staab S. 2009. What is an Ontology? *Handbook on Ontologies*: 1-17.
2. Suárez M. 2010. Scientific representation. In *Blackwell's Philosophy Compass*.
3. Giere R. 2010. An agent-based conception of models and scientific representation. *Synthese* 172 (2).
4. Knuuttila T. 2011. Modelling and representing: An artefactual approach to model-based representation. *Studies In History and Philosophy of Science Part A* 42 (2):262-271.
5. Sojic A., Kutz O. 2012, Open Biomedical Ontologies-Formalising Knowledge About Breast Cancer Phenotypes. *Journal of Biomedical Semantics,* Forthcoming.
6. Faratian D, Robert G. C, Crawford W. J, and Harrison J. D. 2009. Systems pathology - taking molecular pathology into a new dimension. *Nat Rev Clin Oncol* 6 (8):455-464.
7. Kutz O, Mossakowski T, Hastings J, Garcia Castro A, and Sojic A. 2011. Hyperontology for the Biomedical Ontologist: A sketch and some examples. *Working with Multiple Biomedical Ontologies Workshop, ICBO 2011*:399-408.
8. Perou M. C, Sorlie T, Eisen B. M, van de Rijn M,Jeffrey S. S, Rees A. C, Pollack J. R, Ross D. T, Hilde J, Akslen L. A, Fluge O, Pergamenschikov A, Williams C, Zhu S. X, Lonning P. E, Borresen-Dale A, Brown P. O, and Botstein D. 2000. Molecular portraits of human breast tumours. *Nature* 406 (6797): 747-752.
9. Viale G, Ghioni M, and Mastropasqua M. G. 2010. Traditional molecular markers and response to adjuvant endocrine or trastuzumab-based therapies. *Current Opinion in Oncology* 22 (6):541-546.
10. Goldhirsch A., Wood W. C, Coates A. S, Gelber R. D, Thürlimann B, and Senn H. J.. 2011. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology*.
11. Sotiriou C, and Pusztai L. 2009. Gene-Expression Signatures in Breast Cancer. *New England Journal of Medicine* 360 (8): 790-800.