Some Remarks on Modelling from a Computer Science Perspective

Görz, Günther

Some Remarks on Modelling from a Computer Science Perspective

Günther Görz [*]

Abstract: *»Einige Bemerkungen zur Modellierung aus der Sicht der Informatik«.* One of the basic tasks of computer science is to rewrite models derived from other scientific disciplines so that they can be represented and processed on computers. If such a reconstruction process is only partially successful or fails entirely, the modification of the initial model becomes an interdisciplinary research task. The modelling task is to be seen as an application of knowledge representation and processing. We distinguish between aiming at models of something or models for some purpose. Modelling of given domains starts with the construction of a formal ontology. To support issues such as modularity and interoperability, in particular in a web-based environment, the idea of reference ontologies came up. For object-based research in the humanities, the Conceptual Reference Model (CRM) by ICOM/CIDOC is such a reference ontology which has become an ISO standard.

Keywords: Modelling, computer science, ontologies.

1. Methodological Preliminaries

My general assessment of modelling results from a general view on the humanities and the sciences, which appear in two forms: a propositional ("textbook", theory) form and a research form. The latter is a form of (methodological) action. Based on experience from interdisciplinary research, we can observe that problem orientation replaces disciplinary constrictions, which can lead to a reconstitution of the unity of science – i.e. of scientific rationality rather than systems. This unity, transcending disciplinary borders, can be seen as a unity of language, where in both cases we have similar procedures of verification and justification (*giving reasons*), which constitute meaning. That is a practical (pragmatic) unity, including the distinction between the context of discovery and the context of justification. So, it is essential for modelling to provide a linguistic framework for conceptual modelling and justification. That will

[*] Günther Görz, FAU Erlangen-Nürnberg, Computer Science Department, AG Digital Humanities, Konrad-Zuse-Str. 3-5, 91058 Erlangen, Germany; guenther.goerz@fau.de. Bibliotheca Hertziana, Max-Planck-Institut für Kunstgeschichte, Via Gregoriana 28, 00187, Roma, Italy.

Historical Social Research Supplement 31 (2018), 163-169 │ published by GESIS
DOI: 10.12759/hsr.suppl.31.2018.163-169

comprise the identification of common generic concepts and relations / properties from an event, process or action perspective, respectively.

2. Computer Science and the Concept of Model

One of the basic tasks of computer science is to rewrite models derived from other scientific disciplines so that they can be represented and processed on computers. More specifically, this means that given models have to be transformed into versions for which effective procedures can be given. A key application area of reconstructed models is simulation (see Wedekind et al. 1998). If an initial model cannot be translated directly into the language of computer science, a reconstruction step is required. In the humanities in particular, understanding and explanation of actions in terms of reasons and intentionality provide challenges to operationalized representations.

If such a reconstruction process is only partially successful or fails completely, the modification of the initial model becomes an interdisciplinary research task. Key issues for success include making the disciplinary terminology more precise, modifying the modelling approach, and extending the range of computer science methods. This offers opportunities not only to refine existing knowledge, but also to develop new epistemological interests ("Erkenntnisinteressen") in the respective disciplines as well as in computer science. That is the basic meaning of *computational science* and *computational humanities*, whereby it is essential that both parties speak a common language.

When studying the creation of models and its methodological and technical foundations, the precise introduction of metalevel terms such as *model* and *simulation* becomes indispensable. In computer science we find two elaborate uses of the term *model*, one derived from mathematics and physics, the other influenced by empirical applications in natural and social sciences and in engineering.

In mathematics and theoretical computer science, the concept of *model* is used only in the context of structuring theoretic approaches. Such structural models sensu stricto exist only in logic and mathematics. The idea is often applied to physics, but there the axiom systems contain not only schematic, but also interpreted parts from the very beginning. The second understanding of *model* deriving from application areas means a certain way of describing empirical processes, mostly within a naive realistic epistemic framework. Talking about modelling some *external reality* is in fact dealing with discipline specific experimentation and observation contexts, i.e., about descriptions of relevant states of affairs. The claim that a model simplifies some part of reality does in fact refer to the simplification in such descriptions – linguistic means can only be applied to linguistic objects.

The level of detail of the initial description is determined by the research program (Lakatos) of the respective theory in which context the model is to be created. Such a research program provides the language of the description, the description standards, and the explanatory schemata. Last but not least the expectations which precede the construction of a model are formulated within this framework. Hence, modelling is an activity that has two aspects: A model must satisfy the empirical descriptions, but also the theoretical specifications. For model construction, comprehensive descriptions of states of affairs as they result from observations, etc., run through certain modifications, so called *idealizations*: values are being smoothed, different expressions of a feature are replaced by mean values, and certain influences are regarded as negligible. These modifications aim to achieve relatively simple and clear representations of empirical states of affairs as well as adaptability to pertinent theoretical structures. The result of such an idealization is called a descriptive idealized model. If it is a structural model of a theory at the same time, then the hypothetical assertive claim ("Geltungsanspruch") is regarded as confirmed and the state of affairs is seen as described, or explained, by the theory. In engineering the theoretical structure is often complemented by technical (functional) standards–this is the case of technical idealized models such as construction plans and schedules.

Simulation models are special descriptive models of technical or natural systems, which are confined to certain material restrictions. A system in general is given by a set of elements, which are bound together by certain relations and are separated by clear boundaries from its environment. A system is a technical system if its external effects as well as its internal relations are determined by objective functions ("Zielfunktionen"). The abstraction steps involved in the construction of a simulation model do not aim at the generation of a class of cases (as in the descriptive idealized model), but at the generation of a class of variations of a base case.

3. Modelling, Knowledge Representation, and Formal Ontologies

Computer science has a special role in the construction of idealized or simulation models: First, it has to organize the initial descriptions of the models in such a way that the required modelling steps can be carried out, and the descriptions and models have to be transformed into an appropriate representation. This comprises determining how objects can be represented by features or feature groups in general, which relations can be set up among them and how certain feature values are assigned meanings such that these can be processed as data; i.e. the design of data structures and processing rules.

In more detail, this means that the modelling task is to be seen as an application of knowledge representation and processing, which in my view consists of a purpose-driven formal reconstruction of a body of knowledge and its implementation in a (logical) language. Initially we can already distinguish whether our theoretical enterprise aims at *models of* something or *models for* some purpose(s). Hence, the construction of a knowledge base ("knowledge engineering") requires at least the determination and delimitation of the domain of discourse, a determination of the relevant concepts and properties ("what?" as opposed to "how?"), where properties are represented by relations ("has-"), and a hierarchical ordering of concepts and properties ("is-a"). This simple framework already allows for the representation of particular objects (individuals) as instances of concepts. Of course, in most cases there are many desirable extensions to what can be expressed such as constraints restricting properties in various ways, or the specification of properties of properties. To express further relationships between concepts and between particulars, rules ("axioms") are often introduced. Although it sounds trivial that implicit knowledge cannot be processed algorithmically, ontology construction is a good exercise to enforce the development of methods for the explicit representation of implicit knowledge. To summarize: modelling of a given domain starts with the construction of a formal ontology, which in turn can serve as the basis for the construction of a theory, often in the form of a critical reconstruction (see Görz 2016).

At this point, a short remark about semantics seems appropriate: The logical framework provides the structural part, and the meaning of content words (*concepts*) is given by a network of relations even if we include controlled vocabularies; but in an empirical setting reference must be provided by *external* grounding. Nevertheless, semantics is meant to refer to the logical framework, i.e., an inference relation. Reasoning should be performed by sound and complete inference rules as in, e.g., Description Logic. How to deal with vagueness and imprecision in such a framework is still a research question. Another challenge is to take account of conflicting information, such as diverging ascriptions of dates or places or authorship. And, of course, deductive reasoning is only one side of the coin, and must be complemented by an *ars inventoria*, i.e. heuristic procedures, as Leibniz has already stated (1679/1999).

With formal ontologies, several issues arise such as modularity and interoperability, in particular in a web-based environment. Therefore, the idea of reference ontologies, which contain generic concepts and properties relevant for many applications, came up. Specifically for object-based research in the humanities, the Conceptual Reference Model (CRM) by ICOM/CIDOC is such reference ontology, which is now an ISO standard. Its characteristic feature is that it is event-based and easily extensible: a series of extensions for geographic data, archaeology, and scientific observations have been suggested as well as many domain-specific ontologies, e.g. from the museum, library (FRBR) and archive domain, fostering the development of standardized components and libraries. Methodologically, the CRM allows for a chronotopological modelling of data. With CRM as a common reference a high level of interoperability can be achieved.

With the CRM, semi-formal representations are also possible as it allows the inclusion of (uninterpreted) text in its representations, which are in principle open for semantic disclosure at a later point in time. Our implementation in OWL-DL (Görz et al. 2008) allows us to deploy CRM-based ontologies to the Semantic Web and to publish Linked (Open) Data. Because OWL is a very expressive Description Logic, powerful reasoning engines are available. Therefore, such models can serve as explanation models, as opposed to pure functional models which result from existing popular and successful machine learning algorithms. To support cooperative research and communication, in particular for object documentation and object-based knowledge generation and processing, so-called *Virtual Research Environments* (VREs) have been developed. WissKI (Görz 2011; Scholz and Görz 2012; Scholz, Merz and Görz 2016) is such a VRE with special support for data acquisition in the mentioned event-base style through its ontology-based modelling component. Actually it has more than 20 applications in the field of cultural heritage, mostly in museum documentation, but also in providing semantic frameworks for epigraphy (Scholz et al. 2014) and the history of cartography (Görz and Scholz 2013).

To conclude with a remark on simulation: Logic-based models can serve for discrete (qualitative) simulation in a rather immediate fashion using the reasoner. From a theoretical perspective, there is an immediate connection via *proofs as programs* (Curry-Howard correspondence). The domain of continuous simulation models (System Dynamics, etc., as used in social sciences) is beyond the scope of this presentation.

4. Discussion

In the discussion with my opponent Francesca Tomasi, clarification questions have been raised which I hope to have answered by rewriting some formulations present in the text of the presentation. I had the impression that our positions are quite similar. However, I would like to take up some questions that remain. First, Francesca questioned my emphasis on justification. To me it seems clear that giving reasons is fundamental to scientific discourse from the perspective of the philosophy of science, and this includes the humanities.

The conditions and rules of argumentation may be different, but justification is essential for knowledge, as opposed to pure opinion. She then asked why understanding and explanation of actions in terms of reasons and intentions are a challenge for operationalized representations.

Besides a few clarification questions which I hope to have answered by rewriting some formulations in the text of the presentation. I had the impression that our positions are quite similar. The following questions by Francesca remain:

Francesca Tomasi's questions

1) Why do you emphasize justification?
   Answer: Giving reasons is fundamental for scientific discourse from the perspective of the philosophy of science, and this includes the humanities. The conditions and rules of argumentation may be different, but justification is essential for knowledge, as opposed to pure opinion.
2) Why is understanding and explanation of actions in terms of reasons and intentions a challenge for operationalized representations?
   Answer: The question is whether and, if so, how intention can be operationalized. Of course, this depends on our definition of intention. Although there is a highly controversial philosophical discussion about intentionality, I cannot see any viable method to formalize intention completely in line with physical terms. So we need some way to deal with it on the computational level: that's the challenge.
3) Is a model not an explanation?
   Answer: Yes, in a sense. I would prefer to say that an operational model such as a deductive or simulation model can provide explanations.
4) Do you think that the construction of the theory is next to the construction of the ontology?
   Answer: Yes, at least insofar as a formal ontology is a necessary condition for the construction of a theory in the strict – not postmodern – sense. The ontology defines the concepts, at least.

References

Görz, Günther, Martin Oischinger, and Bernhard Schiemann. 2008. An Implementation of the CIDOC Conceptual Reference Model (4.2.4) in OWL-DL. In *The Digital Curation of Cultural Heritage. Proceedings of ICOM CIDOC 2008*, ed. Angelos Delivorrias et al., 1-14. Athens: ICOM/CIDOC <http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/archives/ConferencePapers/2008/7_papers.pdf> (Accessed April 19, 2017).

Görz, Günther. 2011. WissKI: Semantische Annotation, Wissensverarbeitung und Wissenschaftskommunikation in einer virtuellen Forschungsumgebung. *Kunstgeschichte. Open Peer Reviewed Journal*: 1-17. Accessed April 19, 2017. doi: urn:nbn:de:bvb:355-kuge-167-7.

Görz, Günther, and Martin Scholz. 2013. Semantic Annotation for Medieval Cartography: The Example of the Behaim Globe of 1492. *e-Perimetron* 8 (1): 26-36 <http://www.e-perimetron.org/> (Accessed April 19, 2017).

Görz, Günther. 2016. Sind formale Ontologien Orthosprachen? In *Paul Lorenzen und die konstruktive Philosophie*, ed. Jürgen Mittelstraß, 175-88. Münster: mentis Verlag.

Leibniz, Gottfried Wilhelm. 1679/1999. Consilium de Encyclopaedia nova conscribenda methodo inventoria. In *Sämtliche Schriften und Briefe, Sechste Reihe Philosophische Schriften, Vierter Band,* 338-49. Berlin: Akademie-Verlag.

Scholz, Martin, and Günther Görz. 2012. WissKI: A Virtual Research Environment for Cultural Heritage. In *Proceedings of the European Conference on Artificial Intelligence, ECAI 2012*, ed. Luc De Raedt et al., 1017-8. Amsterdam: IOS Press.

Scholz, Martin, Marvin Holdenried, Boris Dreyer, and Günther Görz. 2014. EDEN – *An Epigraphic Web Database of Ancient Inscriptions*. Paper presented at CIDOC 2014, September 11-14, Dresden, Germany.

Scholz, Martin, Dorian Merz, and Günther Görz. 2016. Working with WissKI – A Research Environment for Object Documentation and Object-Based Research. In *Digital Humanities 2016. Conference Abstracts*, ed. Maciej Eder and Jan Rybicki. Krakow, 944-5.

Wedekind, Hartmut, Günther Görz, Rudolf Kötter, and Rüdiger Inhetveen. 1998. Modellierung, Simulation, Visualisierung: Zu aktuellen Aufgaben der Informatik. *Informatik-Spektrum* 21: 265-72.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Supplement 31 (2018): Models and Modelling between Digital Humanities – A Multidisciplinary Perspective