Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

CRISIS, WHAT CRISIS – DOES REPRODUCIBILITY IN MODELING & SIMULATION REALLY MATTER?

Simon J. E. Taylor

Modelling & Simulation Group
Department of Computer Science
Brunel University London
Uxbridge, UB8 3PH, UK

Tillal Eldabi

Brunel Business School
Brunel University London
Uxbridge
UB8 3PH, UK

Thomas Monks

NIHR CLAHRC Wessex
University of Southampton
Southampton General Hospital
Southampton, SO16 6YD, UK

Markus Rabe

Department IT in Production and Logistics
TU Dortmund University
Leonhard-Euler-Str. 5
Dortmund, 44227, GERMANY

Adelinde M. Uhrmacher

Institute of Computer Science
University of Rostock
Albert-Einstein-Straße 22
18059 Rostock, GERMANY

ABSTRACT

How important it is to our discipline that we can reproduce the results of Modeling & Simulation (M&S) research? How important is it to be able to (re)use the models, data, and methods described in simulation publications to reproduce published results? Is it really that important or are the lessons and experiences described in a paper enough for us to build on the work of others? At the 2016 Winter Simulation Conference, a panel considered opinions on reproducibility in discrete-event simulation. This article builds on these and asks if there really is a reproducibility crisis in M&S? A diverse range of views on the subject are presented including reflections on the reproducibility in terms of the art and science of simulation, the frustrations of poor reproducibility, perspectives from the industrial production & logistics community, the wider context of open science and artefact sharing, and the role of provenance beyond reproducibility.

1 INTRODUCTION

What does reproducibility mean in terms of Modeling & Simulation? Does reproducibility matter? To what degree is reproducibility possible? Could reproducibility policies have a negative impact on publishing in M&S? These questions were amongst some initially posed to our panelists as a starting point to gather a range of views on this potentially important topic. This panel paper presents a range of views on the subject and contributes to the on-going debate about the role of reproducibility in our discipline.

2 WHO NEEDS REPRODUCIBILITY (TILLAL ELDABI)

2.1 Reproducibility of the Art and Science of Simulation

As far as simulation modeling is concerned, reproducibility can be defined as the reuse of an existing model with or without its data, for the purpose of wholly or partially modeling a new system or even the same system for which the original model is produced. In the 2016 reproducibility panel by Uhrmacher et al. (2016), the first sentence was "Scientific research should be reproducible …". Simulation is a science as well as it is an art. It is, therefore, safe to assume that the *science* part of simulation is reproducible, yet the *art* part is *not*. subsequently, the answer to the question of reproducibly is "yes" if we are only looking at the science of simulation. The art of simulation, on the other hand, is not reproducible, and it is not even relevant to the question of reproducibility. In this part of the panel, we will extract the science part of simulation that can be reproducible and the art side that is not. Shannon (1998) provides a simplistic yet comprehensive delineation between the art and science of simulation that is still relevant today. He suggests that steps such as programming and statistical analysis are part of the *science* of simulation, whilst analysis and modeling artefacts are the *art*. After two decades such views can be slightly modified by adding that the art part is the conceptual modeling where real system parts are collected and subsequently modified to be part of the simulation model. On the other hand, the science parts are the development of a computer model, whether based on programming from scratch or using a dedicated simulation package. The science part also includes data collection methods, analysis and, naturally, replications.

2.2 Why is the Art of Simulation not Reproducible?

A problem formulation is about trying to convert users' questions and their understandings of their current system into specific questions and components that can be transformed into modelled artefacts. The process of capturing the requirements and drawing up conceptual models usually follows semi-standardized practices (Robinson 2008). This process lies at the heart of the art of simulation and it is highly dependent on the tacit skills and expertise of the modelers. There are some basic concepts to follow but generally each model is an outcome of the skills, problem domain, and objectives of the model. No matter how similar the systems, it is almost impossible to come up with the same model if any of these pillars is different. Therefore, conceptual models and models where the skills and the behavior of people are concerned cannot be reproducible.

2.3 Is the Science of Simulation Reproducible?

The discussion above does not mean that we should ignore the need for reproducibility. For example, modules of the computer models could be reproducible. In fact, using the same simulation package has some elements of intrinsic reproducibility where, for example, an arrival or machine component can be reused or reproduced. Moreover, this is valid for simulation models of natural systems, where humans and social activities are almost non-existent within the model or the system, for example, a weather system or a fully automated system. In such cases models could be reproducible and can be reused if the required model is *understood* to be the same.

2.4 Challenges to Reproducing Scientific Models

In cases where reproducibility is possible, it is required to reduce duplication of the model building process and, hence, the overall modeling cost in addition to the cost of collecting data from scratch. Reproducibility – where possible – is useful to compare new arising phenomena with other ones that have already been modeled. However, there are some challenges to that. First, the simulation tool used needs to be the same. If a commercial package is used for building the original model, then the same package will be required to run the reproduced model. Not only the same package, in fact the version used has to be compatible. If this is not the case, the attention will divert to the age-old discourse of interoperability. Second, there are legal issues. This has been discussed in a number of platforms before. To reproduce the data, there needs to be a legal allowance for that. Third, and most importantly, even if the system is the same, it is not possible to reproduce the old model unless the newly modelled problem is the same as the old one and with the same questions

2.5 Reproducibility Matrix: Possible Forward

In Table 1, we present a 2x2 matrix as an attempt to provide a quick guidance to the possibilities of reproducibility of simulation models without delving into the technicalities of how to do that. The four elements of the matrix are detailed below and will be based on the art and science of simulation arguments above:

- The art of simulation: tacit knowledge and expertise
- The science of simulation: Data collection; analysis; computer modules; measurements etc.
- Social systems: non-replicable behavioral; business models; people-oriented etc.
- Natural systems: replicable physical sciences; chemical process; automated systems

Table 1: The M&S reproducibility matrix.

Natural	**Partially reproducible**: Non-standardized modeling and standardized replicable systems.	**Reproducible**: standardized modeling and standardized replicable systems.
Social	**Not reproducible**: Non-standardized modeling non-standardized non-replicable systems.	**Not reproducible**: Standardized modeling and non-standardized problems non-replicable systems.
	Art	Science

2.6 Grab and Glue Simulation: Dream Forward

It is time to take the brakes of the Grab-and-Glue concept. The aspiring yet daring idea of the Grab-and-Glue concept was discussed by Eldabi et al. (2004). The core of the idea was that simulation models can be built by grabbing different modules from any source (of existing models) and glue them together to build a new model, doing this repeatedly until an acceptable model is reached. At the time of the paper, this looked like a dream. After almost 15 years, interest seems to be rising to reconsider it. The recent years have seen a significant increased attention on hybrid simulation (Eldabi et al. 2016) and on reproducibility (Uhrmacher et al. 2016). More research in these two areas will ultimately give rise to the concept of Grab-and-Glue simulation. My prediction is that this will take place without the need for standards, documentations, or even protocols, as all of these will be replaced by Simulation 2.0, i.e., user-generated modules.

3 THE PROBLEM DOES NOT MATTER TO EVERYONE (TOM MONKS)

3.1 Is This a Familiar Story?

Imagine you are a young researcher in computer simulation. During a particularly challenging simulation research project, a literature search identifies a simulation study with results that would greatly enhance your own modeling. You believe that you can directly reuse and build on results of the study and gain insight into how others have tackled a particularly difficult applied modeling problem. However, on a detailed review of the work you experience that it is not visible how the authors have actually built their model. You have fun tracing citations the authors have made to document vague optimization procedures that they have included in the model – only to find a recursion of papers that never describe these vital details. Your first thought is that this is not a problem, because you can always contact the author of the model. Unfortunately, it turns out that the lead author was a PhD student who did not pursue a career in research and has since disappeared into a lucrative career in finance. The student's supervisor and co-authors might be able to help, of course. However, they have long since retired; or cannot answer the questions because they are not documented in the thesis; or may genuinely have had no idea what the PhD student was doing (only in rare cases, of course). The outcome of all this effort is that it is not possible or at best extremely difficult to reproduce the simulation model and its results. You begin from scratch again.

You may or may not identify with aspects of my story. Either way I recommend reviewing some of the research about the reproducibility crisis in simulation and related areas (Kurkowski et al. 2005; Rahmandad and Sterman 2012; Boylan et al. 2015; Janssen 2017). I think it is also legitimate to ask yourself: Does anything of this really matter or am I being overdramatic? After all, you might argue that our hypothetical researchers simply need to work harder and resolve the issues themselves. The answer depends on your personal view of science. One view is that credible science that is of potential value to society is *transparent and open*. Thus, it enables results to be verified and others to build on this work either in research or industry. This is important for both theoretical and applied results. There is also an ethical argument for reproducibility, as much research is publicly funded. If results and knowledge do not have the potential to offer benefits for science or society, is this an appropriate use of public money?

3.2 Simple Actions Can Help

Communication of computer simulation design and implementation is often difficult due to the complexity of models. A big part of solving the reproducibility crisis is, therefore, publication of the model code. There are relatively simple modern ways to share code (Taylor et al. 2017), but in research, this is rarely done, and there is little incentive to share models within industry. Even with code, there are circumstances where results cannot be verified. Within modern applications of computer simulation, study results should be reproducible based on experimental lab conditions, i.e., the model, software, code libraries, and the computer system specification need to be precisely reported. This is where support for the reporting of simulation models comes in.

Many disciplines outside of computer simulation have adopted a reporting guidelines approach. These are often simple checklists that can be used both during the study to standardize internal documentation and during writing up for an academic journal. Some disciplines such as medicine enforce their usage while others only recommend that authors make use of them. There are already several guidelines available within optimization (Kendall et al. 2016) and computer simulation (Gass 1984; Grimm et al. 2006; Grimm et al. 2010; Rahmandad and Sterman 2012; Monks et al. 2018). The guidelines that I was involved in developing (Strengthening the reporting of empirical simulation studies: STRESS) provide a standardized approach across Agent-based Simulation, Discrete-event Simulation and System Dynamics (Monks et al. 2018). If you are writing up a simulation study, I believe they are worth a read, if you have a spare five minutes (freely available from https://doi.org/10.1080/17477778.2018.1442155).

3.3 Improvements and Limits

There are signs of a shift in the status quo about the openness of simulation research. System Dynamics Review, for example, now requires authors to submit their full model as supplementary material and ACM TOMACS has voluntary processes in place to verify computational results. However, as a community, we need to recognize that reproducibility is extremely challenging and that steps to tackle it are not going to be acceptable for everyone. It is, therefore, not going to be solved quickly or perhaps at all. It has been interesting to talk to colleagues from both academia and industry about reproducibility and the steps that I have outlined above. Responses vary, but here I will stereotype them to make a simple point. My academic colleagues ask me, '*publication is already a grueling process, why would we increase our workload*?'. While my industry colleagues tell me that '*reproducibility and documentation are essential for my business to survive, how do you guys get away with it*?' The difference is due to incentives and consequences. There is no direct reward or consequence for academics to undertake what they view as additional work to make their study more open. For industry,however, there is potentially a high price if, for example, modelers leaving the team without thorough documentation of the modeling they had conducted for a client. An agenda for openness should, therefore, identify incentives for the current generation of researchers to actively adopt more open practices (beyond altruism). As we learn how to incentivize ourselves, my hope is that we will cultivate a new generation of simulation modelers where openness is second nature. The latter aim may be more achievable than the former.

I hope that the collective views and arguments put forward in this paper encourage those entering careers as simulation modelers to consider the openness of their science. However, I admit that you don't need to follow a strict open science path to have a successful career in simulation or research in general. There are plenty of successful professors and I am confident that reproducing some aspects of their work would prove challenging.

4 IS REPRODUCABILITY A RESEARCH KILLER? A SLIGHTLY PROVOCATIVE VIEW FROM A PRODUCTION AND LOGISTICS PERSPECTIVE (MARKUS RABE)

It is common sense that the goal of science is to contribute to the growth of the world's knowledge – this goal is, obviously, independent of the specific discipline of research. Thus, is seems to be common understanding that published research results that do not include sufficient information how these results have been achieved can hardly be considered science, and also the 'critical' research areas, such as military, logistics, or medicine need to take this rule into account. Nevertheless, there are a few specifics that need to be carefully considered in this debate, which shall be discussed in the following on the sample of the production and logistics domain.

4.1 IPR Issues

First, it is evidently necessary to differentiate reproducibility of results from the *publication* of information that is necessary to reproduce results. The demand of *generating reproducible results* should be a general goal, independent of the research area. In the direct course, this is a challenge to the researcher himself: He or she should always be in the position to reproduce his or her own results, which requires a careful documentation of the model, the utilized tools (including its release numbers etc.), the data, and all other relevant information. In an industrial context, which is common in this research area, principal reproducibility is of high importance – be it just for re-using the models or the data later in a slightly different context with low effort, or, in a very negative case, when study results need to be defended at a law court to prove that the conductor of a professional simulation study has applied the current state of the art to achieve and validate his or her results. However, the same industrial context has a clear opinion on publication, which can often be summarized with the single word "NO". Non-disclosure agreements are often binding partners to keep all input and output data classified for five years or even longer. Publication of (even anonymized) data or results requires a weird release process through several departments and hierarchy levels in the enterprise, not rarely with the result that publication is finally not permitted – even if the data asked for to be published are aggregated and show merely nothing about any specific know-how or financial characteristics of this enterprise.

In production and logistics, it is a part of the basic self-understanding of the discipline that research results should be applicable to real cases. Even in basic research (as, for example, funded by the German national funding organization DFG), frequently a proof of the developed methods and techniques is to be performed on real company data or at least on data that are derived from such real data. Therefore, a strict limit to publish only fully reproducible information would factually lead to prohibiting publication of all these results, which is in diametric contrast to the goals of science in our common understanding. Besides principle (and sometimes irrational) IPR issues, the points that make the full publication difficult – and in many cases impossible – can roughly be differentiated into data issues and model issues.

4.2 Data Issues

For data, the protection of company knowledge is quite evident. Typical data, such as manufacturing times, sales predictions, new products, and many others may be fully reasonable to protect. But, there is much more challenge than just publishing data: The full path of acquiring these data would need to be documented. Otherwise, we come to the risk that results are reproducibly invalid, being of quite limited benefit. Why is this an issue? In many application environments (supply chains are an especially weird example), data originate from a number of different companies with different transactional IT systems, different data definitions, and quite different procedures to achieve (at least some) data quality. Data need to be unified and cleansed, the latter being a difficult process (in practice taking many weeks of work) to eliminate completely wrong records, amend missing attributes or rectify obviously implausible ones, passing through cross-entity plausibility checks (e.g., showing trucks having been loaded by 240% or retail shops located in the middle of the English Channel), and trying to negotiate solutions with the involved companies. All these steps, the original data, the final prepared data used for the simulation runs, and, in principle, all intermediate results (as they might contain manual rectifications) would need to be published. Given the unlikely situation that non-disclosure would not be an issue, it might be questionable if the effort for bringing all this information in readable formats and documentations is worthwhile given a not too high probability that anybody wants to build on exactly this information. However, technically it should be possible, and keeping an archive of these data together with at least a limited documentation how they have been processed should anyway be mandatory, as explained above.

4.3 Model Issues

While non-disclosure requirements are obvious for company data, for the model this might be a question. In supply chains, the structure of a network with retail shops and distribution centers seems to be quite obvious, and interested competitors should be able to collect such information anyway. Nevertheless, the details of the model (and especially of its control elements) typically contain important knowledge and facts that the company definitely does *not* wish to be public. Examples could be sophisticated control strategies in production plants or just prioritization rules (e.g., under which conditions are confirmed orders postponed for the benefit of other “strong” or “valuable” customers?).

On the other hand, there are major technical issues. Due to the complexity of the models in this domain, hardly any model is built from scratch in some kind of simulation language, but built upon commercial-of-the-shelf tools, which are more or less specific to the respective sub-domain. There can even be several shells that finally lead to the model. As a supply chain example, the tool SimChain is a data-driven simulation approach, thus, the advantage is that there is no separate and specific model – you will just need the (prepared) data and the tool. However, such tools are not stable, but evolve and improve over time. For reproducibility, this will require to achieve or store (with the issue of license obligations) the specific version of the tool that had been used for the experiments. Even worse, SimChain itself builds on another tool, namely the tool Plant Simulation provided by Siemens, and each SimChain version will rely on a specific version of this underlying tool. If we look on a slightly longer period, this underlying tool will not run on any version of the Microsoft Windows operating systems, so we would also need to archive this system (and the respective license). Whether this license can then later be activated (say, in 8 years) and whether we have hardware to run the system upon are open questions, unfortunately, with the realistic expectation that finally, in spite of all documentation effort, reproduction might be infeasible.

4.4 What is Happening?

A previous study, conducted by Markus Rabe in preparation of a WSC panel in 2016 (Uhrmacher et al. 2016) has shown that, from a sample size of 193 papers, only ten gave (mostly still incomplete) details about the model construction, while the others gave vague or even literally no information on how the model had been constructed. Only five of these papers gave complete information on the used data – toy applications that, on the other hand, might be questionable under the formulated demand that applicability for real cases should be proven. About a quarter of the papers has described the data at least partially, with the intention to illustrate the relationship of the input data with the simulation results. This investigation does, however, not take into account that the authors of these publications have probably not been required or just encouraged to publish their data, and the publication schemas under study did all not provide any standard mechanism to do this; therefore, the results might be biased by the publication environments.

4.5 Is There a Solution, and Are We Asking for the Right Direction?

It should be clearly said that this is no argumentation not to carefully describe how results have been achieved, but the contrary. Especially when the concrete data cannot be published, it should be the responsibility of the authors to carefully document where data came from, which deficits have already been in these data, which assumptions have been taken and which types of (automatic or manual) changes have been conducted in the specific attributes. Having this information is inevitable to provide credibility for the study and its results. The same applied to models: Information how the model is constructed, which tools or libraries have been used, and which specific building blocks have been created or re-used, should be documented as careful as possible. Obviously, this can only partially be done within conference or journal papers, due to their typically very limited number of pages. Of course, authors could provide such information on their homepages or other media, but experience shows that such sources frequently have a much shorter life time than journals or long-lasting conferences such as the WSC. Therefore, journals and conferences should strongly consider to provide a means to authors to publish additional material together with their paper (which does not imply that this becomes part of the reviewing process).

Going one step further, before targeting a (reasonably not achievable) reproducibility, we should honestly ask for *credibility*. What makes us believe that the authors' results are valid? The above-mentioned study (Uhrmacher et al. 2016) has shown that from the 193 papers under study, two thirds did not mention any validation procedures at all, and only 9 (5%) gave at least some information on how the authors believed to have achieved credibility. There has been research on how to achieve credibility, and how to establish verification and validation processes along a simulation study (see, for example, Rabe et al. 2009), but they seem to be almost ignored in most of these papers.

4.6 Consequences for (not only) the Production and Logistics Field

In general, the scientific goal of reproducible research is, of course, valid for all domains, not excluding production and logistics. However, the domain-specific target of demonstrating the research being applicable in the companies' practice puts specific IPR roles on the publication. Thus, this target needs to be formulated as "all information for reproduction of the results should be given, as long as they are not under non-disclosure regulations". Furthermore, there should be a reasonable estimate of the advantage of public documentation (including both its value when used and its probability for being used) in relation to the effort providing it, avoiding to put formal and useless work on researchers going to publish their results. There might be a process in the scientific community to come to widely acceptable guidelines for such estimates.

A general demand for reproducible results, only, would have a very negative effect: Publications would go down to artificial studies and toy samples – a kind of publications that the author of these lines, when active as a reviewer, would frequently like to reject because of their extremely limited impact. Such approaches may be absolutely valid for specific questions, but they should definitely not be encouraged in general.

Before asking for all model details and data, conferences and journals should intensely ask for detailed information on the verification and validation (V&V): How has credibility of the study been achieved, from the first task description through the modeling and data preparation processes to the final interpretation of results? Giving additional data or (partial) models as accompanying material might be helpful and reasonable, but it does not replace the V&V story, which should be part of the original paper. This goes in line with the demand from provenance that not only the model should be given, but also the way it has been produced.

Therefore, at least in the production and logistics domain, the demand to be discussed should incorporate three steps: (1) Make your V&V procedure transparent and demonstrate why your studies are credible, (2) Offer means to the authors to publish any related data they like together with the published paper, and (3 and lowest priority) discuss about actual demands for reproducibility and their limitations.

5 REPRODUCIBILITY AND OPEN SCIENCE (SIMON TAYLOR)

Reproducibility is a major concern, as the results of many papers, not just in M&S, cannot be reproduced or validated (Baker 2016). In some ways, reproducibility best practices could be synonymous with "openness", i.e., an open approach to the way we do science. "Open Science" encapsulate practices that aim to make scientific research accessible to all, typically in some digital format. For most, this involves open access publishing where a scientific paper is made accessible through a journal or some institutional open access repository. However, Open Science principles go further and aim to make all the artefacts of scientific research openly accessible.

The FOSTER project (Facilitate Open Science Training for European Research) (www.fosteropenscience.eu) defines Open Science as "… the practice of science in such a way that others can collaborate and contribute, where research data, lab notes, and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods." Open Science, therefore, refers to efforts to make the output of research more widely accessible to scientific communities, business sectors, and society in general (OECD 2015). The area consists of strategies that address a wide range of associated topics: open access, open research data, open research protocols and notebooks, open access to research materials, open source software, citizen science, open peer review, and open collaboration. It is often facilitated by digital technology and open access repositories. Sometimes, technologies such as Science Gateways and e-Infrastructures (cyberinfrastructures) are used. Funding agencies across the world are reflecting the need to be more open with respect to the outcomes of publically-funded research programs.

There are many benefits that arise from openness in science and research (OECD 2015, p.18):

- Improving efficiency in science by reducing duplication and the costs of creating, transferring, and reusing data; allowing for more research from the same data; and multiplying opportunities for domestic and global participation in the research process.
- Increasing transparency and quality in the research validation process by allowing for a greater extent of replication and validation of scientific results.
- Speeding the transfer of knowledge from research to innovation.
- Increasing knowledge spillovers to the economy and increasing awareness and conscious choices among consumers.
- Addressing global challenges more effectively by globally coordinated international actions.
- Promoting citizens' engagement in science and research – Open Science and open data initiatives may promote awareness and trust in science among citizens. In some cases, greater citizen engagement may lead to active participation in scientific experiments and data collection.

Taylor et al. (2017) discuss Open Science for M&S in more depth. However, going beyond reproducibility, adopting Open Science in our discipline might help to increase transparency and collaboration as well as making the transfer of knowledge from research to innovation and the increased impact of research on society more effective. However, there is a balance to be made when intellectual property rights or confidentiality are an issue. Overall, to support reproducibility and Open Science in M&S one might:

- Publish openly using Gold or Green open access.
- Adopt good Open Data and Reusability practices that encourage independent verification or standardized reporting checklists such as STRESS.
- Consider making your data, results, software, etc. openly accessible (and trackable) by submitting your works to Open Access Repositories that support the use of Digital Object Identifiers (DOIs).
- Use Creative Commons licenses to specify how your work should be shared and used.
- Use a Researcher Registry such as ORCID to uniquely identify yourself and link this to your works via DOIs.
- Ensure that you use both DOIs and ORCIDs when publishing or in social media to correctly identify yourself and your works so that these can be tracked through scientometrics and altmetrics.
- Consider deploying your simulations via a Science Gateway or similar portal-based approach to enable the widest possible access to your work.

6 PROVENANCE IN MODELING AND SIMULATION – BEYOND REPRODUCIBILITY (ADELINDE UHRMACHER)

At the WSC 2016 panel on reproducibility (Uhrmacher et al. 2016), the state of the art and prospect of reproducibility in different application fields, i.e., network simulation, modeling and simulation in health care, logistics, and military has been discussed. Despite standard simulation tools and libraries, even in the area of network simulation, most of the work published in M&S cannot be reproduced. The situation is not improved if simulation studies are littered with classified information, as is often the case (e.g., in military, logistics, and healthcare). This was the dismal conclusion being in stark contrast to the common understanding of the vital importance of reproducibility for M&S research. The assessment was shared by the audience: a small survey was executed at the beginning of the panel session. The importance of reproducibility was equivocally agreed upon, three quarters of the audience assessed that only 25% or less of published M&S research would be reproducible and that conferences should do more to support reproducible research. However, less than 25% of the audience have ever tried to replicate simulation results themselves. Given the interest in reproducible research (50% of the audience stated their own work to be replicable), this is a wake-up call, or, as it was put into one questionnaire: “The field is in danger of becoming filled with fragmentary, unreliable pseudo-results”. Consequently, identifying concrete means of improving reproducibility has been and is a major issue, which implies changes in the scientific culture as well as the need for additional methodological support.

6.1 Initiatives of Research Outlets in the Last Two Years

To change the scientific culture in a scientific area, such as M&S, publication research outlets need to establish corresponding reviewing procedures. At the time the 2016 panel paper was written, the Replicated Computation Results (RCR) initiative of the ACM Transactions of Modeling and Computer Simulation was merely 5 months old, and the first paper that successfully took part in this separate reviewing process to replicate computational results and to evaluate the accompanying artifacts had just been published (Feng et al. 2016). Since then, around 20% of accepted papers in TOMACS took part in this initiative and have received ACM badges (see “ACM Artifact Reviewing and Badging”). In 2018, also the ACM SIGSIM PADS conference initiated a formal process for replicating computational results and evaluating the artifacts, thereby adapting and streamlining the submission and reviewing process of the TOMACS RCR initiative. 50% of the accepted papers have been labeled to document that computational results could successfully be replicated, or that the artifacts associated with the published paper have been found to be functional or reusable. With this, two major research outlets of discrete event modeling and simulation research have joined the ACM initiative of “Artifact Reviewing and Badging” improving reproducible results. Authors are increasingly putting effort into carefully designing their artifacts and documenting their research. They do not merely document what are the results, but particularly how results have been generated. Provenance provides “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” (Groth and Moreau 2013). Provenance information is about how a product came into being and as such essential for reproducibility (Brinckman et al. 2018).

6.2 Provenance Information About Simulation Data – Workflows, Scripting, and Domain-specific Languages

To support acquiring and querying the provenance of data, be these data generated in-silico, in-vitro, or in-vivo, has been the subject of major research efforts during the last two decades and, accordingly, a diversity of software tools and platforms are available that provide a rich portfolio of methods to document the generation of data and to replicate or reproduce them. Workflows are widely used to support the documentation of individual simulation experiments (Wolstencroft et al. 2013; Ribault and Wainer 2012; Görlach et al. 2011; Rybacki et al. 2012) and how simulation data have been generated. Specifically in the area of cell biology, standards such as SBML also contain and maintain provenance information about how the simulation data have been created. For this purpose, SED-ML, an external domain-specific language, has been developed (Bergmann et al. 2018). Also, domain-specific languages such as SESSL (Ewald and Uhrmacher 2014) allow for easily replicating simulation experiments (Peng et al. 2016), as do model-based approaches such as the one by Teran-Somohano et al. (2015). The later are part of recent developments to treat simulation experiments as first-class citizen in simulation, and to make the hypotheses they are based upon explicit (Peng et al. 2014; Lorig et al. 2017; Agha and Palmskog 2018). All of them avoid lengthy ambiguous narratives how simulation data and results have been generated by pursuing approaches that are formal, partly graphical, and, most importantly, executable.

6.3 Provenance Information About Simulation Models – Structured Annotations, Roles and Relations

Efforts have been dedicated to making simulation models accessible and to facilitating their reuse. Thereby, suitable annotations play a central role. This applies to standardized guidelines for describing agent-based models, such as the ODD (Overview, Design concepts, and Details) protocol (Grimm et al. 2010), as well as standards in computational biology, such as supported in SBML (Systems Biology Markup Language) (Hucka et al. 2003), whose annotations contain information about the source, the owner, limitations, or purpose of the model. However, current annotations of simulation models focus on the question of “What has been developed?” rather than “How has it been developed?”. To capture the provenance of simulation models, i.e., how a model has been developed, in Ruscheinski and Uhrmacher (2017) data, simulation model, and simulation experiments have been identified as central processes and products within a simulation study and mapped to processes and artifacts within an Open Provenance Model profile (Moreau et al. 2011). In this triangle of interrelated artifacts, exploiting roles in relating artifacts and processes provides additional semantics that can be queried. Data used as inputs, for calibration, or for validation are distinguished. The provenance model structures information about products and processes and their interrelations within and across simulation studies in a compact highly accessible, partly formal manner. The adoption of provenance models if suitably tuned to the requirements of modeling and simulation (Ruscheinski et al. 2018) can, thus, form a valuable complement to current standards such as the ODD protocol (Reinhardt et. al 2018).

6.4 Provenance Beyond Reproducibility

Reproducibility requires additional documentation effort and care referring to the artifacts to ensure their functionality and reusability by third parties. Therefore, to have a lasting positive effect on modeling and simulation, additional efforts are required a) to minimize these efforts and b) to make them worthwhile. Means for acquiring crucial information automatically and approaches that allow for a compact and structured specification will help to reduce the efforts. To make the efforts worthwhile, incentives, such as provided by ACM, and reviewing standards, such as established in TOMACS or the ACM SIGSIM PADS conference, will play a role. In addition, the benefits of the documentation for others and oneself need to be clear. Approaches with a focus on how something has been generated, which entail an unambiguous and executable semantics, allow not only for a convenient replication of modeling and simulation results, but also facilitate the generation of new results. For example, the specification of simulation experiments in SESSL does not only allow for replicating simulation data, but supports – as part of the provenance of simulation models – the automatic generation of new simulation experiments for newly extended or composed models to test specific behavior patterns (Peng et al. 2016). Thus, provenance, information about the past does not only allow for understanding the present, but also for designing the future, in opening up new avenues for generating and analyzing simulation models. These new opportunities will form an additional strong incentive for provenance and, thus, improving reproducibility in modeling and simulation research.

7 CONCLUSIONS

This panel paper has considered whether or not there is a reproducibility crisis in M&S. Views have been presented that have reflected on reproducibility in terms of the art and science of simulation, the frustration of poor reproducibility, perspectives from the industrial production and logistics community, the wider context of open science and artefact sharing, and the role of provenance beyond reproducibility. It appears that reproducibility is indeed important to our discipline, but a practical balanced approach needs to be developed that reflects the needs of both privacy and industry.

REFERENCES

Agha, G. and K. Palmskog. 2018. “A Survey of Statistical Model Checking”. *ACM Transactions on Modeling and Computer Simulation* 28(1), published online doi:10.1145/3158668.

Baker, M. 2016. “1,500 Scientists Lift the Lid on Reproducibility”. *Nature* 533(7604):452–454.

Bergmann, F., J. Cooper, M. König, I. Moraru, D. Nickerson, N. Le Novere, B. Olivier, S. Sahle, L. Smith, and D. Waltemath. 2018. “Simulation Experiment Description Markup Language (SED-ML) Level 1 Version 3 (L1V3)”. *Journal of Integrative Bioinformatics* 15(1), published online doi:10.1515/jib-2017-0086)..

Boylan, J. E., P. Goodwin, M. Mohammadipour, and A. A. Syntetos. 2015. "Reproducibility in Forecasting Research". *International Journal of Forecasting* 31(1):79-90.

Brinckman, A., K. Chard, N. Gaffney, M. Hategan, M. B. Jones, K. Kowalik, S. Kulasekaran, B. Ludäscher, B. D. Mecum, J. Nabrzyski, V. Stodden, I. J. Taylor, M. J. Turk, and K. Turner. 2018. “Computing Environments for Reproducibility: Capturing the Whole Tale”. *Future Generation Computer Systems*, published online doi:10.1016/j.future.2017.12.029.

Eldabi, T., M. Balaban, S. Brailsford, N. Mustafee, R. E. Nance, B. S. Onggo, and R. G. Sargent. 2016. “Hybrid Simulation: Historical Lessons, Present Challenges and Futures”. In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. Roeder et al., 1388-1403. Piscataway, New Jersey: IEEE.

Eldabi, T., M. W. Lee, and R. J. Paul. 2004 "Examining the Feasibility of Constructing Simulation Models Using the Web-based 'Grab-and-Glue' Framework." In *Proceedings of the Winter Simulation Conference*, edited by R. G. Ingalls et al., pp. 1494-1501. Piscataway, New Jersey: IEEE.

Ewald, R. and A. M. Uhrmacher. 2014. “SESSL: A Domain-Specific Language for Simulation Experiments”. *ACM Transactions on Modeling and Computer Simulation* 24(2):11:1-11:25. .

Feng, C., J. Hillston, and V. Galpin. 2016. “Automatic Moment-Closure Approximation of Spatially Distributed Collective Adaptive Systems”. *ACM Transactions on Modeling and Computer Simulation* 26(4):1–26.

Gass, S. I. 1984. "Documenting a Computer-based Model". *Interfaces* 14(3):84-93.

Görlach, K., M. Sonntag, D. Karastoyanova, F. Leymann, and M. Reiter. 2011. “Conventional Workflow Technology for Scientific Simulation”. In *Guide to e-Science*, edited by X. Yang et al., 323–352. London: Springer.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe’er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis. 2006. "A Standard Protocol for Describing Individual-based and Agent-based Models". *Ecological Modelling* 198(1–2):115-126.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. 2010. "The ODD Protocol: A Review and First Update". *Ecological Modelling* 221(23):2760-2768.

Groth, P. and L. Moreau. 2013. “PROV-Overview. An Overview of the PROV Family of Documents”. https://www.w3.org/TR/prov-overview/, accessed June 25th, 2018.

Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang and the rest of the SBML Forum. 2003. “The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models”. *Bioinformatics* 19(4):524–531.

Janssen, M. A. 2017. “The Practice of Archiving Model Code of Agent-based Models”. *Journal of Artificial Societies and Social Simulation* 20(1):2.

Kendall, G., R. Bai, J. Błazewicz, P. De Causmaecker, M. Gendreau, R. John, J. Li, B. McCollum, E. Pesch, R. Qu, N. Sabar, V. G. Berghe, and A. Yee. 2016. “Good Laboratory Practice for Optimization Research”. *Journal of the Operational Research Society* 67(4):676-689.

Kurkowski, S., T. Camp, and M. Colagrosso. 2005. "Manet Simulation Studies: The Incredibles". ACM *SIGMOBILE Mobile Computing and Communications Review* 9(4):50-61.

Lorig, F., D. S. Lebherz, J. O. Berndt, and I. J. Timm. 2017. “Hypothesis-driven Experiment Design in Computer Simulation Studies”. In *Proceedings of the 2017 Winter Simulation* Conference, edited by W. K. V. Chan et al., 1360–1371. Piscataway, New Jersey: IEEE.

Monks, T., C. S. M. Currie, B. S. Onggo, S. Robinson, M. Kunc, and S. J. E. Taylor. 2018. "Strengthening the Reporting of Empirical Simulation Studies: Introducing the Stress Guidelines". *Journal of Simulation*, published online doi: 10.1080/17477778.2018.1442155.

Moreau, L., B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, B. Plate, Y Simmhan, E. Stephan, and J. Van den Bussche. 2011. “The Open Provenance Model Core Specification (v1. 1)”. *Future Generation Computer Systems* 27(6):743–756.

Peng, D., R. Ewald, and A. M. Uhrmacher. 2014. “Towards Semantic Model Composition via Experiments”. In *Proceedings of the 2014 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, edited by R. Fujimoto, et al., 151–162. New York: ACM.

Peng, D., T. Warnke, F. Haack, and A. M. Uhrmacher. 2016. “Reusing Simulation Experiment Specifications to Support Developing Models by Successive Extension”. *Simulation Modelling Practice and Theory* 68:33–53.

Rabe, M., S. Spieckermann, and S. Wenzel. 2009. “Verification and Validation Activities within a New Procedure Model for V&V in Production and Logistics Simulation”. In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti et al., 2509–2519. Piscataway, New Jersey: IEEE.

Rahmandad, H. and J. D. Sterman. 2012. "Reporting Guidelines for Simulation-Based Research in Social Sciences". *System Dynamics Review* 28 (4):396-411.

Reinhardt, O., A. Ruscheinski, and A. M. Uhrmacher (2018). “ODD+P: Complementing the ODD Protocol With Provenance Information”. In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe et al. (to be published).

Ribault, J., and G. Wainer. 2012. "Using Workflows and Web Services to Manage Simulation Studies (WIP)". In *Proceedings of the 2012 Symposium on Theory of Modeling and Simulation-DEVS Integrative M&S Symposium*, edited by G. Wainer et al. San Diego: SCS.

Robinson, S. 2008. "Conceptual Modelling for Simulation Part I: Definition and Requirements". *Journal of the Operational Research Society* 59(3):278-290.

Ruscheinski, A., D. Gjorgevikj, M. Dombrowsky, K. Budde, and A. M. Uhrmacher. 2018. "Towards a PROV Ontology for Simulation Models". In *Proceedings of the 2018 International Provenance and Annotation Workshop*, edited by A. Gehani et al., London: Springer, to be published.

Ruscheinski, A. and A. M. Uhrmacher. 2017. "Provenance in Modeling and Simulation Studies – Bridging Gaps". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 872–883. Piscataway, New Jersey: IEEE.

Rybacki, S., S. Leye, J. Himmelspach, and A. M. Uhrmacher. 2012. "Template and Frame Based Experiment Workflows in Modeling and Simulation Software with WORMS". In *Proceedings of the Eighth IEEE World Congress on Services, 24th-29th June, Honolulu, Hawaii, USA (SERVICES 2012)*, 25–32, IEEE

Shannon, R.E. 1998. "Introduction to the Art and Science of Simulation." In *Proceedings of the 1998 Winter Simulation Conference*, edited by D. J. Medeiros et al., 7-14. Piscataway, New Jersey: IEEE.

Taylor, S. J. E., A. Anagnostou, A. Fabiyi, C. Currie, T. Monks, R. Barbera, and B. Becker. 2017. "Open Science: Approaches and Benefits for Modeling & Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 535-549. Piscataway, New Jersey: IEEE.

Teran-Somohano, A., A. E. Smith, J. Ledet, L. Yilmaz, and H. Oguztüzün. 2015. "A Model-driven Engineering Approach to Simulation Experiment Design and Execution". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz et al., 2632–2643. Piscataway, New Jersey: IEEE.

Uhrmacher, A. M., S. C. Brailsford, J. Liu, M. Rabe, and A. Tolk. 2016. "Panel – Reproducible Research in Discrete Event Simulation - A Must or Rather a Maybe?". In *Proceedings of the 2016 Winter Simulation Conference*, edited by W. K. V. Chan et al., 1301–1315. Piscataway, New Jersey: IEEE.

Wolstencroft, K., R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J.Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A.N. de la Hidalga, M.P.B. Vargar, S. Sufi, and C. Goble. 2013. "The Taverna Workflow Suite: Designing and Executing Workflows of Web Services on the Desktop, Web or in the Cloud". *Nucleic Acids Research* 41(W1):W557–W561.

AUTHOR BIOGRAPHIES

SIMON J. E. TAYLOR is a full Professor and leader of the Modelling & Simulation Research Group in the Department of Computer Science, Brunel University London (https://tinyurl.com/ya5zjh8z). He leads major projects in industry and Africa. He is a member of the ACM SIGSIM Steering Committee and founder of the Journal of Simulation. He has chaired several major conferences and has published over 150 articles. His email address is simon.taylor@brunel.ac.uk and his ORCID is 0000-0001-8252-0189.

TILLAL ELDABI is senior lecturer at Brunel University. He has a B.Sc.in Econometrics and a M.Sc. and Ph.D. in Simulation Modeling in Healthcare. His research is focused on aspects of healthcare modeling and simulation. He developed models to support health economists and clinicians to decide on best treatment programs. He has published widely in the field of modeling in healthcare and edited a number of special issues in related journals. His email address is tillal.eldabi@brunel.ac.uk.

THOMAS MONKS is the Director of Data Science at NIHR CLAHRC Wessex, University of Southampton. His research interests are applied simulation modeling, optimization and machine learning in healthcare. His email address is thomas.monks@soton.ac.uk and his ORCID is orcid.org/0000-0003-2631-4481.

MARKUS RABE is full professor for IT in Production and Logistics at TU Dortmund University. Until 2010 he had been with Fraunhofer IPK in Berlin as head of the corporate logistics and processes department, head of the central IT department, and a member of the institute direction circle. His research focus is on information systems for supply chains, production planning, and simulation. Markus Rabe is vice chair of the "Simulation in Production and Logistics" group of the simulation society ASIM, member of the editorial board of the Journal of Simulation, member of several conference program committees, has chaired the ASIM SPL conference in 1998, 2000, 2004, 2008, and 2015, and was local chair of the WSC'2012 in Berlin. More than 190 publications and editions report from his work. His e-mail address is markus.rabe@tu-dortmund.de.

ADELINDE M. UHRMACHER is Professor at the Institute of Computer Science, University of Rostock and head of the modeling and simulation group. She is editor in chief of the ACM Transactions of Modeling and Computer Simulation. Her interests are in developing methods for multi-level modeling and simulation and their applications in areas such as cell biology, demography, or ecology. Her email address is adelinde.uhrmacher@uni-rostock.de.