CSER 2020

Investigating Model Credibility within a Model Curation Context

Donna H. Rhodes*

Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

Abstract

Model curation can be thought of as a hallmark of digital maturity, signifying that models are highly valuable assets of the enterprise used as a trusted basis for engineering decisions. As enterprises begin to develop large model repositories, model credibility becomes a central concern. Recent studies show the decision to use, reuse and repurpose models is contingent on the model consumer's perception of validity and trustworthiness of the model. This paper discusses an investigation of selected foundational works on credibility of models, simulations and websites as part of a larger research effort on model curation. The objective is to leverage findings and strategies from prior work, and identify useful heuristics that can inform model credibility within the context of model curation.

Keywords: model curation, model credibility, model discovery, model consumer, trust

1. Introduction

The transformation to digital engineering has brought model curation to the forefront of systems engineering research. Model Curation can be defined as the *lifecycle management, control, preservation and active enhancement of models and associated information to ensure value for current and future use, as well as repurposing beyond initial purpose and context*. Curation activities include model governance, acquisition, accession, valuation, preservation, active enhancement, model discovery, and archiving. Curation practices promote formalism and provide for the strategic management and control of models and associated digital artifacts, particularly when managed as a collection at the enterprise level. Model curation infrastructure will better enable an enterprise to establish and actively enhance the collection of models that are of value to the

* Corresponding author. Tel.: +1-617-324-0473
E-mail address: rhodes@mit.edu

enterprise. [1, 2]. As evidenced by curation practice in institutional collections (e.g., museum, historical society, libraries), dedicated leadership and support roles are needed to carry out curation processes. Moving into the future, model curation can be expected to involve unique responsibilities at the enterprise level, motivating a new leadership role of model curator. [3]. The lack of access to models, mistrust of models, and perception of legitimacy of models are all barriers to model reuse and longevity, which can potentially be mitigated by model curation. The terms *model* and *simulation* are used in this paper in a broad manner; useful definitions are specified by NASA [14].

> **Model**: A description or representation of a system, entity, phenomenon, or process (adapted from Banks (1998). **Note:** A model may be constructed from multiple sub-models; the sub-models and the integrated sub-models are all considered models. Likewise, any data that goes into a model are considered part of the model.
>
> **Simulation:** The imitation of the behavioral characteristics of a system, entity, phenomenon, or process.

The most recently published version of the NASA Standard for Models and Simulations (2016) illustrates the life cycle of M&S as shown in Figure 1. As completed models are selected for curation, model curation can be thought of as overlapping the latter two phases, model use/operations and model archiving. Additionally, an enterprise model collection would come into play at the start the life cycle, where reuse of a model is an option that replaces the development of a new model.

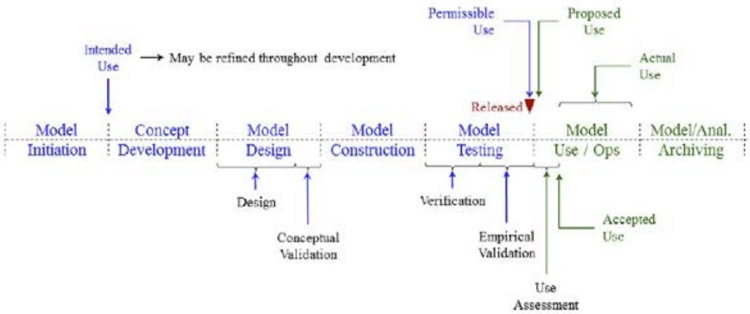

Figure 1 M&S Life Cycle from NASA-STD-7009A W/CHANGE 1 [12, p. 70]

2. Model curation

Not all models are suitable for curation. Curation applies to longer duration models, rather than those developed for a quick study or to simply work out a problem. Two broad categories of longer duration models that necessitate model curation are: (1) models used throughout the lifespan of a program; and (2) models to be intentionally reused for other purposes/contexts. The former includes, for instance, the set of models comprising a digital twin. The latter includes reference models and "platform" models, which enable enterprise to reuse and re-purpose models. [2].

Model curation use cases need to be generated through investigating the myriad situations under which curation is applicable and valuable to the enterprise and its stakeholders. This will depend uniquely on the enterprise, program characteristics and specialized circumstances, including its strategic technology and business roadmaps. Model credibility is a key consideration because there is minimal value in placing a model under curation if the credibility is in question. This investigation is motivated by the belief that the foundational ideas and approaches for model credibility are worth re-examining respective to a model curation context. This context has many facets, including: curation practice, requisite leadership, supporting infrastructure, required competencies, curating for model consumer needs, and application of innovations to enhance model discovery [3]. Model credibility is a construct that underpins all of these facets.

3. Model credibility

Model credibility and its associated constructs (model confidence, model trust, model validation, model value, etc.) have been investigated and discussed in the literature for more than four decades. The earliest works come from the operations research and simulation communities [1, 2, 3]. Our ongoing research is re-examining the prior work on credibility within the context of model curation. As revealed through past investigation of model credibility, there are actions that can be taken in curation practice and the supporting infrastructure to increase the likelihood that a model consumer will perceive a model as trustworthy and valid.

The following sections review selected works on model and simulation credibility. In addition to model credibility, a body of work on website credibility is discussed. It is worth noting that websites at the time of the studies were dense and content-rich static information, as contrasted with today's highly visual and more interactive experience. As such, this is an informative body of work for digital engineering in that the websites in the early website credibility studies are more akin to digital artifacts.

3.1 Credibility of models and simulations

Kahne (1976) proposed a new approach for examining model credibility for large-scale systems, asserting "model credibility is separated into two distinct, if not independent, issues: validity and value" (verification is assumed in is paper). Kahne states, "To be a bit more specific, we are contrasting verisimilitude (having the appearance of truth) with worth (value). For convenience we will use the words validity and value." He notes that credibility of a model will depend, among other things, upon the quality of the match between the model and the model user", noting that the model reflects the biases and outlook of the modeler. A novelty of his approach is to take the viewpoint of buyer/seller, with a subjective approach to credibility-type questions where credibility is defined as *capable of being believed*. [5]

In 1979, the SCS Technical Committee issued a report on *terminology for model credibility*. This was motivated by the desire to develop a standard set of terminology to facilitate effective communication between the builder of a simulation model and its potential users, believed to be the "cornerstone for establishing the credibility of a computer simulation" [6]. This committee provided a framework to review credibility of a simulation (Fig. 2). This framework divides the simulation environment into three basic elements. Inner arrows describe processes that relate the elements to each other. The outer arrows refer to procedures that are used to evaluate credibility of the processes.

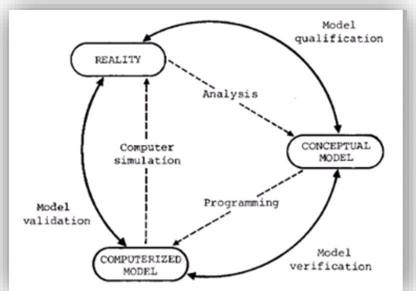

Figure 2 SCS Framework to Review Credibility of a Simulation [6].

Gass & Joel (1981) investigated the concepts of *model confidence*, showing model confidence to be not an attribute of a model, but of the model user. [7] Seven confidence criteria the authors proposed are model definition, model structure, model data, computer model verification, model validation, model usability, and model pedigree. The latter (originally called model demographics) is especially pertinent to perception of credibility, given its subjectivity. They state pedigree "should enable the decision maker to determine the model's status with respect to past achievements, theoretical and methodological state-of-the-art, and the expert advice that went into its development [7]. Gass (1993) states that critical to use of a model is "the credibility or confidence that the decision maker has in the model and its ability to produce information that would be of value to the decision makers" [8].

Balci (1986) proposed comprehensive guidelines for assessing credibility of simulation results. He characterizes a life cycle of simulation study as richly characterized with 10 phases, 10 processes and 13 credibility assessment stages (Figure 2).

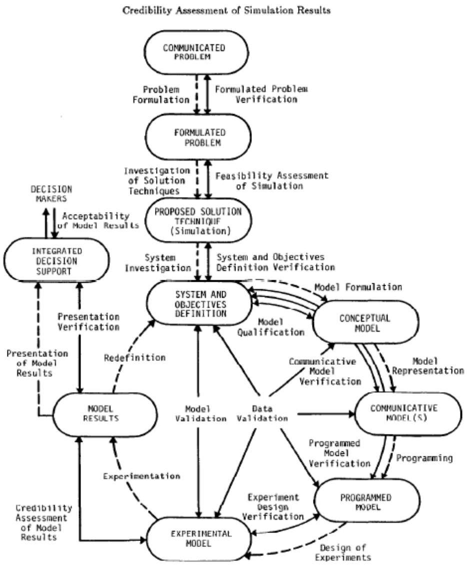

Figure 3. Balci's Life Cycle of a Simulation Study [10]

Balci's important work demonstrates that credibility assessment is complex and involves staged assessment through the lifespan of a model or simulation. [10]. His work demonstrates that during development, the acceptance of the model is a result of the model consumer's cumulative perception of validation efforts. This suggests the importance of giving a model consumer transparency into the series of validation activities that went into the original development, not only the end result.

3.1.1 Assessment of credibility of models and simulations

Steele (2008) reveals the insights and thinking behind the NASA's standard for models and simulations (M&S). Eight relevant factors of credibility were identified during the development of this standard, which defines credibility as *the quality to elicit belief or trust in M&S results* [10]. The evolution of the NASA standard surfaced various dimensions of credibility, and more recently an assessment approach. State of the practice on model credibility assessment in the systems field has emerged as part of the NASA efforts over more than a decade. A method for M&S credibility assessment is described in Appendix E of the 2016 update of *NASA Standard for Models and Simulations* (2016). [4, pp. 55- 72]. Ahn et al. (2013) propose a formal procedure based on the NASA standard to assess the credibility of an M&S in an objective way using the opinions of an expert group for credibility assessment and a Delphi approach [11], initially piloted on an M&S platform called SpaceNet by Ahn & de Weck in 2007 [12].

3.2 Website credibility

The Stanford University Persuasive Technology Lab, founded and directed by BJ Fogg, has investigated captology (Computers as Persuasive Technologies) over the past two decades, with early studies on website credibility. The larger body of research of the lab seeks to create insights into how computing products can be designed to change people's beliefs and behaviors. An early study (Fogg, et al., 2000) aimed to assess a broad range of elements that impact varying aesthetic, context, and technical factors on credibility of websites. [13].

Fogg et al. (2001) state "simply put, credibility can be defined as believability" and is a perception based on two factors: trustworthiness + expertise. Additional clarifying points are: (1) credibility is a perceived quality; it does not reside in an object, person or piece of information; and (2) when discussing credibility of a computer product, one is always discussing a perception of credibility. [14]

Several findings of this work are insightful for model curation. First, web credibility was found to increase when users perceive a real world organization and real people behind a website. Second, small errors had a large negative impact on credibility of a website. Third, the users view websites as less credible if they experience technical problems (e.g. delays in download of information). Fogg states, "if users think a site lacks credibility – that the information and services cannot be trusted – they will abandon the site and seek to fill their needs in other ways". [14] This appears to suggest that a poorly designed model repository would be a significant deterrent to the success of model curation in an enterprise. The research by Fogg et al. led to a collection of website design guidelines; several of these appear to be applicable for design of model curation infrastructure and enablers.

3.3 Recent research on model confidence and trust

Research performed by Flanagan (2012) uses case studies and a web-based experiment to investigate key challenges to model-based design: distinguishing model confidence from model validation. [15] The objective of her research is to understand factors that cause perception of model quality to differ from actual quality. She proposes eight factors as the key variables to misaligned model confidence, and tests hypotheses for six of these in the experiment to illustrate the effect of the factors on perception of model credibility. According to Flanagan, these factors can potentially help explain behavior of decision makers, especially in the situation where "the model would be a good tool to help solve a problem; however, the decision-maker does not agree and continues without input from the model, effectively dismissing its predictions" [15]. One of the hypotheses that was validated in the experiment, and is most relevant to this model curation, concerns source and transparency of the model. This hypothesis is: *A more trustworthy model author and transparent governing equations will improve model perception*. Her finding was that for cases where the source was important to the decision, there was a significant difference in the decision outcome where untrustworthy sources caused reduced confidence. Flanagan's research, while preliminary, demonstrates the value of further research of this nature.

These findings are consistent with findings of a more recent empirical study by German and Rhodes (2017) on model-centric decision making and trust [16]. Model credibility was found to be a perceived quality, positively impacted by tailorable transparency and available model pedigree (detailing who originated the model, who subsequently enhanced the model, assumptions made, expertise of modelers, etc. They also found that while not always needed, model consumers must have available model transparency when determining if a model should be trusted in making a specific decision.

3.4 Recent research on overall credibility

Recent work by De Vin {2015) provides a significant discussion on credibility of simulation, stating it "…is thus influenced by three factors: Credibility of the model, credibility of the data, and credibility of the model use." [17, p. 152]. He notes that without credible data (also called Data Pedigree as discussed in [4]), it will not be possible to carry out trustworthy validation of the model. De Vin's paper "uses the NASA CAS model for credibility assessment of simulations to arrive at a schematic representation of how overall credibility as composed of aspect related to the model, the data, and the model's use". [17]

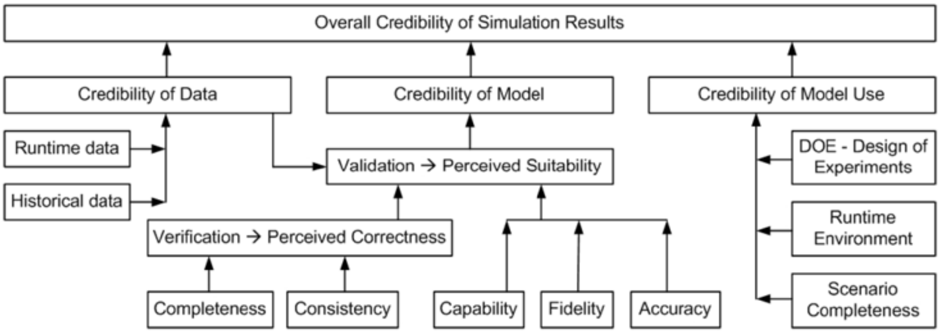

Figure 4 DeVin's schematic of factors influencing the overall credibility of simulation results [17, p 154]

4. Toward design guidelines for model curation

Model curation requires both initial and ongoing investment, which must be outweighed by the payoffs (e.g., lifecycle model usability, re-purposing of models for new contexts, re-use of models at the enterprise level, etc.). This would be a poor investment if model credibility is neglected. In the spirit of Kahne's buyer/seller approach, it may be useful to think about model curation from a model acquirer (consumer)/model curator approach. The model acquirer's perception of the expertise and authority of the model curator will have influence on perception of credibility.

Model curation practices and infrastructure must be designed to support quality of individual model consumer experience and the capacity to foster perceived trustworthiness of the model. Model curation infrastructure (model repositories, interfaces for repository access, etc.) need to designed for cost effectiveness and security, as well as for the quality of the experience of human interaction. As can be inferred by Fogg (2000, 2003), a poorly designed model repository would negatively impact perceived credibility, as would a poorly designed user experience with access and interaction with the repository. The design implications resulting from the website credibility studies of Fogg et al. offer practical guidance for model curation. For example, including markers of expertise and markers of trustworthiness could be implemented through model pedigree information.

4.1 Heuristics for Model Curation

The technological challenges for creating model collection repositories are quite significant. The social, cognitive and perceptual challenges are equally – if not more – challenging, yet are more difficult to comprehend and address. These challenges must all be addressed for the future success of digital engineering.

Model credibility is now generally accepted as a property of the model perceiver, and there are observed behaviors and useful strategies that the systems community can adapt in support of digital engineering goals. The foundational papers and on credibility suggest a derived set of heuristics toward establishing design guidelines that can be used by the systems community in evolving model curation practice and enabling technology. Continuing investigation is expected to contribute to expanding and refining these heuristics, and formulating respective design guidelines. An initial set of nine heuristics are proposed below.

1. Model credibility is an attribute of the model consumer, not the model.
2. Model credibility is positively influenced by communication between modeler and model consumer, both active and passive.
3. Credibility of digital artifacts is influenced by both trustworthiness and expertise of a model consumer.
4. Acceptance of a model for (re)use is influenced by a model consumer's belief that the model has the ability to produce information of value to them.
5. Credibility of models in a collection influences a model consumer's trust in the enabling infrastructure.
6. A model consumer's experience in discovering and retrieving models from a repository influences perceived credibility of the model.
7. Model credibility is influenced by a model consumer's trust in the expertise of the model originator, as well as modelers who subsequently enhance and maintain the model over time.
8. Model credibility is influenced by a model consumer's capacity for transparency into the validation activities throughout its development and enhancement.
9. Credibility of the model collection is influenced by a model consumer's perception of expertise of the governance authority that accepted the model into the collection.

5. Conclusion and further research

Significant research is needed to realize the promise of model curation for the systems field [1, 2, 3]. The ongoing larger research project is investigating four facets of model curation. The first is model curation implementation practices for performing model curation activities at the enterprise level. These must be harmonized with existing model management practices and supporting practices such as configuration management and data management. A second facet of the research investigates precursors, enablers and barriers to model curation practice, infrastructure, and approaches to the curating of models. The third facet concerns innovations through myriad newer methods and technologies that could be applied in model curation, especially in model discovery and curating for consumer needs. The fourth concerns competencies, roles and responsibilities. Model credibility is an overarching consideration across these four areas of research. The formulation of heuristics that draw from prior work and other fields can be used to inform the strategies and enablers for model curation such as model pedigree, accession records, model valuation.

This paper considers selected prior research on model credibility, simulation credibility, and website credibility. Data credibility is not explored in this paper but is also of central importance to model curation. The findings suggest that prior work on credibility has significant relevance for model curation and useful strategies and practices from other fields can be adapted for digital engineering. Further investigation into the substantial body of work on credibility within the model curation context is ongoing. The linkage between model credibility and data credibility needs further examination within the model curation context. Additionally, other areas remain to be explored, including credibility in regard to information retrieval, data curation, human-computer interaction, and augmented intelligence.

Acknowledgements

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract HQ0034-13-D-0004. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Department of Defense.

References

[1] Rhodes D.H. (2018). Interactive Model-Centric Systems Engineering Technical Report, Phase 5. (No. SERC-2018-TR 104). Systems Engineering Research Center Hoboken NJ.

[2] Rhodes D.H. (2018). Interactive Model-Centric Systems Engineering Technical Report, Phase 6. (No. SERC-2019-TR 003). Systems Engineering Research Center Hoboken NJ.

[3] Rhodes, D.H. (2019). Model Curation: Requisite Leadership and Practice in Digital Engineering Enterprises. 17th Conference on Systems Engineering Research. Washington, DC.

[4] NASA. (2016). NASA-STD-7009A w/CHANGE 1, Standard for Models and Simulations https://standards.nasa.gov/standard/nasa/nasa-std-7009, accessed 30 Oct 2019.

[5] Kahne, S. (1976). Model Credibility for Large-Scale Systems. *IEEE Transactions on Systems, Man and Cybernetics.* Aug. 587- 590.

[6] SCS (1979). Terminology for Model Credibility. *Reports of the SCS Technical Committees*.

[7] Gass, S.I. and Joel, L. (1981). Concepts of Model Confidence. *Computers and Operations Research*. 8(4), 341-346.

[8] Gass, S. (1993). Model accreditation: A rationale and process for determining a numerical rating. *European Journal of Operational Research*, 66(2), 250-258.

[9] Balci, O. (1986). Credibility Assessment of Simulation Results. Proceedings of the 1986 Winter Simulation Conference.

[10] Steele, M. J. (2008, June). Dimensions of credibility in models and simulations. Proceedings of the 2008 Summer Computer Simulation Conference (57). Society for Modeling & Simulation International.

[11] Ahn, J., de Weck, O, Steele, M. (2014). Credibility Assessment of Models and Simulations Based on NASA's Models and Simulation Standard Using the Delphi Method. *Systems Engineering,* 17 (2), pp. 237 – 248.

[12] Ahn, J., de Weck, O.L. (2007). Pilot study: Credibility assessment of SpaceNet 1.3 with NASA-STD-(I)-7009, Tech Rep. MIT, Cambridge, MA.

[13] Fogg, B.J., Marshall, J., Osipovich, A., Varma, C., Laraki, O., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P. and Treinen, M., (2000). Elements that affect web credibility: early results from a self-report study. In *CHI'00 extended abstracts on human factors in computing systems* (pp. 287-288). ACM.

[14] Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P. and Treinen, M., (2001). What makes Web sites credible?: a report on a large quantitative study. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 61-68). ACM.

[15] Flanagan, G. (2012). Key Challenges to Model-Based Design: Distinquishing Model Confidence from Model Validation. Master's Thesis. Massachusetts Institute of Technology.

[16] German, E.S. and Rhodes, D.H. (2017). Model-Centric Decision-Making: Exploring Decision-Maker Trust and Perception of Models. 15th Conf. on Systems Engineering Research. Los Angeles, CA.

[17] De Vin, L. (2015). Simulation, Models, and Results: Reflections on their Nature and Credibility. Proceedings of FAIM 2015. (pp. 148-155).