Simulation Studies of Social Systems – Telling the Story Based on Provenance

Oliver Reinhardt [1], Toby Prike [1], Martin Hinsch [1], Jakub Bijak [1], and Adelinde M. Uhrmacher [1]

[1]Affiliation not available

October 30, 2023

Abstract

Social simulation studies are complex, because they typically combine various sources of data and hypotheses, that are integrated by intertwined processes, of model building, simulation experiment execution, and analysis. Various documentation approaches exist that support transparency and traceability of social simulation studies. The exploitation of provenance standards allows for making the information about what sources and activities contributed to the generation of an entity, e.g., simulation model, queryable and computationally accessible. Therefore, provenance patterns have been defined to capture central activities and entities. Activities include model building, calibration, analysis, and validation. Entities are simulation model, simulation experiment (its specification), and research question. Here we refine and extend this approach to address specific challenges of social agent-based simulation studies, i.e., activities such as collecting and analyzing primary data about human decisions, or collecting and assessing the quality of secondary data. This allows us to tell the whole story of these simulation studies in a comprehensive manner. We illustrate the potential of the approach by applying it to central activities and results of the Bayesian Agent-Based Population Studies project and implementing it in a web-based provenance tool.

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Simulation Studies of Social Systems – Telling the Story Based on Provenance

OLIVER REINHARDT[1], TOBY PRIKE[2], MARTIN HINSCH[3], JAKUB BIJAK[4], AND ADELINDE M. UHRMACHER[1]

[1]Institute for Visual and Analytic Computing, University of Rostock
[2]School of Psychological Science, University of Western Australia
[3]MRC/CSO Social and Public Health Sciences Unit, University of Glasgow
[4]Department of Social Statistics and Demography, University of Southampton

Corresponding author: Oliver Reinhardt (e-mail: oliver.reinhardt@uni-rostock.de).

This research was funded by the European Research Council (ERC) via the research project Bayesian Agent-based Population Studies (grant number 725232) and the German Research Foundation (DFG) via the research project Modeling and Simulation of Linked Lives in Demography (grant number UH-66/15).

ABSTRACT Social simulation studies are complex, because they typically combine various sources of data and hypotheses, that are integrated by intertwined processes, of model building, simulation experiment execution, and analysis. Various documentation approaches exist that support transparency and traceability of social simulation studies. The exploitation of provenance standards allows for making the information about what sources and activities contributed to the generation of an entity, e.g., simulation model, queryable and computationally accessible. Therefore, provenance patterns have been defined to capture central activities and entities. Activities include model building, calibration, analysis, and validation. Entities are simulation model, simulation experiment (its specification), and research question. Here we refine and extend this approach to address specific challenges of social agent-based simulation studies, i.e., activities such as collecting and analyzing primary data about human decisions, or collecting and assessing the quality of secondary data. This allows us to tell the whole story of these simulation studies in a comprehensive manner. We illustrate the potential of the approach by applying it to central activities and results of the Bayesian Agent-Based Population Studies project and implementing it in a web-based provenance tool.

INDEX TERMS Computational modeling, data, provenance model, simulation experiments, social simulation.

I. INTRODUCTION

REPRODUCIBLE and interpretable simulation studies require a thorough documentation of activities, sources and products involved in this process [1], [2]. Simulation studies involve complex modeling and analytical processes in which different activities such as model building and refinement, conducting various simulation experiments, data processing, and interpretation are closely intertwined (Fig. 1). These studies often span several years. Their documentation therefore requires significant effort and has been subject of reporting guidelines such as [1]–[3]. Computational support for recording crucial information, such as data, simulation models, assumptions, research questions etc., about simulation studies includes adopting archives [4], Wikis [5], electronic notebooks [6], as well as provenance standards (and a graph-based database) [7].

Generally speaking, all efforts in documenting simulation studies are - at least implicitly - concerned with provenance, i.e., providing "information about entities, activities, and people involved in producing a piece of data or thing" [8]. The benefit of adopting a provenance standard such as W3C PROV [8] is that the various sources, activities and products of a simulation study are put into well-defined relations to each other. Its graph-based abstraction provides a historical and causal delineation of what contributed to a simulation model and how it did so in a simple and formal manner [9]. It can be mapped into a graph data base which allows - in addition to storing the information - filtering and querying the stored information on demand [10]. Its graph-based visualization, e.g., in a web-based tool, makes it possible to easily access and assess dependency structures within and across simulation studies [7]. Provenance standards have already been applied to cell biological simulation studies [7], [11] and to documenting a migration model in demography [12].

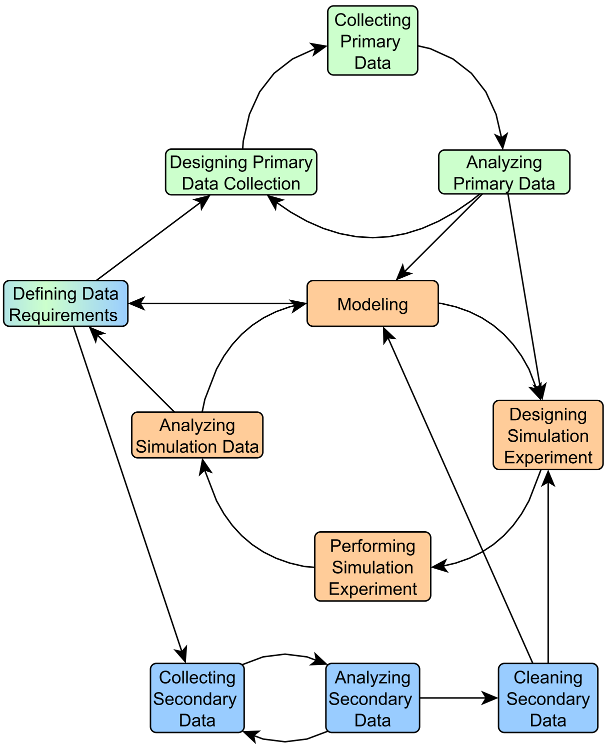


FIGURE 1. Central activities of the modeling and simulation lifecycle - including the procurement of data.

In demography, as is common for a social science discipline, the need to combine hypothesis-driven and data-driven modeling adds to the complexity of simulation studies [13], and thus to the effort required for their thorough and systematic documentation. The situation becomes even more complicated whenever primary data about human behavior are collected, e.g., through interviews or psychological experiments as part of a broader agent-based simulation study, or once uncertainty of the used secondary data is taken into account, which requires specific evaluation schemes. This is especially relevant in the context of the paradigmatic shift in demography towards more micro-level and multi-level studies [14], and the recognised need for greater use of simulation models to enhance the theoretical base of the discipline [15]. This is also in line with the general developments in social simulation more broadly, which explicitly recognize issues such as data quality, necessity of collecting bespoke primary data for simulations, and so on [16]. So far, the diversity of sources, products and processes has hampered a systematic and accessible documentation of entire demographic simulation studies.

To systematically and accessibly document entire demographic simulation studies, which also include extensive data evaluation, analysis, adaptations and their collection, we will present an approach based on provenance patterns specified in the W3C PROV standard. In [17], provenance patterns have already been identified for the documentation of and reasoning about central activities of simulation studies such as model building, refinement, verification, calibration, and validation. These patterns shall now be extended to capture data evaluation schemes to assess the quality or uncertainty of data sources, and the conducting of psychological experiments or interviews to support the agent-based modeling of human decision processes. The patterns will take reporting guidelines in the respective areas into account. We will use and adapt the tool WebProv, which combines a web-based visual interface and the graph-data base Neo4J to store and retrieve provenance information based on these patterns.

We will demonstrate our approach by applying it to the research project BAPS (Bayesian Agent-Based Population Studies[1]), which aims at "transforming the study of migration – one of the most uncertain population processes – in the way it can be understood, predicted, and managed". The development of a simulation model to analyze the formation of migrant routes from Syria to Europe, is complemented by a framework for assessing existing secondary data and their quality [18], and the acquisition of primary data by psychological experiments on human decision making under uncertainty [19], and by carrying out ethnographic interviews to provide richer contextual information [20].

The contributions of this study are threefold: 1) to identify crucial activities and entities for documenting data acquisition, quality assessment and psychological experiments and to encode them as patterns in a provenance standard, 2) to integrate this information with previously identified patterns for conducting simulation studies, 3) to apply the patterns to the activities and results achieved within a major research project on migration studies to provide a comprehensive documentation of the research done in this project. As a proof of concept we implemented the the approach based on a web-based provenance tool accessible at https://doi.org/10.5281/zenodo.6786191.

II. CASE STUDY: DEVELOPING AN ABM OF MIGRATION ROUTE FORMATION

Migration is a highly complex and uncertain population process, being driven by the decision making of individuals and various levels of institutions. Migration routes are highly volatile, with the flows responding to sometimes very rapid changes in various migration drivers, their broader environments and individual circumstances [21]. In the study shortly presented here, agent-based simulation is applied to improve the theoretical understanding of human migration with a focus on the question of how migration routes are established and sustained.

The core of the study [22] is the development of an agent-based model of migration route formation [23]. Therein, modeled migrants try to traverse an abstract landscape based on limited and uncertain information about locations on the way, potential paths, and the involved risks. As model development is an iterative process [24], multiple model

[1] https://baps-project.eu/

versions were designed in succession, informed by knowledge from the scientific and non-scientific literature on the migration process, knowledge about decision making, and lessons learned from previous iterations. For the latter, extensive simulation experiments with the model were necessary. For example, Gaussian Process emulators where fitted to the model version to assess sensitivity to the input parameters, and the uncertainty of the results. While earlier model versions were very abstract and theoretical, later versions were designed and calibrated to capture the reality of migration routes in the Mediterranean. Thereby, a considerable amount of data was integrated into the model.

Data in general, especially migration data [25], tends to be difficult to compare, and may in some cases be incomplete or of dubious quality. Hence, an important part of the project was the assessment of available data on asylum migration. To this end, an assessment framework was designed, and applied on various potentially useful sources of migration data [18], so that the data were supplemented with the meta-information about its quality, that is necessary to use it in the simulation study. This migration data from secondary sources is complemented with information on the migrants' decision processes elicited in psychological experiments and interviews, designed to answer specific questions that arose during the modeling work. For example, the models highlighted information sharing and the trust in such information as a key influence in the formation of migration routes. In a psychological experiment, data on the subjective judgment of migrants based on different kinds of information and sources was collected. The results were then used to inform the parameterization of the model [26].

This case study highlights that simulation studies of complex social systems are themselves complex and intertwined processes that include the modeling work itself, the execution of simulation experiments, the collection and assessment of secondary data sources, and the collection of new data to inform the model, as shown in Fig. 1. Broader philosophical underpinnings of such a model-based approach, within which the iterative model development is situated, are discussed in more detail in [22, Chap. 2].

Thus, diverse activities, data as well as information sources contribute to the products of such a simulation study, which again depend on each other. Each of these products can only be interpreted if their context of generation is fully taken into account. Therefore, an accessible and thorough documentation of simulation studies becomes of utmost importance, also for replicability purposes.

III. REPORTING GUIDELINES

The wish to reproduce, interpret, and reuse the results of simulation studies has led to various reporting guidelines in different application fields [3], [27].

Some focus on specific activities or products of a simulation study: on the simulation model, e.g., MMRR (Minimum Model Reporting Requirements) respectively PMRR (Preferred Model Reporting Requirements) for systems dynamics models [3], or ODD for agent-based models [28], [29]), or on simulation experiments, e.g., MIASE (Minimum Information about a Simulation Experiment) [30] and MSRR (Minimum Simulation Reporting Requirements) respectively PSRR (Preferred Simulation Reporting Requirements) [3]. Others, such as TRACE [1] or STRESS, [2] aim at documenting the entire simulation study and thereby covering all of the essential steps, sources and products of a modeling and simulation life cycle [31], [32]. TRACE structures the documentation according to crucial activities in conducting a simulation study, or *how* the simulation study has been executed, such as, problem formulation, model description, data evaluation, conceptual model evaluation, implementation verification, model output verification, model analysis, and model corroboration. In contrast, STRESS emphasises the *what* of a simulation study in terms of objective, (model) logic, data, experimentation, implementation, and software availability. In these specifications other reporting guidelines for activities or products can be reused. For example, in the case of an agent-based model, the model description in TRACE can rely on the ODD protocol, and for the model analysis, such as sensitivity analysis, on MSRR.

Wilsdorf et al. [17] mapped those considerations into the provenance standard W3C PROV, by identifying entities such as research question, simulation model, simulation experiment (specification), simulation data, data, requirement, qualitative model, and assumptions, and related those entities by activities such as creating simulation model, refining simulation model, reimplementing simulation model, calibrating, analyzing, and validating simulation model.

In all reporting guidelines of simulation studies information about the used data is required, e.g., in its checklist STRESS asks for details of data sources, input parameters for base runs of the model and scenario experiments, assumptions and data pre-processing. The latter refers to any manipulation of the data that occurred. In the TRACE documentation the data evaluation should provide insights into the quality and sources of numerical and qualitative data that have been used to parameterize the model.

While the use of data is generally included in these documentation standards - and even put into the focus by some [33], [34], the procurement of data is usually not. For more detailed provenance information on the data, the approaches for documenting and recording information about simulation studies, can rely on reporting guidelines for data acquisition and generation in the respective application field. In the social sciences this may include quantitative and qualitative data about various subjects, e.g., data from psychological experiments, interviews, surveys, or from official sources. The replication crisis in psychology, and the subsequent focus on uncovering questionable research practices in psychology and empirical research more broadly, led to the development of several suggestions and guidelines for how to document and improve rigour in empirical research [35], [36]. For primary data collection, these practices include: making collected data and analysis code publicly available, publicly

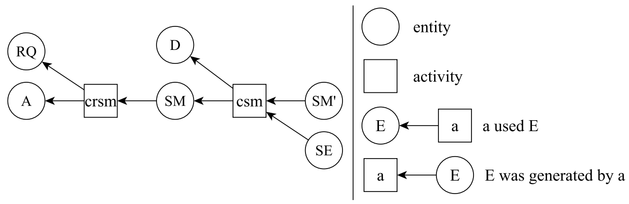


FIGURE 2. An example of a provenance graph as defined by PROV. The graph shows two typical activities in a simulation study. First (crsm) a simulation model is created, based based on a research question (RQ) and an assumption about the modelled system (A). This produces a model (SM). The model is then calibrated (the activity csm) against some data (D), producing a calibrated model (SM') and a specification of the performed calibration experiment (SE).

posting the study materials and procedure, and preregistering study protocols and analysis plans ahead of time [37]–[39]. Although some of these practices are not directly applicable to secondary data collection and analysis, practices such as sharing analysis code and clearly specifying analysis plans ahead of time are also strongly recommended for improving the transparency and rigour of research relying on secondary data [40], [41].

IV. PROVENANCE MODELS AND PROVENANCE PATTERNS

Provenance of a simulation model documents the process of creating the model, e.g., what questions it was designed to answer, on which underlying theory and data it is based, how it was constructed, and how it was experimented with. This back-story of a model is crucial to interpret and reuse a model, and to assess its quality, and the quality of the results generated with it.

Following the W3C PROV standard [8], provenance information can be represented as a directed acyclic graph with two types of nodes: entities and activities. Edges between entities and activities relate the two (see Fig. 2), specifying which entities were generated by or used by which activities.

Applying PROV requires specializing the PROV Data Model by specifying types of entities and activities, and possible relations between them. For the modeling and model analysis, the central part of a simulation study (see Fig. 1), important entities and processes have been identified in the literature [7], [9]. Building on this, Wilsdorf et al. [17] identified *provenance patterns* arising in a simulation study: certain activities within a simulation study will always use and produce certain types of entities. A pattern consists of an activity at its center, and the types of entities that are used and produced by this activity. For example, creating a simulation model (crsm in Fig. 2) will always produce a model, and calibrating a simulation model (csm) will always use a model and a calibration target, and will always produce a calibrated model and a specification of the calibration experiment. This example matches the patterns *Creating Simulation Model* and *Calibrating Simulation Model* (see Fig. 3).

We annotate entities with meta information that contains the documentation of the entities themselves. We recommend this meta information to follow established reporting guidelines for this type of entity, or refer to a document following such guidelines, e.g., to an ODD document for an agent-based model. Also, the meta information should include references to all relevant artifacts, e.g., the implementation of model or the data set for a data entity.

However, the modeling and model analysis itself, while being central, is only part of a simulation study in the social sciences, which also needs to grapple with the agency of the objects of the scientific enquiry (human beings) and the resulting high levels of uncertainty of the related social processes. In demography, migration is the one component of demographic change which – unlike fertility or mortality – does not have explicit biological underpinnings, and is thus much more challenging to analyze due to the high levels of agency of various actors, and high complexity of the underlying factors and drivers [22].

Another important ingredient is the data that grounds the model in reality - and the process of its collection. We distinguish primary and secondary data collection as follows:

Primary data was collected specifically by the conductors of the simulation study for the purpose of the simulation study itself. This collection may take the form of surveys, interviews, psychological experiments, etc.

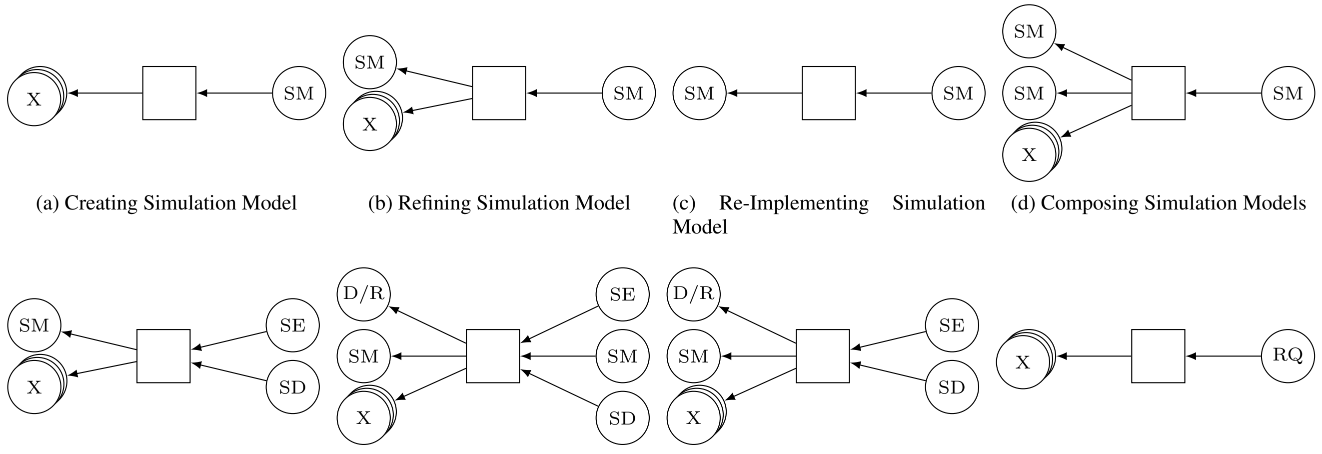
Secondary data was collected for another purpose, potentially by someone else. Hence, its suitability must be assessed, and the data may potentially need to be cleaned, to account for various sources of uncertainty and biases.

In this work, we extend the method of provenance patterns to take these processes into account. In the following, we summarize the entities and activities, and the arising patterns, in modeling and model analysis, based on [17]. Building on that, we then extend the scope of the approach by identifying such entities, activities and patterns for primary and secondary data collection.

V. MODELING AND MODEL ANALYSIS

Based on [17], we distinguish the following entities in the modeling and analysis part of the study; *Research Question* (RQ), *Simulation Model* (SM), *Simulation Experiment Specification* (SE), *Simulation Data* (SD), *Requirement* (R), *Assumption* (A) and *Other* (O). In the patterns, we also make use of an entity type *Data* (D), as a general placeholder for any kind of data, including simulation data produced in an experiment, but also primary and secondary data (which we otherwise distinguish due to different documentation requirements). Research questions may also appear in other parts of the simulation study, as they also form the basis of primary and secondary data collection.

Fig. 3 shows the patterns based on Wilsdorf et al. [17] graphically. The upper four patterns describe activities in the modeling process. When a new simulation model is created from scratch, the pattern *Creating Simulation Model* (a) applies. That activity uses various inputs, e.g., a research



(a) Creating Simulation Model (b) Refining Simulation Model (c) Re-Implementing Simulation Model (d) Composing Simulation Models

(e) Analyzing Simulation Model (f) Calibrating Simulation Model (g) Validating Simulation Model (h) Identifying Research Question

FIGURE 3. Provenance patterns for model development and model analysis activities (a to g from Wilsdorf et al. [17]).

question, assumptions, theories or data, represented by the wild-card (X) in the pattern, and produces a simulation model (SM). When an existing model is refined, the pattern *Refining Simulation Model* (b) applies instead, which has an additional input in form of the existing simulation model. For example, the model from the case study was refined when new data from psychological experiments became available, that could be used to improve the decision-making mechanisms. The pattern *Re-Implementing Simulation Model* (c) refers to an activity, where a simulation model is re-implemented in another language or tool, without refining or extending it. For example, in our case study, we re-implemented the model, originally implemented in Julia, in the modeling language ML3 to cross check both models [42]. Finally, the pattern *Composing Simulation Models* (d) describes the composition of two simulation models, e.g., when we combined the Julia and the ML3 implementation of models to gain a new Julia implementation that follows a similar rule-based approach as the ML3 implementation.

The patterns (e) to (g) describe activities during the analysis of and experimentation with a simulation model. When a simulation model is analyzed (the pattern *Analyzing Simulation Model* (e)), e.g., via a sensitivity analysis or uncertainty quantification, a simulation model (SM) is used as well as potentially some other inputs (X). The result is some simulation data (SD), and a simulation experiment specification (SE), e.g., a script or description that allows the analysis to be repeated. For example, when one of the models developed in the case study was subject to uncertainty and sensitivity analysis (see [22] Chapter 8.3), that model served as the (SM) input. As this analysis was a repetition of an earlier analysis performed on an earlier model version, that earlier simulation experiment specification was used as an additional input (X). The uncertainty and sensitivity characteristics determined through the experiment are referred to in the simulation data entity (SD). The cited chapter may serve as a reference in the simulation experiment specification entity (SE). If there were scripts or other software artifacts to repeat the experiment, it would be preferable to reference these in the entity (SE). However, this analysis was performed with a GUI-based tool, so there were no scripts produced. The pattern *Calibrate Simulation Model* (f) for model calibration is similar, but requires an additional input: some data or a requirement (D/R) that serves as the calibration target. A calibrated simulation model is produced as an additional output. The pattern (*Validating Simulation Model* (g)) is defined in a similar way to the calibration pattern. The only difference of the pattern is that a validation does not produce a simulation model. For validation, the model behavior is compared with the data or requirement (D/R) and the results are stored as simulation data (SD).

Compared to Wilsdorf et al. [17], we added one additional pattern *Identifying Research Question* (h). Newly identified research questions are often a major driver of long term simulation studies - as well as important results. For example, the modeling work may identify gaps in the data, that lead to research questions for data collection efforts or the collected data may show interesting properties that pose new questions. In general any entity (or combination of entities) might lead to new questions. Hence, the pattern for *Identifying Research Question* allows any input (X) to produce a research question (RQ). For example, the results of a sensitivity analysis of the model highlighted the importance of information sharing for the model behavior. This lead to a new research question for a psychological experiment, which, in this case is, how migrants judge information recieved from different sources.

As with other reporting guidelines for simulation studies, the provenance patterns can and should be combined with documentation guidelines for describing central products of simulation studies, e.g., ODD or MMRR for simulation models, or MIASE or MSRR for simulation experiments. As meta information or attributes for all entities can be defined on demand, an entity's attributes can easily be adapted to reflect the considerations of other documentation guidelines.

In addition (or in the preferred case [3]), meta information should be combined with accessible and executable versions of simulation models and simulation experiments. Thus, each entity of type simulation model or simulation experiment would refer to an openly accessible code repository. The accessibility and transparency of code can be enhanced by exploiting developments in specifying executable simulation experiments, including domain-specific languages [43], [44] or model-based experiment designs [45].

As in TRACE or STRESS, the purpose of the simulation study has been identified as a central piece of information for assessing and reusing the products of a simulation study, called research question, problem formulation, and objective respectively. According to TRACE, suitable attributes of the entity research question RQ would be 1) *context* in which the model will be used, 2) the types of *model clients* or stakeholders, 3) research *questions* to be answered, 4) *model outputs* to be observed and, ideally, 5) an *applicability statement* that explains the domain to which the model can be applied to.

Whereas TRACE and STRESS documentations assume one purpose and one single simulation model, we assume that different versions of simulation models and even research questions can belong to the documentation of a simulation study. The provenance graph can be seen as a "model" describing the generation process of a simulation model, in terms of defining activities such as model creation, refinement, and composition, as well as the generation of research questions and their interrelations with sources and (intermediate) products of the simulation study. The unique perspective on the simulation study that our approach offers also becomes evident if we look at activities such as model analysis, calibration and validation. It should be noted that STRESS guidelines deliberately exclude the validation of the simulation model, and calibration is not on the check list of STRESS guidelines. In contrast, TRACE acknowledges the importance of calibrating (output verification), analyzing (model analysis), and validating (output corroboration) the simulation model by dedicating specific sections of the documentation to these activities. However, TRACE does not clarify how to document calibration, analysis or validation activities. Here, our provenance patterns provide further structure and support. They require identifying the simulation model, and any information sources or (intermediate) products that were used for analyzing the simulation model. Data or requirements are specifically requested for calibration and validation patterns. In addition to the simulation data, and, in the case of calibration, the simulation model, the simulation experiment specification constitutes a vital product of calibration, analysis and validation, and thus demands the careful attention of the modeller (see discussion above).

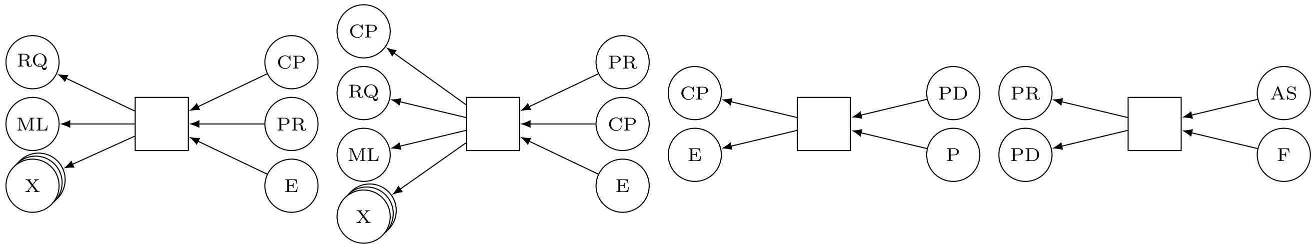
VI. PRIMARY DATA COLLECTION

Primary data collection includes the design of a collection procedure, the execution of the collection procedure to gather data, and the analysis of the data to gain insights.

We begin by defining the relevant types of entities, using the case study experiment on migrant judgments of information from different sources as an exemplar, and then continue with defining patterns for the activities.

- **Methodology Literature (ML):** When designing a data collection procedure, researchers often rely on reusing or adapting methodologies from existing research. By including information about key papers that have informed the data collection procedure, other researchers are better able to understand, reproduce, and assess the data collection procedure as well as the primary data and findings that are generated. For example, in the psychological experiment on migrant judgments and decisions there were two key papers that informed the methodology. One was a review paper by [46] on how source impacts how people assess and make judgments and decisions based on information. The other was a paper by [47] that examined how people convert different verbal likelihood statements into numerical judgments.
- **Data Collection Procedure (CP):** The data collection procedure determines what data will be collected and how. Depending on the form of data collection, e.g., a survey, interviews, psychological experiments, this entity may take different forms, e.g. a questionnaire, interview questions and instructions for interviewer, or even a piece of interactive software that is presented to participants. In any case, when presented to the participants, the Data Collection Procedure allows data collection to be undertaken. The primary data collection conducted in the case study used a psychological experiment setup within the Qualtrics[2] survey software. This Data Collection Proceduce collected data on how people make risk related judgments and decisions in response to migration related safety information presented by a variety of sources. For an example of the data collection procedure used please see https://southampton.qualtrics.com/jfe/form/SV_20kQsSP0cyi6o06.
- **Participant Information (PI):** To allow for the assessment and reproduction of primary data collection, it is crucial to provide information about the participants included within the study. This includes information such as which populations they were recruited from, how they were recruited, and any specific requirements or exclusions that were used (e.g., language requirements, demographic characteristics, attention check questions etc.). Providing this information also allows other researchers to assess the primary data collection (e.g., whether the participants were appropriate to address the research question and support the findings) and to decide whether the data and/or findings are appropriate for other researchers to rely on or reuse (e.g., if they can be transferred to a new population of interest). Participant information about the exemplar experiment included in the current study can be accessed by looking at the demographic information included within the primary

[2]https://www.qualtrics.com



(a) Designing Data Collection (b) Designing Follow-Up Data Collection (c) Collecting Primary Data (d) Analyzing Data

FIGURE 4. Provenance patterns for secondary data collection.

data[3] as well as by looking at the preregistration[4] for information about how many participants were included and what (if any) inclusion and exclusion criteria were used.

- **Preregistration (PR):** A preregistration is a document outlining several key aspects of a study methodology and analysis plans. Some of the key details included within a preregistration are: the specific research questions and/or hypotheses, the methodology that will be used (e.g., dependent variables and independent variables/experimental conditions), the participant sample size to be collected along with exclusion or inclusion criteria, and the planned analyses that will be used to answer the research questions/test the hypotheses (for an example, see https://osf.io/3qrs8).
- **Ethical Approval (E):** Primary data collection from human participants, be it through interviews or psychological experiments, requires adherence to ethical standards that are set by the funders and institutions carrying out the data collection. Here, the Ethical Approval documents the final, approved by the relevant body, version of the research ethics application, interview/experiment schedules (Data Collection Procedure), and Participant Information and Consent forms, which sets out the conditions and standards of data collection, storage, use and re-use. The experiment conducted as part of the case study underwent ethical review and approval by the University of Southampton Ethics Committee prior to data collection being conducted (ERGO Approval: 56865).
- **Primary Data (PD):** The data is the principal result of primary data collection. Depending on the kind and scope of data collection, it may take the form of a table or a set of tables, interview transcripts or their summaries (excerpts, codes), or a data base. In any case, this entity is a representation of the raw output of the data collection, potentially anonymized or pseudonymized if that is necessary, that may then be analyzed in further steps. The primary data for the psychological experiment in the case study is available in an OSF repository[3].
- **Findings (F):** The findings refer to the key conclusions or results that are generated by analyzing the data. These can take a variety of formats, such as a written results section, graphs, tables, or descriptive statistics (e.g., means, medians, correlations etc.). These findings can subsequently be used as inputs for modeling activities in a variety of ways, including to set or inform model parameters, to help specify the direction of relationships between model variables, or to test the broader implications of findings (e.g., how a masking intervention influences disease spread through a societal network). The results section that outlines the findings from the experiment on migration related risk judgments and decisions can be found within the following paper [19].
- **Analysis Specification (AS):** To reproduce the findings, it must be possible to repeat the analysis of the data precisely, either with the same data, or with comparable data, e.g., from a follow-up study. Hence, a specification of the conducted analysis is required. Often, the analysis will be conducted with some statistics programming language or library, e.g., in R or in Python. In this case, analysis scripts, are a natural result of the analysis process, and will allow for easy analysis repetition. For example, the analyses for the case study experiment were conducted in R and the analysis scripts have been made publically available at the following link: https://osf.io/ws63f/files. If such scripts do not exists, e.g., if a GUI-based analysis software is used, or if the scripts are not sufficient on their own, the specification of the analysis may also be textual.

We identified the following patterns for the activities involved in primary data collection (see Fig. 4).

(a) **Designing Data Collection:** Before any data can be performed, the data collection must be designed. We identify three core products of the design phase: the data collection procedure, the preregistration, and the ethics document. One example of (X) is the literature from the substantive area of relevance (in this case, cognitive studies of human decision making)

(b) **Designing Follow-Up Data Collection:** Some data col-

[3]https://osf.io/cqh6j
[4]https://osf.io/3qrs8

lection efforts are designed as to follow up on a previous one, e.g., to replicate the result, to refine the procedure, or to answer new questions raised by the findings. In this case, the data collection procedure (CP) of the original experiment is an additional input when designing the follow-up experiment. Including the original data collection procedure as an input connects the follow-up data collection to the original data collection and show how the data collection procedure has been refined across multiple rounds.

(c) **Collecting Primary Data:** Once the data collection is designed, the data can then be collected. Participants are recruited, and the data collection is executed with them, e.g., they are given the survey, or are interviewed. This process is based on the previously designed data collection procedure (CP) and must conform to the ethics document (E). Hence, both are inputs to this activity. The product are the collected data (PD), as well as information about the recruited participants (PI).

(d) **Analyzing Data:** When the data was collected, it must be analyzed. Apart from the data itself (PD), the preregistration (PR), containing the planned analyses, is an input for this activity. The activity produces two outputs: the findings (F), and analysis specification (AS). While the preregistration contains plans for the analysis, the actual analysis may still differ from it, especially in the case of exploratory studies or if unexpected issues emerge (e.g., parametric analyses are not appropriate so non-parametric analyses are used instead). Researchers may also wish to explore additional research questions or test the robustness of their results by using additional unplanned analyses. Changes to analysis are perfectly understandable and often recommended, but it is important there is clear delineation between pre-planned confirmatory analyses and exploratory analyses. As it only becomes apparent what is actually analyzed - and how it is done precisely - during the activity, the analysis specification (AS) is produced as part of this activity.

As previously highlighted, there are multiple existing proposals and recommendations for how to improve the rigour and transparency of primary data collection (e.g., [38]; [39]) and recent years have seen considerable advancement in the extent to which various aspects of the primary data collection process are made open and transparent [37]. However, there is still a long way to go and considerable variation exists in the extent to which these practices have been adopted by different research fields, subareas, lab groups, and even across different studies by the same researchers. For example, although there are many advantages to preregistration (PR) it is not a mandatory practice and therefore may not always be present within a primary data collection process. Similarly, although the primary data (PD) and analysis specification (AS) entities are always generated from a primary data collection process, these are not always made publicly available or available to other researchers at even upon request (e.g., see [48] about the low response rates of authors to data requests). Nonetheless, the inclusion of as many of these entities as possible within a provenance model greatly increase the ability of researchers, including those who conducted the primary data collection, to assess the robustness, reliability, and relevance of the data collected as well as any findings that were generated. This also has important flow on effects for subsequent modeling activities that incorporate and rely on the primary data. For example, further assessment, new data collection, and/or new information coming to light (e.g., failed replications; [35]), may lead the reliability and robustness of a primary data collection process to be called into question. If this primary data collection has been incorporated within a provenance model(s) then researchers can quickly and easily discover which further processes relied or built upon the questionable primary data collection. This makes it much easier to discover and reexamine or reassess whether subsequent pieces of work need to also be updated or adjusted in light of the questions raised about a primary data collection process.

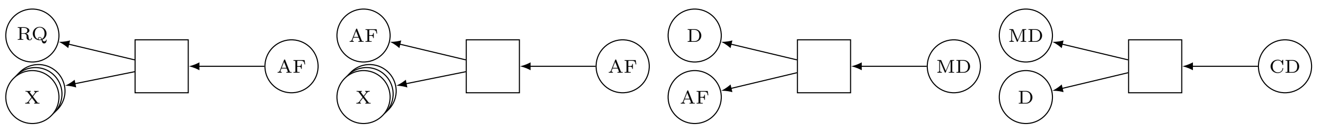
VII. SECONDARY DATA COLLECTION

Unlike primary data, secondary data is more generic - it is not collected for the purpose of a specific study. Still, such data can of course be useful (and used) for modeling. However, to make the quality of secondary data apparent, it must be assessed, based on criteria dependent on the study. Based on the results of the assessment the data might then be used as is, or it might require cleanup or transformations to address the shortcomings.

We identified the following three entity types:

- **Assessment Framework (AF):** The assessment framework defines the criteria of the data assessment, dependent on the specific simulation study. For example, the criteria for our case study were specified in [18]. Therein, a set of criteria is defined (such as fitness for purpose, trustworthiness, level of disaggregation, timeliness, completeness, accuracy, and so on). There are five levels of evaluation for each criterion, ranging from "green" where a desirable criterion is met in full, through "amber" when it is met in part, to "red" where this criterion is not met (see e.g. [49]), in our case additionally including in-between ratings (green-amber and amber-red). Some criteria are general in nature, determining the extent a given source may be useful, others are linked to the bias and variance inherent in the data source, which need to be considered for the modeling process.
- **Metadata (MD):** Metadata are properties of the dataset in question, including the values of specific evaluation ratings from the Assessment Framework (AF) given to the data source. In the migration study presented above, this meta-information is available in the online Data Inventory on Syrian Migration to Europe, 2011-21[5]. As

[5] https://www.baps-project.eu/inventory/data_inventory



(a) Creating Assessment Framework (b) Refining Assessment Framework (c) Assessing Secondary Data (d) Cleaning Secondary Data

FIGURE 5. Provenance patterns for secondary data collection.

an example, the meta-information for UNHCR data on asylum registration includes source (UNHCR), a short description, a url, time detail (daily), source type (registrations), topic (destination population of interest), data types (quantitative, process-related and macro-level, i.e. aggregate numbers), as well as seven individual aspects of data assessment, ranging from "green" to "amber", with an "amber-green" rating overall.

- **Cleaned Data (CD):** A product of transforming the initial data (D) taking into account their properties (MD), aiming at creating new variables with desired properties, such as devoid of explicit bias or with reduced variance. A migration-related example can be: if migrant registration data (D) are known to be under-reported (one of properties of MD, completeness, is rated "amber", indicating a presence of bias), then CD can include daily rates of change in registrations rather than volume of registrations, because the former would be less sensitive to the presence of systematic bias.

And the following four patterns (Fig. 5):

(a) **Creating Assessment Framework:** As the assessment framework is specific to the simulation study, its creation is the necessary first part of the assessment process. The connection to the rest of the study is realized by using the research question (RQ) as an input. Other inputs (X) may include, but are not restricted to, earlier assessment framework(s) or knowledge about limitations of the data relevant in the field, e.g., about typical problems with migration data. The product is the assessment framework (AF).

(b) **Refining Assessment Framework:** At some point during the study the existing assessment framework may need refinement, e.g., when the research question has shifted enough that the previously defined criteria no longer fit. This activity uses the previous assessment framework (AF), as well as potentially other sources (see Creating Assessment Framework). It produces a refined assessment framework (AF).

(c) **Assessing Secondary Data:** The assessment of some data is the application of the assessment framework to that data to determine the properties of the data. Hence, the assessment framework (AF) and the data (D) are used by the activity, while the metadata (MD) are produced.

(d) **Cleaning Secondary Data:** The transformation of the data (D) in the light of the data properties (MD) identified during the process of applying the assessment framework (AF), in order to produce cleaned data (CD). The process may involve e.g. removing the identified biases, smoothing data to reduce variance, applying a variable transformation to reduce other issues identified in the assessment process (such as log-transformation for strictly positive variables which exhibit exponential patterns of change), and so on.

There were existing frameworks for comprehensively assessing the different aspects of the quality of migration data according to different criteria, including those relying on the traffic-lights operationalisation (e.g. [49]). The inclusion of data assessment in a provenance model not only allows for quality checks and corrections to be formally embedded as a necessary element that secondary data need to undergo as part of the modeling process, but also enables identifying which parts of the model may be affected by potential problems with a particular data source. In the case of migration, where available data sources are notorious for their imperfections (e.g. [25]), this makes the ensuing modeling and analysis explicitly conditional on the information used and cleaning activities undertaken. It also means that, where needed, uncertainty from the data can be propogated to the model results along the paths of the provenance graph, helping with the transparency of the analysis and with honest reporting of the results and their limitations. Alternatively, the provenance sub-graphs related to data analysis and cleaning (Fig. 5) may describe a piece of analysis in its own right, should data-related question be of specific interest to the analysts or the users of a particular data source.

In TRACE the use of secondary data is documented in the "data evaluation" element; in STRESS in the "data" section. However, there, the focus is on what data was used, and for which purpose it was used in the model (which we treat as part of the modeling and model analysis, where many of the patterns either allow or require data inputs). Assessment of data, with explicit criteria, and data transformation are not considered in detail by either TRACE or STRESS.

VIII. PROOF OF CONCEPT

To demonstrate the approach, we realized a provenance model of the case study project outlined in Section II. In terms of software, we extended WebProv [7], provenance editor for simulation studies. It allows for the creation and editing of a provenance graph with a web-based interface. The provenance graph is stored in a graph database (Neo4j), which not only allows for simple and efficient storage, but also includes a powerful language for retrieving information from the database. Documents and artifacts referenced in the meta information are stored online in appropriate repositories, e.g., on GitHub or OSF.

Our extended version of WebProv and the provenance graph presented here are available at https://doi.org/10.5281/zenodo.6786191 and https://doi.org/10.5281/zenodo.6786226.

A. OVERVIEW

Fig. 6 shows an overview over the provenance graph as a whole. Fig. 7 shows a part of it in detail, in particular an instance of primary data collection: a psychological experiment to elicit subjective probability judgments migrants make based on information they gain from different sources. While the figure only shows the graph, the interactive UI displays detailed information about each entity and activity when it is selected (see Fig. 8) (often giving a concise description, some key properties and referencing the document or piece of software represented by an entity). In the activity irq3 a research question is identified (see the pattern *Identifying Research Question* Fig. 3h), based on some entities in the "Modeling and Model Analysis" area that are not displayed here. Starting from this question, a psychological experiment was designed (*Design Data Collection*, ddc2), using (ML2; referring to Briñol and Petty [46]) and ML3 (referring to Wintle et al. [47]) as methodology literature (ML) inputs - one satisfying the ML input of the pattern, and the other serving as an optional additional input (X). The results of this activity are a data collection procedure (CP2; referring to the survey[6]), the preregistration (PR2; linking to the preregistration stored on OSF[7]) and the ethical approval (E2; referring to the University of Southampton Ethics Committee, ERGO number 56865). Similarly, cpd2 and ad2 match the patterns *Collecting Primary Data* and *Analyzing Data*.

As demonstrated, the provenance graph can serve as a high level overview of the various activities of a simulation study, connecting the various inputs and outputs. For large-scale studies with many interconnected parts the graph will become increasingly large and complex, reflecting the complexity of the documented study. However, the semi-formal structured approach allows for computational processing of the provenance graph, as we demonstrate in the next section.

B. RETRIEVING DETAILED INFORMATION: QUERYING

The provenance graph does not only give an overview about the conducted simulation study, it is also rich in detailed information, linking various artifacts produced in the study. Querying allows for the retrieval of detailed information on demand. Using a dedicated graph database for storing the provenance graph, we can exploit the included querying language, in this case Neo4J's Cypher, to formulate queries effectively and have them executed efficiently. In practice, retrieving some detail requires two steps: First, a Cypher query to retrieve the provenance nodes of interest must be formulated and executed. Second, the meta-information of the nodes of interest may be inspected, either for the information itself, or to follow references to the relevant documents. In the following we show some typical questions that may be asked of a simulation study, and demonstrate how they can be answered with queries on the provenance graph.

Often not only single entities, but their context within the study is of interest - after all, putting the entities into the context of their generation and use is the point of provenance models. For example, we might want to ask for research questions that were newly asked within the study - and what they are based upon. This context can be specified in the query as a graph pattern:

```
MATCH (n {definitionId: 'Research
    Question'})-[]->(m {definitionId:
    'Identifying Research
    Question'})-[]->(k:ProvenanceNode)
RETURN n,m,k
```

Here, we query for all Research Question entities `n`, the *Identifying Research Question* activities that generated them `m`, and any entities `k` that were used in these activities. The result is displayed in Fig. 9. Please note that the initial research questions of the study are not displayed, as we specifically asked for research questions generated within the study.

The same approach also allows for querying of complex graph patterns, i.e., asking questions about relationships between entities and activities of the simulation study. One might want to know how a certain finding from a psychological experiment, e.g., the findings of a psychological experiment on the subjective judgment of migrants concerning different kinds of information and sources (the entity labeled F2), influenced the simulation models. In terms of the provenance graph this means asking for simulation models from which a path (possibly via multiple intermediary steps) leads to F2, as well as for the nodes on this path:

```
MATCH p=shortestPath(({definitionId:
    'Simulation Model'})-[*]->({label:
    'F2'})), (n)
WHERE n IN NODES(p)
RETURN n
```

Here, we use the shortest path, to only see the most direct path from any simulation model, hiding more indirect relations. The result of the query, a sub-graph of the provenance graph can be seen in Fig. 10. This shows the findings informed the building of the simulation model SM4, as well as later model versions via SM4.

As a final example, one might be interested in how the modeling work in the study was grounded on the other work conducted in the study, e.g., on the collected data. In the query, we are looking for any links from nodes in the "Modeling and Model Analysis" area to other areas of the

[6] https://southampton.qualtrics.com/jfe/form/SV_20kQsSP0cyi6o06
[7] https://osf.io/3qrs8

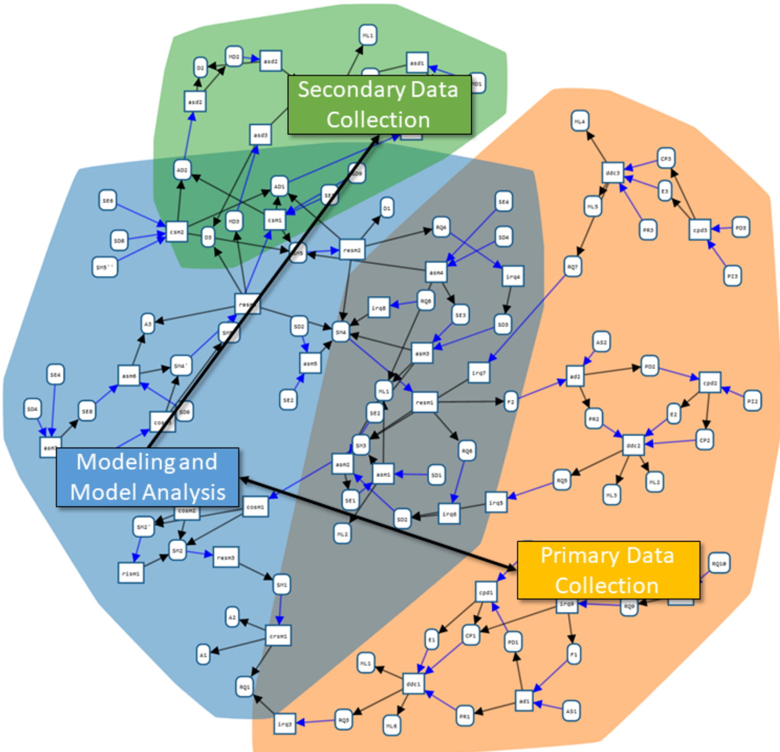


FIGURE 6. Overview of the case study provenance graph in WebProv. Each node is associated with one part of the project as distinguished in this paper: the modeling and model analysis (blue), the primary data collection (orange) and the secondary data collection (green).

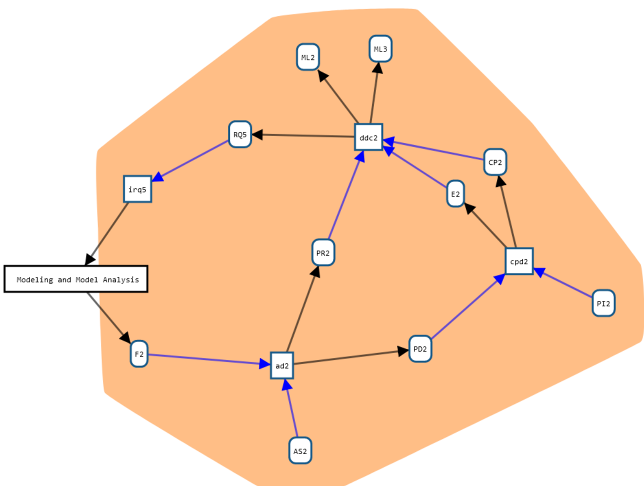


FIGURE 7. Detail of a part of the primary data collection: the experiment on subjective probability judgments. The box on the left labelled "Modeling and Model Analysis" refers to that part of the project (the blue area in Fig. 6). Arrows pointing to or from it represent provenance relations ("used" or "was generated by") with nodes in the "Modeling and Model Analysis" part. In WebProv, this box may be "opened" to display the actual relevant nodes.

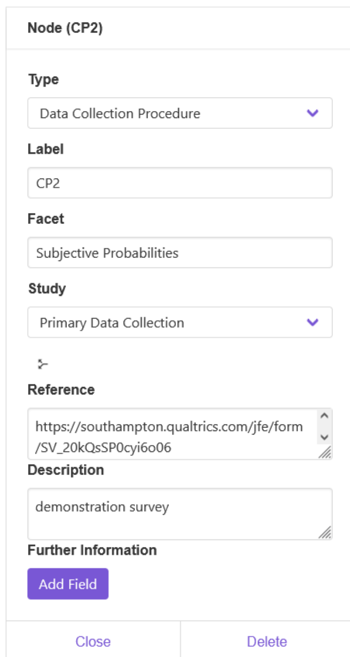


FIGURE 8. Information about the entity CP2, the data collection procedure of the psychological experiment described in the text, as displayed in WebProv. The field "Facet" allows a keyword to be entered, to aid in searching for related entites, e.g., all entites related to this experiment (see Fig. 7) have the facet "Subjective Probabilities". The "Study" field refers to the different colored areas, which are called studies in WebProv. "Reference" contains a reference to the collection procedure itself, in this case in the form of a demonstration survey identical to the one given to the participants. The "Description" field allows additional information to be summarized.

study. For the sake of clarity, we only want to display the first entity or activity outside of the "Modeling and Model Analysis" area:

```
MATCH (s:Study {label: "Modeling and
    Model Analysis"}), p=shortestPath((n
    {studyId: s.id})-[*]->(k)), (m)
```

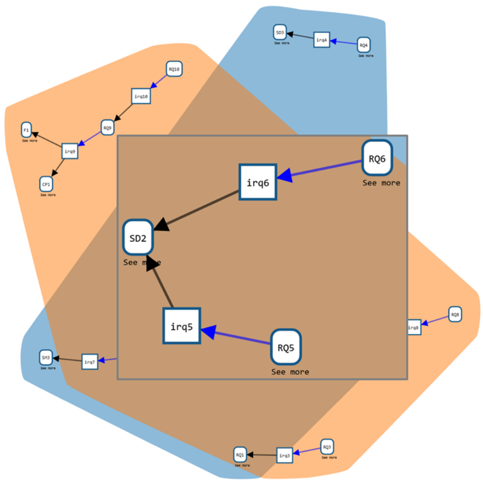


FIGURE 9. Result of a query for research questions identified in the project. The grey box shows a part of the result: RQ5 ("How do migrants make likelihood judgments?") and RQ6 ("How do risk perception and risk avoidance affect the formation of migration routes?") both follow from SD2, the result of an analysis of the model that shows that the parameters related to risk are most sensitive, requiring more research on this subject. The background shows the complete result of the query.

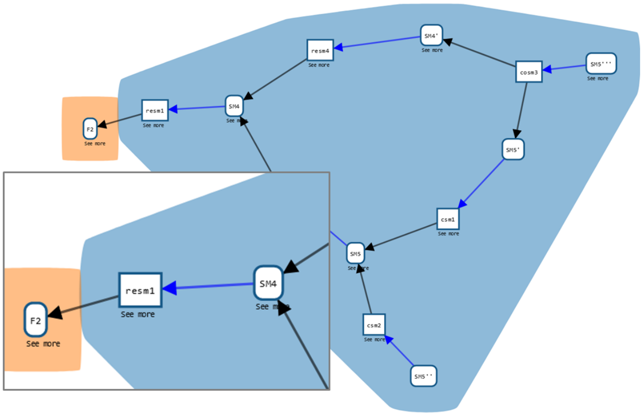


FIGURE 10. Result of the query for models that use the findings of experiment on the subjective judgment of migrants on different kinds of information and sources (F2). The findings were used in creating the model version SM4 (see the enlarged box). Through SM4, they also influenced later model versions.

```
WHERE k.studyId <> s.id AND
    exists(({studyId: s.id})-[]->(k))
    AND m IN NODES(p)
RETURN m
```

These areas are called "study" in WebProv (which is a different use of the term study then elsewhere in this paper). We identify `s` as the WebProv-study "Modeling and Model Analysis". Then we search for paths from a node `n` within the WebProv-study `s` to other nodes `k` that have a different WebProv-study id, i.e., are in a different area. Further, these nodes shall have a predecessor with the same WebProv-study id as `n`, i.e., which is in the same area as `n`. Fig. 11 shows that the modeling work was grounded on the F2 findings from the primary data collection as well as several entities from secondary data collection.

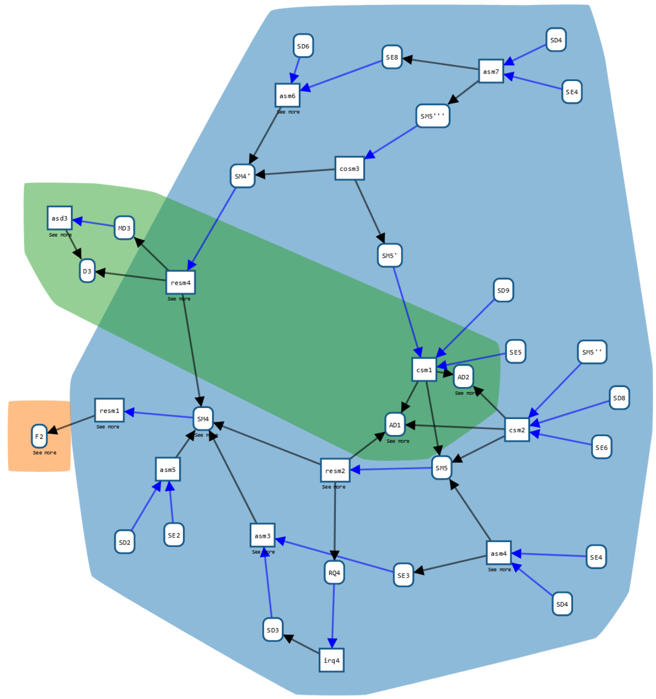


FIGURE 11. Result query for sources of simulation models outside the "Modeling and Model Analysis" area. The node on the left is F2 (as in Fig. 10), the nodes in the green area are from the "Secondary Data Collection" area.

IX. CONCLUSION

The documentation approach presented in this paper, based on provenance graphs and patterns, differs from existing documentation guidelines for simulations studies such as TRACE or STRESS in three key aspects: scope, subject and degree of formalization.

Unlike the aforementioned guidelines we treat data collection, both of primary and secondary data, as an integral part of the simulation study. Not only the the existence of data, but also its quality and suitability must be assessable to judge the foundations of a simulation model. Unlike TRACE [1] or STRESS [2], which are primarily concerned with what data was used, and where it is used in the modeling process, the provenance approach presented in this paper makes collection of primary data, and the assessment and cleaning of the secondary data explicit. The detailed documentations of both collection procedures (for primary data), and assessment criteria also aims to make the limitations of the data visible.

The provenance graph documents the process, not the product, of a simulation study. The provenance patterns we suggest do not describe a simulation model, simulation experiment, or piece of data, but it describes how they were created, what steps were undertaken, and how they relate to specific research questions they are designed to answer, to data they are based upon, and to the results they generated. Consequently, the provenance graph is not intended to replace other documentation, but to complement it. The whole graph documents the process of the study. Single entities document individual (intermediate) products, for which exisiting documentation standards such as ODD [28], [29] (for an agent based model) or MIASE [30] (for a simulation experiment) should be employed.

Unlike most documentation standards and protocols used or suggested for social simulation, which are textual, we propose a more structured semi-formal approach. This aims to make the documentation more accessible for computational processing. We demonstrate some of the benefits in Section VIII, by using graph-queries to retrieve information about the simulation study. This benefit is not restricted to the consumer of the documentation: Wilsdorf at al. [17] demonstrate how the provenance graph can be used to automatically generate new versions of simulations experiments for new model versions. The provenance graph can be automatically generated during the simulation study, e.g., by exploiting workflow systems to conduct the simulation study [10]. However, this requires the modeller to get acquainted with the respective workflow system. Therefore, an automatic and less intrusive solution for automatically capturing provenance is desirable, and shall be the subject of future research.

ACKNOWLEDGMENT

The authors would like to thank the original developers of the tool WebProv, Jacob Smith and Kai Budde.

Author contributions[8]: Conceptualization: O.R., A.M.U.; Methodology: O.R., A.M.U.; Software: O.R.; Investigation: O.R., T.P., J.B., A.M.U.; Data Curation: O.R., T.P, M.H., J.B.; Writing original draft: O.R., T.P., J.B., A.M.U.; Writing review and editing: O.R., T.P., M.H., J.B., A.M.U.; Visualization: O.R.; Supervision: J.B., A.M.U.; Project administration: O.R.; Funding Acquisition: J.B.

REFERENCES

[1] Volker Grimm, Jacqueline Augusiak, Andreas Focks, Béatrice M. Frank, Faten Gabsi, Alice S. A. Johnston, Chun Liu, Benjamin T. Martin, Mattia Meli, Viktoriia Radchuk, Pernille Thorbek, and Steven F. Railsback. Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling, 280:129–139, 2014.

[2] Thomas Monks, Christine SM Currie, Bhakti Stephan Onggo, Stewart Robinson, Martin Kunc, and Simon JE Taylor. Strengthening the reporting of empirical simulation studies: Introducing the stress guidelines. Journal of Simulation, 13(1):55–67, 2019.

[3] Hazhir Rahmandad and John D Sterman. Reporting guidelines for simulation-based research in social sciences. Systems Dynamics Review, 28(4):396–411, 2012.

[4] Frank T Bergmann, Richard Adams, Stuart Moodie, Jonathan Cooper, Mihai Glont, Martin Golebiewski, Michael Hucka, Camille Laibe, Andrew K Miller, David P Nickerson, et al. Combine archive and omex format: one file to share all information to reproduce a modeling project. BMC bioinformatics, 15(1):1–9, 2014.

[5] Pia Wilsdorf, Fiete Haack, and Adelinde M. Uhrmacher. Conceptual models in simulation studies: Making it explicit. In Proceedings of the 2020 Winter Simulation Conference, pages 2353–2360, Piscataway, New Jersey, 2020. IEEE.

[6] Daniel Ayllón, Steven F Railsback, Cara Gallagher, Jacqueline Augusiak, Hans Baveco, Uta Berger, Sandrine Charles, Romina Martin, Andreas Focks, Nika Galic, et al. Keeping modelling notebooks with trace: Good for you and good for environmental research and management support. Environmental Modelling & Software, 136:104932, 2021.

[8] Contributor Roles Taxonomy (CRediT): https://credit.niso.org/

[7] Kai Budde, Jacob Smith, Pia Wilsdorf, Fiete Haack, and Adelinde M. Uhrmacher. Relating simulation studies by provenance—Developing a family of Wnt signaling models. PLOS Computational Biology, 17(8):e1009227, 2021.

[8] Paul Groth and Luc Moreau. PROV-Overview – An Overview of the PROV Family of Documents. Technical Report, World Wide Web Consortium, 2013.

[9] Andreas Ruscheinski, Dragana Gjorgevikj, Marcus Dombrowsky, Kai Budde, and Adelinde M. Uhrmacher. Towards a PROV Ontology for Simulation Models. In Provenance and Annotation of Data and Processes, pages 192–195. Springer International Publishing, 2018.

[10] Andreas Ruscheinski, Pia Wilsdorf, Marcus Dombrowsky, and Adelinde M. Uhrmacher. Capturing and Reporting Provenance Information of Simulation Studies Based on an Artifact-Based Workflow Approach. In Proceedings of the 2019 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, pages 185–196, New York, NY, USA, 2019. Association for Computing Machinery.

[11] Fiete Haack, Kai Budde, and Adelinde M. Uhrmacher. Exploring the mechanistic and temporal regulation of LRP6 endocytosis in canonical WNT signaling. Journal of Cell Science, 133(15):jcs243675, 2020.

[12] Oliver Reinhardt, Andreas Ruscheinski, and Adelinde M. Uhrmacher. ODD+P: Complementing the ODD Protocol With Provenance Information. In M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, editors, Proceedings of the 2018 Winter Simulation Conference, pages 727–738, Piscataway, New Jersey, 2018. Institute of Electrical and Electronics Engineers, Inc.

[13] Jakub Bijak, Daniel Courgeau, Robert Franck, and Eric Silverman. Modelling in demography: From statistics to simulations. In Methodological Investigations in Agent-Based Modelling, pages 167–187. Springer, 2018.

[14] Daniel Courgeau and Robert Franck. Demography, a Fully Formed Science or a Science in the Making? An Outline Programme. Population, 62(1):39–45, 2007.

[15] Thomas K. Burch. Model-Based Demography: Essays on Integrating Data, Technique and Theory. Demographic Research Monographs. Springer Nature, Cham, 2018.

[16] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J. P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, and D. Helbing. Manifesto of computational social science. The European Physical Journal Special Topics, 214(1):325–346, 2012.

[17] Pia Wilsdorf, Anja Wolpers, Jason Hilton, Fiete Haack, and Adelinde M. Uhrmacher. Automatic Reuse, Adaption, and Execution of Simulation Experiments via Provenance Patterns. arXiv:2109.06776 [cs], 2021.

[18] Sarah Nurse and Jakub Bijak. Meta-Information on Data Sources on Syrian Migration into Europe. Technical Report, University of Southampton, 2019. https://www.southampton.ac.uk/baps/inventory/data-sources.page.

[19] Toby Prike, Jakub Bijak, Philip A. Higham, and Jason Hilton. How safe is this trip? Judging personal safety in a pandemic based on information from different sources. Journal of Experimental Psychology: Applied, 2022.

[20] S. Belabbas, J. Bijak, A. Modirrousta-Galian, and S. Nurse. From conflict zones to Europe: Syrian and Afghan refugees' journeys, stories and strategies. 2022. In Preparation.

[21] J. Bijak and M. Czaika. Assessing uncertain migration futures – a typology of the unknown. Quantmig project deliverable d1.1, University of Southampton and Danube University Krems, 2020. Available: https://www.quantmig.eu.

[22] Jakub Bijak. Towards Bayesian Model-Based Demography, volume 17 of Methodos Series. Springer, Cham, 2021.

[23] Martin Hinsch and Jakub Bijak. Rumours lead to self-organized migration routes. In The 2019 Conference on Artificial Life: How Can Artificial Life Help Solve Societal Challenges? (29/07/19 - 02/08/19), 2019.

[24] Claudio Cioffi-Revilla. A Methodology for Complex Social Simulations. Journal of Artificial Societies and Social Simulation, 13(1):7, 2010.

[25] Michel Poulain, Nicolas Perrin, and Ann Singleton, editors. THESIM: Towards Harmonised European Statistics on International Migration. Presses Universitaires de Louvain, Louvain-la-Neuve, 2006.

[26] Toby Prike, Philip A. Higham, and Jakub Bijak. The Boundaries of Cognition and Decision Making. In Jakub Bijak, editor, Towards Bayesian Model-Based Demography: Agency, Complexity and Uncertainty in Migration Studies, Methodos Series, pages 93–112. Springer International Publishing, Cham, 2021.

[27] Ahmet Erdemir, Trent M Guess, Jason Halloran, Srinivas C Tadepalli, and Tina M Morrison. Considerations for reporting finite element analysis studies in biomechanics. Journal of biomechanics, 45(4):625–633, 2012.

[28] Volker Grimm, Gary Polhill, and Julia Touza. Documenting social simulation models: the odd protocol as a standard. In Simulating social complexity, pages 349–365. Springer, 2017.
[29] Volker Grimm, Steven F Railsback, Christian E Vincenot, Uta Berger, Cara Gallagher, Donald L DeAngelis, Bruce Edmonds, Jiaqi Ge, Jarl Giske, Juergen Groeneveld, et al. The odd protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. Journal of Artificial Societies and Social Simulation, 23(2), 2020.
[30] Dagmar Waltemath, Richard Adams, Daniel A Beard, Frank T Bergmann, Upinder S Bhalla, Randall Britten, Vijayalakshmi Chelliah, Michael T Cooling, Jonathan Cooper, Edmund J Crampin, et al. Minimum information about a simulation experiment (miase). PLoS computational biology, 7(4):e1001122, 2011.
[31] Osman Balci. A life cycle for modeling and simulation. Simulation, 88(7):870–883, 2012.
[32] Stewart Robinson. Simulation: The Practice of Model Development and Use. MacMillan, 2nd edition, 2014.
[33] Sebastian Achter, Melania Borit, Edmund Chattoe-Brown, and Peer-Olaf Siebers. RAT-RS: A reporting standard for improving the documentation of data use in agent-based modelling. International Journal of Social Research Methodology, 0(0):1–24, 2022.
[34] Ahmed Laatabi, Nicolas Marilleau, Tri Nguyen-Huu, Hassan Hbid, and Mohamed Ait Babram. ODD+2D: An ODD Based Protocol for Mapping Data to Empirical ABMs. Journal of Artificial Societies and Social Simulation, 21(2):9, 2018.
[35] Open Science Collaboration. Estimating the reproducibility of psychological science. Science, 349(6251):aac4716, 2015.
[36] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science, 22(11):1359–1366, 2011.
[37] Garret Christensen, Zenan Wang, Elizabeth L Paluck, Nicholas Swanson, David J Birke, Edward Miguel, and Rebecca Littman. Open Science Practices are on the Rise: The State of Social Science (3S) Survey, October 2019.
[38] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. A manifesto for reproducible science. Nature Human Behaviour, 1(1):0021, January 2017.
[39] B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. Science, 348(6242):1422–1425, 2015.
[40] Marcin Miłkowski, Witold M. Hensel, and Mateusz Hohol. Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. Journal of Computational Neuroscience, 45(3):163–172, 2018.
[41] Victoria Stodden, Peixuan Guo, and Zhaokun Ma. Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. PLOS ONE, 8(6):1–8, 2013. Publisher: Public Library of Science.
[42] Oliver Reinhardt, Adelinde M. Uhrmacher, Martin Hinsch, and Jakub Bijak. Developing Agent-Based Migration Models in Pairs. In Proceedings of the 2019 Winter Simulation Conference, pages 2713–2724, Piscataway, New Jersey, 2019. IEEE.
[43] Tom Warnke and Adelinde M. Uhrmacher. Complex simulation experiments made easy. In Proceedings of the 2018 Winter Simulation Conference, pages 410–424, Piscataway, New Jersey, 2018. IEEE.
[44] Jan Salecker, Marco Sciaini, Katrin M Meyer, and Kerstin Wiegand. The nlrx r package: A next-generation framework for reproducible netlogo model analyses. Methods in Ecology and Evolution, 10(11):1854–1863, 2019.
[45] Pia Wilsdorf, Jakob Heller, Kai Budde, Julius Zimmermann, Tom Warnke, Christian Haubelt, Dirk Timmermann, Ursula van Rienen, and Adelinde M. Uhrmacher. A Model-Driven Approach for Conducting Simulation Experiments. TechRxiv, 2021.
[46] Pablo Briñol and Richard E. Petty. Source factors in persuasion: A self-validation approach. European Review of Social Psychology, 20(1):49–96, 2009.
[47] Bonnie C. Wintle, Hannah Fraser, Ben C. Wills, Ann E. Nicholson, and Fiona Fidler. Verbal probabilities: Very likely to be somewhat more confusing than numbers. PLOS ONE, 14(4):e0213522, 2019.
[48] Mirko Gabelica, Ružica Bojčić, and Livia Puljak. Many researchers were not compliant with their published data sharing statement: mixed-methods study. Journal of Clinical Epidemiology, 2022.
[49] D. Vogel and V. Kovacheva. Classification report: Quality assessment of estimates on stocks of irregular migrants. Technical report, HWWI Hamburg Institute of International Economics, 2008.

OLIVER REINHARDT is a Ph.D. student in the Modeling and Simulation Group at the University of Rostock. He holds an MSc in Computer Science from the University of Rostock. In his research, he is concerned with domain-specific modeling languages and the methodology of agent-based simulation. His email address is oliver.reinhardt@uni-rostock.de.

TOBY PRIKE is a Research Associate at the University of Western Australia where he works on the ARC funded project: Combating Misinformation – Designing a Toolkit to Address a Global Problem. His broader research interests and experience include non-evidence based beliefs, reasoning, cognitive bias, memory, and migration decision-making. His email address is toby.prike@uwa.edu.au.

MARTIN HINSCH is a Research Associate at the MRC/CSO Social and Public Health Sciences Unit at the University of Glasgow. He is interested in emergent structures and complex systems and has done research in theoretical biology, bioinformatics, machine learning, epidemiology and swarm robotics. His email address is martin.hinsch@glasgow.ac.uk.

JAKUB BIJAK is Professor of Social Demography and currently a Joint Head of Department of Social Statistics and Demography at the University of Southampton, UK. At present, he leads an ERC funded project on Bayesian Agent-Based Population Studies and a H2020 project QuantMig: Quantifying Migration Scenarios for Better Policy. His email address is j.bijak@soton.ac.uk.

ADELINDE M. UHRMACHER is Professor at the Institute for Visual and Analytic Computing of the University of Rostock and head of the Modeling and Simulation Group. She holds a PhD in Computer Science from the University of Koblenz and a Habilitation in Computer Science from the University of Ulm. Her email address is adelinde.uhrmacher@uni-rostock.de.