PROVenance Patterns in Numerical Modelling and Finite Element Simulation Processes of Bio-electric Systems*

Max Schröder[1,2], Hendrikje Raben[3], Frank Krüger[1], Andreas Ruscheinski[4], Ursula van Rienen[3,5], Adelinde Uhrmacher[4,5], Sascha Spors[1]

***Abstract*— The reproducibility of scientific results gains increasing attention. In the context of biomedical engineering, this applies to experimental studies of three different kinds: *in-vivo*, *in-vitro*, and *in-silico*. Numerical modelling and finite element simulation of bio-electric systems are intricate processes involving manifold steps. A typical example of this process is the electrical stimulation at alloplastic reconstruction plates of the mandible. During the bio-electric modelling and simulation process, diverse methods realised in various software tools are exploited. To comprehensibly render how the final model has been developed requires a thorough documentation. We exploit the W3C provenance model *PROV* to structure this process and to make it accessible for modellers and for automatic analyses. Different entity types, such as data, model, software, literature, assumptions, and mathematical equations are distinguished; roles of entities within an activity are revealed as well as the involved researchers. In addition, we identify five process patterns: 1) *information extraction* from the literature; 2) *generation* of a geometrical model which uses data as input; 3) *composition* of several geometrical or mathematical models into a combined model; 4) *parameterisation*, which augments the input model by additional properties; and, finally, 5) *refinement*, which uses a model in addition to an assumption and generates an enhanced model. By modelling provenance information of a typical bio-electric modelling and simulation process as well as identifying provenance patterns, we provide a first step towards a better documentation of academic investigations in that scientific field.**

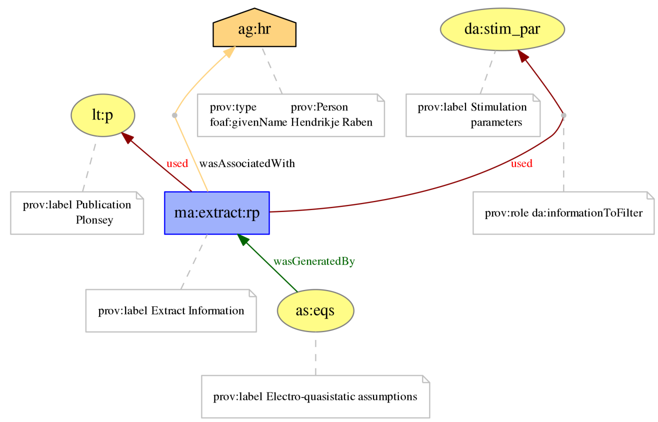

Fig. 1: Instance of the extraction pattern encoding the extraction of assumptions out of extant literature by filtering the information with known details on the stimulation method. Ellipses represent entities, rectangles represent the activity, the pentagon represents the agent. Relations are illustrated by connections between the elements; additional information are shown within white boxes. Entity prefixes encode the type of the entity: `da` – data, `lt` – literature, and `as` – assumption.

I. Introduction

The amount of data, parameters, and methods is growing fast in all types of experiments: *in-vivo*, *in-vitro*, and *in-silico* [1]. Keeping track of all these information during an experiment is essential. Also, to understand and interpret the experimental finding, a thorough documentation of processes and sources that contributed to experiments is crucial [2]. For this purpose so-called *provenance models* are employed [3], [4]. These models enable not only to formally describe the generation and usage of data, scripts, and other research artefacts, but also help to evaluate the trustworthiness of them [5], [6]. Especially when preparing medical therapies based on the research findings or applying them to uncontrolled eco-systems, the provenance information is crucial. Provenance models then help to analyse which entities might be affected by these impairments.

In the field of numerical modelling and finite element simulation, *in-silico* experiments are intricate processes involving a multitude of data (formats), models, and software, but also assumptions, literature and equations. Therefore, the formalisation of this process as a provenance model promises to help interpreting and assessing the final results, and at the same time also their replication [7]. A provenance model of a study that combined *in-vitro* and *in-silico* experiments for developing a signalling pathway model [8] revealed already benefits in applying provenance models to *in-silico* studies [9]. The diverse sources, research artefacts, and processes were structured and their role in developing the simulation model made explicit, accessible and available to automatic inference mechanisms. Another application of a provenance formalism to a simulation study, this time in demography for modelling migration [10], underlined the potential but also showed that, although crucial entities re-emerge, such as simulation model, data, theory, and simulation experiment, the structure of provenance models varies significantly. For example in [10], the composition based on diverse theories

* *This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Collaborative Research Centre 1270 ELAINE and UH 66/18 GrEASE, and a grant of the Federal State of Mecklenburg-Vorpommern.*

[1] Institute of Communications Engineering, University of Rostock, Germany (correspondence e-mail: max.schroeder@uni-rostock.de)
[2] University Library, University of Rostock, Germany
[3] Institute of General Electrical Engineering, University of Rostock, Germany
[4] Institute of Computer Science, University of Rostock, Germany
[5] Department Life, Light & Matter, University of Rostock, Germany

First published in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 3377-3382, doi: 10.1109/EMBC.2019.8856841
This author manuscript version: https://doi.org/10.18453/rosdok_id00004210

and their calibration based on one large data set on migration largely shaped the provenance model, while in [9] cross-validation with other models, and interrelating successive simulation model extension, simulation experiments, and new experimental data played a decisive role. Following this line of thought, we analyse the application of provenance by using the example of electrical stimulation at alloplastic reconstruction plates of the mandible described by van Rienen et al. [11]. In particular, we use this example to:

- identify research artefacts that play a role in executing simulation studies in the domain of numerical modelling and finite element simulation (Section III-B);
- translate the process of executing a simulation study to analyse the electric fields that are induced due to electrical stimulation at alloplastic reconstruction plates of the mandible into a provenance model (Section III-C);
- identify patterns in the provenance model and discuss their possible relevance for other studies that rely on finite element methods (Section IV).

The provenance model in this paper represents the particular investigations of van Rienen et al. [11]. However, by identifying process patterns and research artefact categories in that specific example, we enable the re-use, adaptation, and extension of them for the application in other processes in that scientific field.

In the following section we introduce the basic concepts and notation of the W3C provenance model *PROV*, that we use to model the provenance information.

II. W3C Provenance Formalism

To describe the provenance information of research artefacts, their relations in the form of activities, and involved researchers, we exploit the W3C *PROV* formalism [12]. According to it, provenance is defined as "information about entities, activities and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness." [12]

Entities describe "physical, digital, conceptual or other kinds of things" [12] and, thus, correspond to all research artefacts within our simulation study, e. g., data, models, software, and assumptions. These entities were used by *activities*, describing processes in our simulation study, to generate new or modify existing entities, e. g., a geometric modelling process that uses tomographical data and an appropriate software to generate an according geometrical model. Finally, we use *agents* to describe a person or a group of persons associated with an activity.

Example 1: Considering the process of extracting information from literature as an example, *entities* are 1) the article itself; 2) the data that are used to filter the information within that article; and 3) the extracted information such as in the form of an assumption. The reading and extraction process itself represents the *activity* and the reader of the article is the corresponding *agent*.

To define *relations* between such entities, activities, and agents, we use the following types from the W3C *PROV* formalism:

- *used*: an activity used an entity,
- *wasGeneratedBy*: an entity was generated by an activity, and
- *wasAssociatedWith*: an activity was associated with an agent.

Furthermore, we not only use the pure form of each type, but also the qualified equivalents of the relations. A *qualified relation* defines the *role*, in other words the purpose, of an entity, activity, or agent within that relation. Note that additional relation types in the W3C *PROV* formalism exist, which, however, were not needed for the provenance modelling in this paper.

Example 2: Following on Example 1, the reading activity *used* the article and the data. The data in that activity are employed in order to filter the information from the article which corresponds to its *role*: `informationToFilter`. The activity then *generated* an assumption. The reader of the article is *associated* with the activity.

As a result of defining entities, activities, agents, and their relations, we get a so-called *provenance graph*. In this paper, we use the following graphical notation of these graphs: Entities are represented as ellipses, activities as rectangles, and agents as pentagons. Figure 1 shows the provenance graph of Example 2. Relations are illustrated as connections between the graphical elements. The role of a qualified relation is shown in a white box connected to the relation. Additional information about entities, activities, and agents are also shown within white boxes connected to the corresponding element.

III. Bio-electric Provenance Model

A. Bio-electric Stimulation Example

In the following, the modelling and simulation workflow for conducting a finite element simulation of a bio-electric problem is briefly introduced. Details can be found in [11]. Starting from the computer tomographic (CT) dataset of a patient with a defective lower jaw, a 3D computer-aided design (CAD) model was created: Firstly, the different tissue and material types in the region of interest (i. e., the lower jaw, its surrounding soft tissue, and the mandibular reconstruction plate with fixation screws) were segmented based on their CT grey values and initially smoothed within the segmentation software Materialise Mimics® v19 (Mimics, RRID:SCR_012153)[1]. Secondly, the segmented model of the bone was transformed into a coarse polygonal surface model, smoothed again, and repaired (i.e., removing undesirable geometry details like spikes and holes) using the CAD software GEOMAGIC Studio v12 (GEOMAGIC Studio, RRID:SCR_016978). The resulting stereolithography objects of the bone and the additional models of two fixation screws then were further smoothed and converted into a non-uniform rational basis spline (NURBS) surface object.

[1] If available, we use Research Resource Identifiers (RRIDs) for referencing software. The RRIDs aim to identify the correct resource, see https://www.scicrunch.org/resources

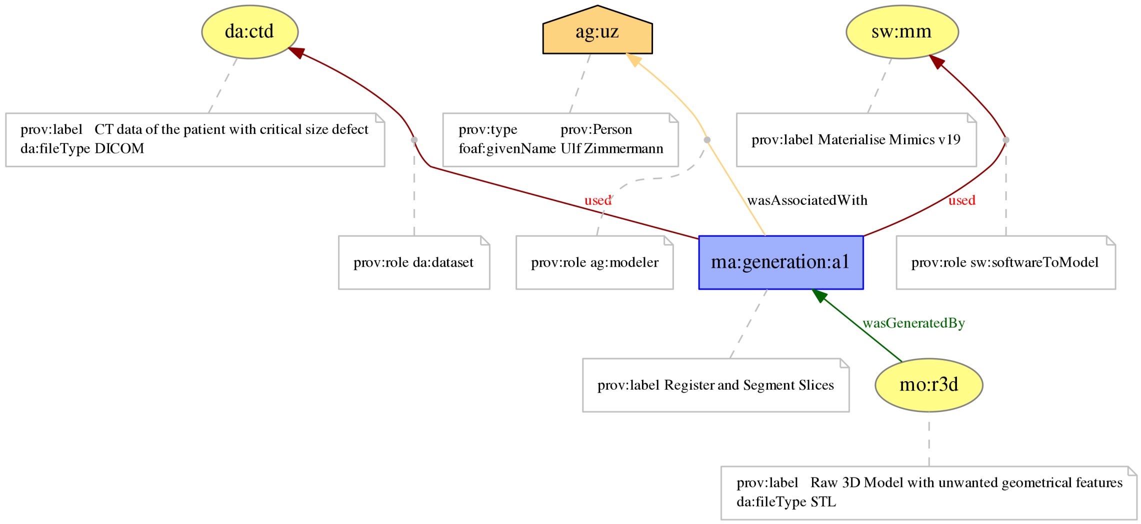

Fig. 2: Instance of the generation pattern encoding the use of CT data and a software tool in order to create a 3D geometrical model.

This can be easily imported into the simulation software COMSOL Multiphysics® v5.2a (COMSOL Multiphysics, RRID:SCR_014767). After importing the geometry object into COMSOL and adding some additional geometrical features (e. g., background box) and assigning the dielectric tissue properties on basis of literature data, the physics describing the system were defined. This includes the partial differential equations and the respective boundary conditions to be considered in the finite element simulations. Next, the tetrahedral finite element mesh has been defined, on which the solution of the finite element problem has then finally been computed.

The provenance model is developed by repeatedly discussing, modelling, and screen observing steps with the researchers of the original study. In the future, version control systems and workflow based systems [13] can be employed to automatically generate the provenance model during the execution of the simulation study.

B. Research Artefacts

Different types of artefacts contribute to executing a simulation study. In our provenance model, the type of an artefact is indicated as a prefix of its ID. Within the bio-electric stimulation example process, we identified the following types of artefacts that have a broader perspective than those used in e. g., [14]:

- **Software** (`sw`): including proprietary software such as COMSOL Multiphysics® and GEOMAGIC Studio.
- **Data** (`da`): generated e.g. by CT scans.
- **Models** (`mo`): including 3D geometry models, physical models, and simulation models
- **Equations** (`eq`): mathematical descriptions, e. g., Maxwell's equations.
- **Assumptions** (`as`): including constraints and conditions, e. g., the smoothness of surfaces in 3D geometries.
- **Literature** (`lt`): which might refer to theories, models and data of relevance to the simulation study, e. g., Gabriel et al. [15] and Kraus [16].

In addition, we also keep track of the researcher and institutions, called the agents (`ag`), who are responsible for generating the artefacts. This is especially interesting, as the simulation model was developed by two scientists who build upon the simulation models of each other. Activities are prefixed by `ma` representing "manual activity" which is performed by a researcher.

C. Provenance Model

The overall number of entities in the provenance model for the example process is 27; 13 activities are needed and two agents are involved. The provenance model is publicly available at `http://dx.doi.org/10.5281/zenodo.2650138` It is structured into three sub-models[2]:

a) Mathematical and Theoretical Knowledge Extraction (`context_models`): Theoretical as well as mathematical background knowledge is needed in order to create the final simulation model. The extraction and refinement of this knowledge is depicted in this provenance model. Note, this provenance model also includes the COMSOL process, even though this is reduced to a single activity. Figure 1 illustrates the extraction of information from the literature.

b) Geometrical Modelling (`geometry`): In this provenance model, the process of extracting the geometry model of the mandible from CT data is defined. Although, a technical model for the screws will usually be used, the screw geometries could be extracted from the CT data using the same process. Figure 2 shows an excerpt of the provenance information, where the first version of a geometry model is created from the CT data using the software Materialise Mimics® according to the description in Section III-A.

[2]Note that brackets contain the file name within the GitHub repository.

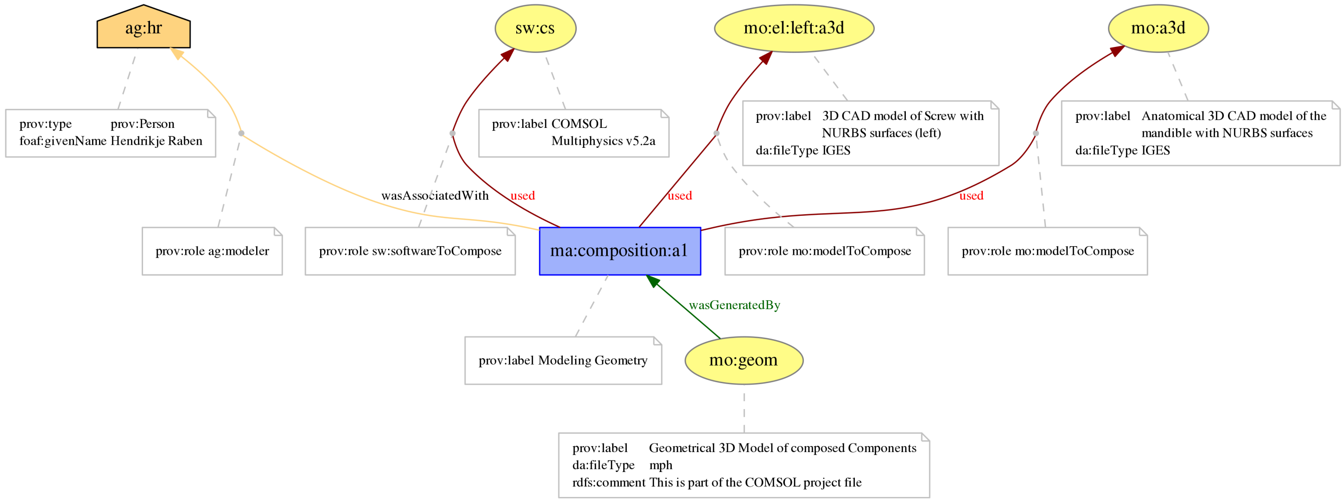

Fig. 3: Instance of the composition pattern encoding the combination of two geometrical models with the software COMSOL.

c) Preparation of the Simulation Model (`simulation_models`)*:* Composing and parameterising the geometrical model with dielectric tissue and material properties into a computable simulation model is performed using the proprietary software COMSOL Multiphysics®. Note, as we focus on model provenance, this part of the provenance model involves not the simulation activity itself, but only the final simulation model which needs to be computed. Figure 4 shows an excerpt of the model, where tissue properties are added to the 3D geometry model using the software COMSOL.

IV. Provenance Patterns

After modelling the bio-electric stimulation process, we identified four recurring patterns with respect to the modelling process and another one that is used in order to extract context information. Below, we introduce each of these pattern based on their objective, the used and generated entities as well as a description along with an example. Due to the categorisation of the entities (see Section III-B), the role of the entities within an activity is implicitly defined.

In order to provide a comprehensive view on the modelling process, the origin of information entities such as assumptions or equations has to be defined. This is the task of the following information extraction pattern:

Name	Information Extraction
Objective	Extract information from the literature
Entities	*used:* literature *optional use:* data, assumption *generated:* data, assumption, or equation
Description	The literature is read with the purpose of extracting information in the form of data, assumptions, or equations. By doing so, the information might be filtered due to some input data.

Figure 1 illustrates one instance of the information extraction pattern. This instance shows the use of stimulation parameters to filter information from an article written by Plonsey in order to extract electro-quasistatic assumptions. For the overall example process, we identified two additional instances of the extraction pattern which extract data instead of an assumption.

The following four patterns are responsible for the generation and modification of a model.

Name	Generation
Objective	Create an initial model from raw input data
Entities	*used:* raw data, software *generated:* resulting model
Description	The raw data are loaded into the software and an initial model is created from the data.

Figure 2 shows one instance of the generation pattern which encodes the generation of a raw 3D model from the CT data of the patient using the software Materialise Mimics®. The generation pattern, in general, is usually the first step in modelling processes. Thus, in our overall provenance model, we used it for the extraction of the bone geometry only, as illustrated in Figure 2.

Name	Composition
Objective	Combine multiple models into a comprehensive model
Entities	*used:* multiple models, software *generated:* composed model
Description	The models are loaded into the software and combined with respect to their purpose.

An instance of the composition pattern is displayed in Figure 3. This instance encodes the composition of a 3D model of the mandible with a 3D model of the left screw within the software COMSOL Multiphysics®. Such a composition of multiple models is frequently needed, as dividing the modelling problem into more feasible sub-models is a regular approach. Within our overall provenance model this pattern appears twice: composing the geometry model and

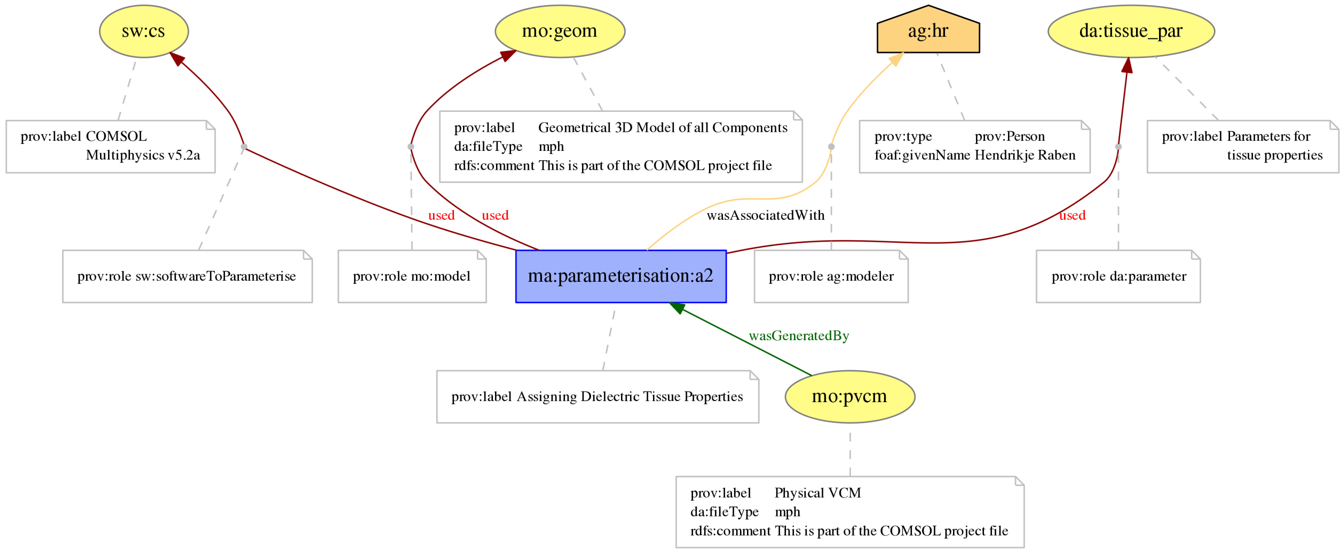

Fig. 4: Instance of the parameterisation pattern within our example process. It models the parameterisation of the geometrical model with dielectric tissue properties within the COMSOL software.

composing the physical electrical model.

Name	Parameterisation
Objective	Integrate parameter values into a model template
Entities	*used:* model, parameter data *generated:* parameterised model
Description	The model is loaded into the software and parameterised.

The augmentation of the geometrical model with tissue properties is one example for the appearance of the parameterisation pattern (Figure 4). All in all, the parameterisation of a model is used four times in our process. In general, the instantiation of a model by parameter values for a specific purpose is often used in numerical modelling.

Name	Refinement
Objective	Refine model using software and assumptions
Entities	*used:* model, assumption *optional use:* software *generated:* refined model
Description	The model is modified into a “better” model by optionally exploiting some assumptions and using software. This includes also refinement of mathematical models in the form of equations.

Figure 5 shows one instance of the refinement pattern. Within this instance, the raw 3D model is refined in the software GEOMAGIC Studio under the assumption, that the surfaces are clean, e. g., they do not contain holes. As a result, a smoothed and cleaned 3D model is created during this activity. In our provenance model this pattern is used two additional times.

V. Conclusion

We have successfully translated the process of the numerical modelling and finite element simulation study of electrical stimulation at alloplastic reconstruction plates into a W3C *PROV* provenance model by repeatedly discussing, modelling, and screen observing with the researchers of the study. During the provenance modelling, we revealed several types of research artefacts and categorised them into software, data, models, equations, assumptions, and literature. Furthermore, we identified four process patterns with respect to the simulation model development and another one for the information extraction. Though, the presented provenance patterns and research artefact categories are identified from the investigations of van Rienen et al. [11], they provide a first step for re-use or adaptation respectively, in order to represent other processes in that scientific field as provenance models. By providing such provenance models, further (semi-)automatic analyses can be applied, e. g., parameter usage and consistency checks. As such, the analysis in this paper can be seen as the first two phases of the PrIMe methodology proposed by Miles et al. [17] which describes how to develop provenance-aware applications. However, we developed the provenance model retrospectively and did not integrate the automatic recording of the provenance information in the application, which corresponds to phase three within the PrIMe methodology. In order to automatically record such provenance information in the future, the templating mechanism proposed by Moreau et al. [18] can be employed. The identified patterns, then, correspond to individual templates.

References

[1] B. Hanson, A. Sugden, and B. Alberts, “Making Data Maximally Available,” *Science*, vol. 331, no. 6018, pp. 649–649, Feb. 2011.

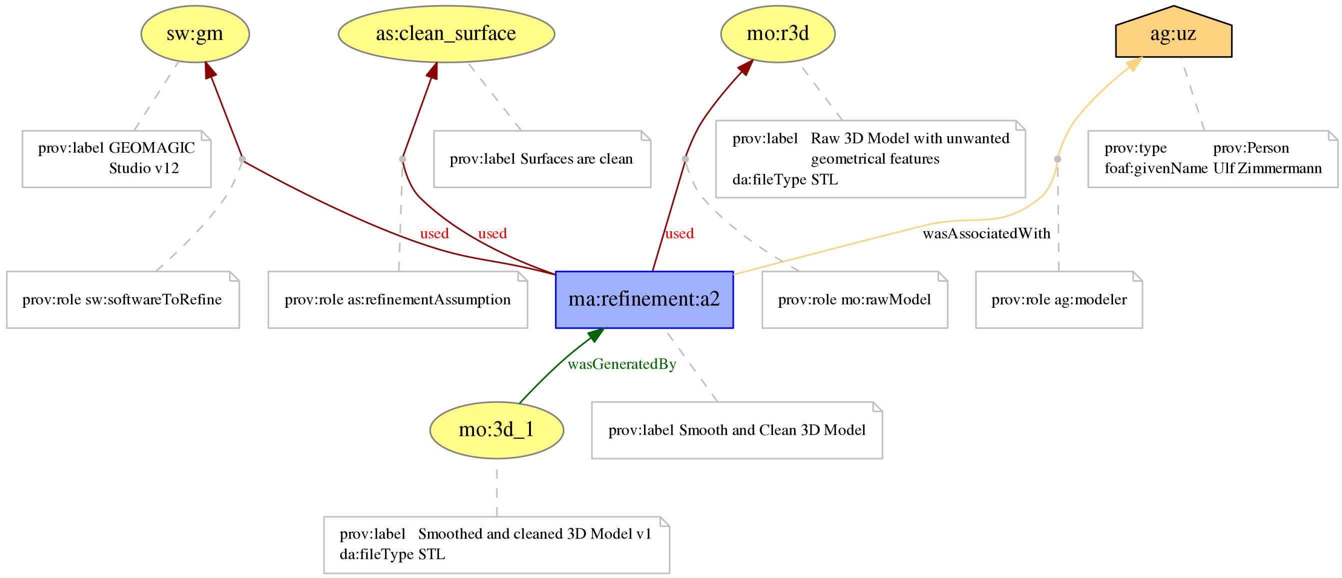

Fig. 5: Instance of the refinement pattern. The raw geometry model is refined using the software GEOMAGIC Studio under the assumption that surfaces are clean, e. g., there are no holes in the surface.

[2] R. Bose and J. Frew, "Lineage retrieval for scientific data processing: A survey," *ACM Comput. Surv.*, vol. 37, no. 1, pp. 1–28, Mar. 2005. [Online]. Available: http://doi.acm.org/10.1145/1057977.1057978

[3] P. Groth, S. Miles, W. Fang, S. C. Wong, K.-P. Zauner, and L. Moreau, "Recording and Using Provenance in a Protein Compressibility Experiment," in *Proceedings of the 14th IEEE International Symposium on High Performance Distributed Computing (HPDC'05)*, Research Triangle Park, North Carolina, July 2005, pp. 201–208.

[4] M. Greenwood, C. Goble, R. Stevens, J. Zhao, M. Addis, D. Marvin, L. Moreau, and T. Oinn, "Provenance of e-Science Experiments - experience from Bioinformatics," in *Proceedings of the UK OST e-Science second All Hands Meeting 2003 (AHM'03)*, Nottingham, UK, Sept. 2003, pp. 223–226.

[5] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability," *Commun. ACM*, vol. 51, no. 6, pp. 82–87, June 2008. [Online]. Available: http://doi.acm.org/10.1145/1349026.1349043

[6] T. D. Huynh, M. Ebden, M. Venanzi, S. D. Ramchurn, S. Roberts, and L. Moreau, "Interpretation of crowdsourced activities using provenance network analysis," in *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

[7] L. Moreau, "Provenance-based reproducibility in the semantic web," *Journal of Web Semantics*, vol. 9, no. 2, pp. 202–221, jul 2011.

[8] F. Haack, H. Lemcke, R. Ewald, T. Rharass, and A. M. Uhrmacher, "Spatio-temporal Model of Endogenous ROS and Raft-Dependent WNT/Beta-Catenin Signaling Driving Cell Fate Commitment in Human Neural Progenitor Cells," *PLOS Computational Biology*, vol. 11, no. 3, pp. 1–28, 03 2015. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1004106

[9] A. Ruscheinski and A. M. Uhrmacher, "Provenance in Modeling and Simulation Studies - Bridging Gaps," in *Winter Simulation Conference (WSC 2017)*. Electronic ISSN: 1558-4305: IEEE, 2017, pp. 872–883. [Online]. Available: https://doi.org/10.1109/WSC.2017.8247839

[10] O. Reinhardt, A. Ruscheinski, and A. M. Uhrmacher, "ODD+P: Complementing the ODD Protocol with Provenance Information," in *Winter Simulation Conference (WSC 2018)*, 2018.

[11] U. van Rienen, U. Zimmermann, H. Raben, and P. W. Kämmerer, "Preliminary Numerical Study on Electrical Stimulation at Alloplastic Reconstruction Plates of the Mandible," *Scientific Computing in Electrical Engineering*, Jan. 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-75538-0_1

[12] P. Groth and L. Moreau, "PROV-Overview. An overview of the PROV Family of Documents," 2013. [Online]. Available: https://www.w3.org/TR/prov-overview/

[13] A. Ruscheinski, T. Warnke, and A. M. Uhrmacher, "Artifact-based Workflows for Supporting Simulation Studies," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019.

[14] A. Ruscheinski, D. Gjorgevikj, M. Dombrowsky, K. Budde, and A. M. Uhrmacher, "Towards a PROV Ontology for Simulation Models," in *Provenance and Annotation of Data and Processes*, K. Belhajjame, A. Gehani, and P. Alper, Eds. Cham: Springer International Publishing, 2018, pp. 192–195.

[15] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2251–2269, nov 1996.

[16] W. Kraus, "Magnetic field therapy and magnetically induced electrostimulation in orthopedics," *Der Orthopäde*, vol. 13, no. 2, pp. 78–92, April 1984.

[17] S. Miles, P. Groth, S. Munroe, and L. Moreau, "PrIMe," *ACM Transactions on Software Engineering and Methodology*, vol. 20, no. 3, pp. 1–42, aug 2011.

[18] L. Moreau, B. V. Batlajery, T. D. Huynh, D. Michaelides, and H. Packer, "A templating system to generate provenance," *IEEE Transactions on Software Engineering*, vol. 44, no. 2, pp. 103–121, feb 2018.