Preparing the past for the future

Curating a daylight simulation model of Hagia Sophia for modern data infrastructures

Andreas NOBACK, Technische Universität Darmstadt, Germany

Lars Oliver GROBE, Lucerne University of Applied Sciences and Arts, Switzerland

Keywords: *Data management — FAIR Principles — simulation model*

CHNT Reference: Andreas Noback and Lars Oliver Grobe. 2020. Preparing the past for the future: Curating a daylight simulation model of Hagia Sophia for modern data infrastructures. Proceedings of the CHNT2020 Conference.

Introduction

A detailed simulation model of Hagia Sophia emerged from more than 20 years of research in its day-lighting. The challenging publication of its heterogeneous data set is planned in compliance with the FAIR Principles to foster findability, accessibility, interoperability, and reusability of its digital assets for both people and machines including AI applications.

Platforms such as those initiated by the Deutsche Forschungsgemeinschaft's Specialised Information Services program provide opportunities for such data publication. They also ask for contribution to their further development. The FAIR Principles themselves introduce new requirements for cultural heritage data. Internal workflows, separation of different types of data, annotation, and external referencing have to comply with modern research practice.

A case study for machine-readable data

The model offers itself as a case to study the role of data, its citation and interoperability in cultural heritage. It allows to evaluate the internals of a data-set that presents a realistic, non-ideal case for the curation. Approaches to transform the model from its current state are discussed as well as insights for the development of research data infrastructure and AI application.

The existing heterogeneous data-set

The data-set is currently split into static content stored on a file server, and the version-controlled simulation environment. The former comprises a photographic documentation, digitised survey plans, numeric simulation results and derived visualisations for publication. The simulation environment forms a directory tree that reflects the building's spatial structure. Its geometry is separated into objects, stored as meshes (DXF and OBJ). To assemble the model, objects are instantiated, positioned, and oriented according to triangular placeholders stored in so-called marker files. Complex objects are represented by sub-trees. Reflection and transmission properties are stored as text-based parametrisations of the improved Ward-model, or as data-driven light scattering models in XML files. Non-uniform appearance and colour are represented by calibrated image files. Routines to process the model, and to start simulations are formulated as a dependency graph interpreted by GNU make. The simulations are further guided by viewpoints and time-steps. The entire simulation environment except the imagery is stored in text-files and under version control - initially by CVS, later SVN, and now git - that tracks the history of changes, including comments and contributors, combined with fragmented observations, sources, and thoughts in text files within the directory tree.

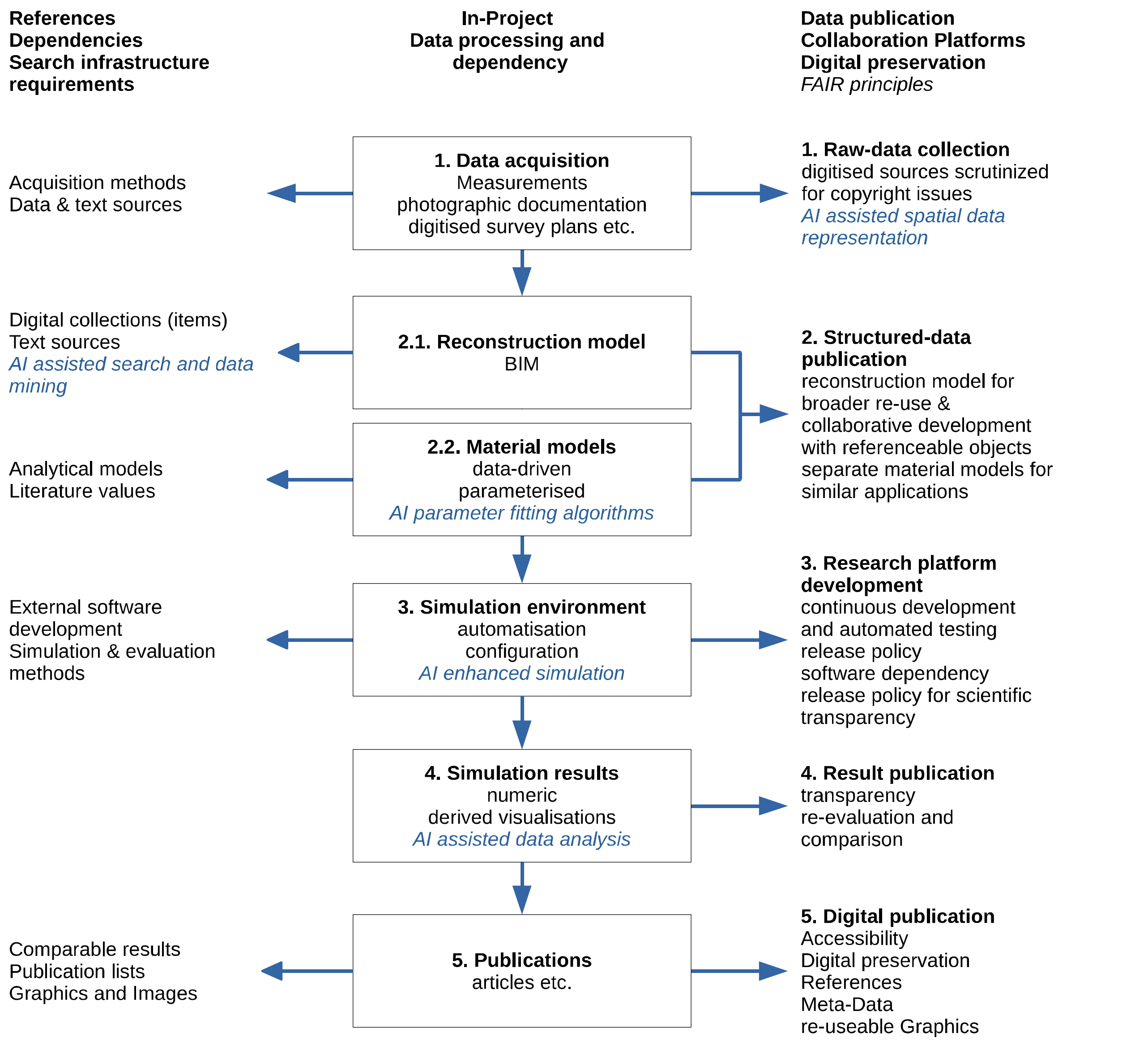
The simple, text-based triangular and polygonal mesh representation of geometric objects has supported version control, and has avoided compatibility problems over the decades. However, the geometric constraints of the floor plan and the volumetric properties of the interior space can be efficiently edited only by more complex geometric entities. Therefore, editable CAD files of multiple proprietary formats were kept in parallel to the simulation model. These do not support versioning and separation into the model’s tree-structure. Data has been lost due to conversion between different versions of the CAD-Software. This further hindered a structured documentation of the reconstruction efforts.

A concept with five data scopes for data management and AI application

In face of the heterogeneous research data and its development it is evident that no single approach fits all needs, without developing a very specific application -- and ontology -- that lacks generality and hinders access. Further examination of different requirements for collaborative development, scientific transparency and documentation, machine-readable annotation and dependencies to digital collections or external software-development lead to a separation into five data-scopes for the project and digital publications as depicted in Fig. 1: 1. Raw data collections, 2. Reconstruction and material models, 3. Simulation environment, 4. Simulation results and 5. Digital publications.

This separation is compatible with the data-flow and dependencies within the project, and is helpful for the integration into existing infrastructures or ontologies. It provides opportunities to find common solutions to similar problems in other projects and helps to define requirements for multi-purpose publication infrastructures and possible AI applications. For example, digital collections and object catalogues follow similar structures and allow a standardised publication format and AI assisted search tools.

References to objects in the structured data-models and suitable meta-data in all publications will enhance these possibilities and provide additional transparency. Internal and external linking between objects in the different data-scopes enables knowledge graphs as a form of scientific text,



and defines anchor points for search infrastructures. For example, the photographic documentation can be ordered with AI assistance for spatial reference, search and relation to other image sources. Finally, the separation allows a step-by-step approach for the demanded publication of all research objects.

With the goal to enhance the ability of machines to automatically find and use data, the presented concept provides the opportunity for a more general discussion about data management and exchange in similar cultural heritage projects, the application of the FAIR Principles, and the further development of collaboration platforms such as the aforementioned Specialised Information Services.

Fig. 2. Proposed concept of five scopes for in-project data-processing, necessary references, publication demands and possible AI application.

References

Deutsche Forschungsgemeinschaft, DFG (2019). Evaluation of and Statement on the “Specialised Information Service” Programme. Available at: https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/informationservice_science/ (Accessed: 17 September 2020).

Grobe, Lars O., Andreas Noback, and Mehlika Inanici (2020). ‘Challenges in the simulation of the daylight distribution in late antique Hagia Sophia.’ *International Hagia Sophia Symposium*. Istanbul, Turkey. under review.

Wilkinson, Mark D. et al. (2016). ‘The FAIR Guiding Principles for scientific data management and stewardship’. *Scientific data*, 3, pp. 160018– 160018. https://doi.org/10.1038/sdata.2016.18