The following paper will be published and presented at the 2016 IEEE International Systems Conference in Orlando, Florida, 18-21 April, 2016.

The copyright of the final version manuscript has been transferred to the Institute of Electrical and Electronics Engineers, Incorporated (the “IEEE”), not excluding the retained rights of the manuscript authors. Reproduction, reuse, and distribution of the final manuscript is not permitted without permission.
Considerations for Model Curation in Model-Centric Systems Engineering

Lucie Reymondet  
Systems Engineering Advancement Research Initiative  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
lucier@mit.edu

Adam M. Ross  
Systems Engineering Advancement Research Initiative  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
adamross@mit.edu

Donna H. Rhodes  
Systems Engineering Advancement Research Initiative  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
rhodes@mit.edu

Abstract—Contemporary systems are often highly complex sociotechnical systems that require models to help system engineers with sense-making and decision-making. The systems community has developed and instantiated many modeling approaches, practices, formal languages and toolsets, which are all areas of progress. If models and their instantiations could be managed as assets, documented, archived, protected, retrieved and re-used as such, modeling and analysis tasks would likely gain in quality and timeliness. This paper asks the question “What would a curator need to know about models or their instantiations to provide a model curation function”? Considerations of the activities and body of knowledge associated with curation of models are presented. The potential usefulness of a curated system modeling approach is illustrated in an example.

Keywords—socio-technical systems, systems of systems, models, model-centric, model curation

I. INTRODUCTION

Contemporary systems involve strongly interwoven technical systems and social players within an environmental context that influences both social and technical spheres. These types of systems (e.g., Systems of Systems, Engineering Systems, Complex Socio-Technical Systems) display much complexity in the eyes of systems engineers: they are difficult to understand, study, predict, control or change. Model-centric approaches, as discussed in the paper, provide a means of simplifying, understanding, and explaining such systems. While models themselves are key, human-model interaction is an important aspect of model-centric engineering.

The Interactive Model-Centric Systems Engineering (IMCSE) project within the MIT Systems Engineering Advancement Research Initiative (SEArI) is a multi-year project that aims to develop transformative results in engineering projects through advancing the theory and practice of intense human-model interaction. An invited IMCSE Pathfinder Workshop was conducted in January 2015, to foster an initial dialogue on human-model interaction. Research needs were identified from both a model-centric perspective and an interactive perspective [1, 2, 3]. The workshop participants agreed that progress has been made on standards, methods and techniques for model-based systems engineering, yet little attention has been given specifically to human-model interaction needs and challenges. This has subsequently been further confirmed with other systems practitioners and researchers. Model curation was cited as an important topic for investigation in evolving model-centric engineering.

II. MOTIVATION

Given a shift in engineering to a model-centric paradigm, there is a need to better understand where the role of the human (versus automation or “AI”) is essential in the effective management and utilization of model environments. This includes the models (managed as organizational assets), supporting infrastructure, and the associated protocols and practices. Digitized legacy systems information and new digital system models will provide the basis for designing and evolving systems into the future. This drives the criticality of models as assets and necessitates change in model-related policy and practices. Accordingly, there is an urgent need to mature a practice of “model curation” including a “model curator” functional role within engineering organizations.

Model curation has been raised as a topic of growing importance within the systems community. Rouse [4] stresses that the wealth of existing models is often not used because of a lack of knowledge of these resources and the difficulty in accessing them. He asserts there needs to be a single point of access to this body of knowledge, enabled with downloading
computer codes, documentation on assumptions and use, and dissertations on the development and validation of these models.

The 2015 IMCSE Pathfinder Workshop participants agreed that there is no consistent meaning across or among the “sea” of models that are used or exist. There is a lack of precise semantics, especially in the behavior and timing/interactions of models [1]. Many organizations are beginning to develop repositories of models and archive associated data, which is one of the fundamental activities in model curation. Repositories support re-use of building blocks from previous analyses and tailoring to the specific requirements of one’s problem. Potential benefits include improving the quality and timeliness of analysis, by developing a repeatable framework providing guidelines for accessing, designing, and implementing models.

Evolving the science of model curation will involve a partnership of government, industry and academia. There are many challenges and interesting problems related to this topic, and there is great potential benefit for practitioners, researchers and educators. As shown in Table 1, each contributes in many ways. Further consideration is needed to understand the role of model curator and impact this can have in the field.

Table 1: Selected examples of benefits and contributions related to model curation

<table>
<thead>
<tr>
<th>Example Benefits</th>
<th>Example Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Practitioners</strong></td>
<td><strong>Submit validated</strong></td>
</tr>
<tr>
<td>• Have access to state-of-the-art research in modeling and simulation</td>
<td><strong>models and data to in-house and/or shared repositories</strong></td>
</tr>
<tr>
<td>• Have timely access to validated modeling resources and support from curation specialists</td>
<td><strong>Propose model improvement based on feedback from real use cases implementations</strong></td>
</tr>
<tr>
<td>• Evaluate different modeling alternatives before investing heavily in one</td>
<td><strong>Work towards community-shared semantics in model-centric activities</strong></td>
</tr>
<tr>
<td><strong>Researchers and Educators</strong></td>
<td><strong>State-of-the-art research and impact on modeling and simulation tools and methods</strong></td>
</tr>
<tr>
<td>• Have research opportunities to work on interesting model-centric topics</td>
<td><strong>Strengthen the theoretical grounds of model curation and human-model interaction</strong></td>
</tr>
</tbody>
</table>

III. ROLE OF A MODEL CURATOR

As engineering practice becomes increasingly model-centric, models are valuable assets for designing and evolving systems. The need for model curation accordingly becomes a necessary functional role in organizations. This is a relatively new idea in the systems community, though progress has been made on some of the activities involved in curation. Maturing an approach for model curation in the systems engineering field can leverage the work of other related curation practices.

Rusbridge et al. [5] say “Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle.” Various organizations focused on digital curation have been established. For instance, the Digital Curation Centre (DCC) in the UK is a leading center of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK’s higher education research community. As digital curation is closely related to modeling curation, there is much to be borrowed and adapted from this practice. Practices on collaboratively developed model repositories and their management provide additional insights for model curation [6]. Another related area is social curation, focusing on collaborative sharing of Web content organized around one or more particular themes or topics.

Curation of institutional collections, such as the role of the traditional museum curator, offers an analogy for the many potential responsibilities of the model curator. As with a large institution, the curation function is carried out by a team of individuals and the same would be true of the model curation in the systems field. The museum curator’s role is an essential one where highly knowledgeable curators oversee collections of artwork and historic items, with support from archivists who appraise, edit, and maintain permanent records and historically valuable documents. Specialists and technicians are also involved in various capacities.

Extending from the various types of curation roles and activities of other fields, the model curator’s role is envisioned to include a number of major responsibilities and support of various staff. The model curator (curation function) would set and administer model-related policies and practices. The curator would ensure models and related documents are authenticated, preserved, classified and organized accordingly with model metadata standards, to be defined. The curator may own the data management for models and related information, or oversee the ownership by other individuals or organization. As needed, a curator would meet with individuals and teams, who will create, use and re-use models, helping to determine a useful classification of individual and sets of models. At the organization level, the curator may organize training and special projects related to model-based engineering. The curator may also participate in creating and maintaining model-based work environments.

The model curator role needs to become a formal role in engineering enterprises, and curators will need to know about a number of things including model ontologies, model metadata, latest modeling techniques and classes of models,
policies on data rights, code of ethics, and others. Effective model curation necessitates clarity across the systems community in characterizing and handling models. It requires formalizing knowledge of models and determining a distinctive set of model characteristics (purpose, input/output types, logic, assumption types, model incompatibilities, etc.). A model classification drawn from prior use is a first step towards generating metadata for curating models. There are many facets of model curation needing further investigation and elaboration. The next section discusses preliminary work toward one of these facets.

IV. DEVELOPING A SHARED UNDERSTANDING OF MODELS

Successful model curation will require a shared understanding of numerous aspects of models. A goal of this paper is to provide preliminary ideas to inspire further dialogue in the systems community, focusing in on the topics of model purpose, model classification, model selection, and model composition guidelines.

A. Model Purpose

The selection of an appropriate modeling approach is determined by the question or problem being addressed. It is recognized that a model is not a full representation of the system. A model should be a useful representation, and draws its goodness from being simple, yet evocative [7, 8]. Therefore, the purpose for modeling sets the boundary for model depth and breadth.

Various types of models have been enumerated by the systems engineering community, but there appears to be insufficient attention given to model purpose itself. In the field of informatics, McBurney [10] proposed nine model purposes:

1. To understand, predict or control natural reality.
2. To understand or predict an existing human phenomenon or system.
3. To understand, predict or control future human phenomena or artificial systems in design or development phase.
4. To serve as a locus for discussion between stakeholders, to enable alternative exploration in a structured and shared way.
5. To identify, articulate and potentially resolve trade-offs, action options, and their consequences.
6. To enable rigorous, structured and justified thinking about assumptions and their relationships to one another.
7. To train people in expedited and focused experiences of reality.
8. To enable stakeholders to learn about and assess the assumptions, reasoning process and action plans of the modelers.
9. To play, to enable the exercise of human intelligence, ingenuity and creativity, in developing and exploring the model.

When tackling the complexity of a technical system such as an aircraft, modelers have the intent to classify, explain, predict, control, detect or diagnose the system’s input/output behavior. As for human behavioral and social phenomena, different modeling approaches are appropriate at the people, process, organization and ecosystem scales [4]. For example, at both ends of the spectrum, models at the people level will seek to detect, explain and resolve anomalies in human behaviors (e.g., consumer behavior), while models at the ecosystem level will seek to detect, explain and resolve anomalies in societal behaviors (e.g. intra-firm competitive relations). Lower-level model outputs serve as inputs for higher-level models and higher-level outputs are constraints on lower levels of modeling.

Zacharias et al. [11] point out that individual, organizational and societal models, do not predict exactly what humans will do, as individuals or in groups, but rather help forecast a range of potential action outcomes, draw attention to potential unintended consequences, and highlight variables that are overlooked in a particular situation. Accordingly, model purposes include: to analyze fragmented information and develop courses of action based on the likelihood of desired outcomes; to train personnel, simulating the environment, dynamics and providing performance feedback; and to design and evaluate a technical system, predict its performance and make decisions based on cost-benefit tradeoffs.

A shared understanding of model purposes would facilitate the critical transition from the motivation for modeling (e.g. perceived complexity in a system, or issues perceived in an organization), to the choice of an adequate set of models (e.g. to tame the system complexity, or to help decide upon management strategies). Empirical observations suggest the manner in which this is performed today is mostly ad hoc, or biased by practice anchoring within a field or a group.

B. Model Classification

The systems community has instantiated many different models to support designing, evaluating, and testing technical systems (e.g. requirements/function diagram, computer-aided design, design structure matrix, multi-attribute trade space exploration, cross impact analysis, fault tree analysis, hardware-in-the-loop/human-in-the-loop simulation…), analyzing social networks and phenomena (e.g. system dynamics, queuing models, agent based models, graph theory models…), and understanding human behavior (e.g. mental models, Petri net models, task models, microeconomics models…) among others.

A suggested first step in making the transition from the motivation for, to the selection of, a model would be to examine one’s issue and expectations from the model against previous instances. In this endeavor, a modeler would use a classification of previously adopted modeling approaches, as shown in Figure 1, with a profile of each model instance along
several characteristics, including model purpose. It is also expected that such a survey would highlight limitations of each modeling approach in its context of use, which are equally important as the purpose.

Model characteristics would provide additional elaboration to the metadata that is needed in organizing the envisioned classification, to enable case-specific selection of models. Table 2 lists selected examples for different model characteristics.

### TABLE I. SELECTED EXAMPLES OF MODEL CHARACTERISTICS

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outputs</td>
<td>Conceptual/Quantitative, Statistic/Deterministic, Visualization format, Sensitivity diagram, Error estimation</td>
</tr>
<tr>
<td>Inputs</td>
<td>Data necessary and available, Initial conditions, Parameterization, Boundary conditions, Required format</td>
</tr>
<tr>
<td>Tool</td>
<td>Computer environment used? Language used? Open source code?</td>
</tr>
<tr>
<td>Assumptions</td>
<td>Variable definitions, Time horizon, Variables approximated as steady or quasi-steady, Variables restricted by upper/lower bounds, Discrete vs. continuous state functions, Mathematical properties (continuity, differentiability etc.), Set of important stakeholders, Markov logic</td>
</tr>
<tr>
<td>Simulation</td>
<td>Enabled or not</td>
</tr>
<tr>
<td>Time</td>
<td>Static or Dynamic, Dynamic: Continuous or Discrete time</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Deterministic, Stochastic, Fuzzy variables; Probability distributions</td>
</tr>
</tbody>
</table>

#### C. Model Selection

Traditional, often computational, methods for systems engineering are inadequate for a complex system such as a sociotechnical system, which includes soft and hard problems, a very large number of technical components and autonomous actors, and emergent behavior created by soft-hard, social-technical, non-linear interactions. Using a set of models enables insight into various aspects of system complexity.

In contrast with consolidative computational modeling of deterministic systems (models as surrogates for real systems), Bankes [8] espouses exploratory modeling of systems, which tend to be plagued with uncertainty and unknown unknowns (models as means of testing hypotheses and exploring ranges of possible outcomes). However, it is argued that exploratory modeling can only produce useful results through a constellation of alternative models. Instead of expending effort to find the “right” model, leveraging multiple different models and comparing their results can support cross-validation of each model and increase decision maker confidence in their results. Ross et al. [9] demonstrate this model-trading concept on evaluative models (e.g. capability, cost, performance models) underlying trade space generation. By using multiple simple models, complexity is exported outside the models to the ensemble of model outcomes, from which modelers and stakeholders must make sense. [8]

Selecting one or several models for a complex case study requires breaking down the problem into scoped questions such that each question translates into a model purpose. For example, designing an urban smart power grid is a complex problem taken as a whole. It could be broken down into the following scoped problems (1) Tradeoff energy production mix design alternatives, predict reliability and vulnerability of each design against dynamic demand loads (technical level); (2) Predict consumer preferences regarding energy bills, effort involved, and value attached to contributing to environmental sustainability (people level); and (3) Optimize dynamic pricing of utilities – organization level - within the rules set by local government – environment. Rouse [4] proposes such a multi-level modeling approach for dealing with complex socio-technical systems. Each problem would then call for a tailored modeling approach, for example: (1a) Trade space exploration; (1b) Fault tree analysis; (2) Multi-attribute utility functions, and (3) Microeconomics model of key stakeholders. This model set (1a, 1b, 2 and 3) is neither unique, nor exhaustive and a modeler might only be interested in one aspect of the problem. Therefore, identifying the problem clearly and formulating a clear model purpose is essential in setting the boundaries for the modeling effort.

#### D. Model Composition Guidelines

Recent work on hybrid modeling [12, 13, 14, 15] and multi-scale modeling [4] point towards the usefulness of using not a single but a set of models to study a complex system. LaTour [12] implements a system dynamics model interactively with a trade space exploration model for investigating the impact of time on the lifecycle and procurement of GPS satellites. Mathieu et al. [13, 14] implement a system dynamics model, a Petri net process model and an agent-based model for simulating the response timeliness and effectiveness of the Air.
and Space Operations Center to a series of critical events, at the mission, process and operator levels. Zulkepli [15] implements a system dynamics model and a discrete-event simulation to simulate the interactions between healthcare personnel stress, and patient non-recovery and readmission rates. Synergies between models are seen to increase modeling capabilities and insights into problem solving while reducing the limitations of individual techniques.

Solving the original overarching problem requires being able to integrate different models so as to make sense of the whole. Issues arise especially with computational models, when the integration is automated. Assumption consistency between models and between models and the reality they represent is one issue [4]. For example, assuming a Markovian process in one model when behavior is strongly memory-driven (e.g. autonomous agents) in another is, a priori, inconsistent. Additionally, interfacing two models, especially two computational models, raises issues such as incompatibilities in data types, in naming schemes, in logic mechanisms.

Parallel models run independently from one another, whilst hybrid models are coupled such that the output of one serves as the input for another, and vice versa. Parallel models can be developed and tested individually. Hybrid modeling requires more validation effort, since the models must first be tested individually and then together. However, the risk in aggregating many models is to end up with a model of high resolution but low practical utility and transparency. [8]

When the output of one model serves as the input for another (model chaining, Fig.2A), the link between the two models may be straightforward (e.g. sequential, passing a variable, triggering an event…), involve a transformation (e.g. translating from one modeling language to another, from hard to soft variables, converting units…) or require the intervention of the modeler. In a different configuration, two models are run in parallel and the modeler confronts and analyzes outputs (Fig.2B). The confrontation of results might, again, be straightforward (e.g. variables plotted against each other on a graph), or involve some transformation (e.g. of units), or require evocative visualization. If a model provides unexpected outputs, feedback allows reviewing assumptions and modifying inputs, in a goal-seeking manner (Fig.2C). Such feedback could be automated (e.g. parameter fitting to data) or require human intervention (e.g. hypothesis testing). Combinations of the above can be constructed, increasing the overall modeling effort and complexity.

Guidance for model composition is a part of the larger body of knowledge needed for model curation, in particular for modeling complex systems, where multi-modeling approaches are useful.

Figure 2: Possible Compositions of Models

V. ILLUSTRATIVE CASE EXAMPLE

The number, impact, and geographical dispersion of humanitarian disaster relief operations have been increasing over the past decades. Incidentally, the associated logistics grow in complexity. Transportation and logistics are major components of the operations of relief organizations, and improving efficiency of transportation and logistics systems has potential to significantly decrease operational costs and to improve humanitarian relief services. However, the large number of organizations involved and the diversity of players make effective logistical coordination a challenging task. This often includes United Nations agencies, international and local nongovernmental organizations, the Red Cross Movement, governments of the affected region and military. [16]

Humanitarian logistics is different from commercial supply chain logistics because demand is highly unpredictable until a disaster strikes, and because of unusual constraints such as substantially degraded infrastructure, terrain roughness, political instability, and volatile funding. Regardless of what is being supplied (e.g. food, water, sanitation, housing, shelter), the goals of relief logistics are to optimize inventories ahead of time and throughout the intervention, and to deliver rapidly the correct amount of goods to the needed locations when disasters strike. Essentially all players pursue these same objectives and there is wide consensus that coordination of the effort is important and mutually beneficial, but its implementation is problematic [17]. Organizations each have their own governance structure, imperatives and political interests. The push for profile necessary to attract funding is antagonizing and diverts personnel from the coordination effort. The pressure to distribute vital supplies in the immediate response phase is prioritized over forming collaborative relationships between relief participants, unless such relationships are established prior to the disaster.
Information sharing about the specific needs, available commodities, transport routes, and infrastructure status is key. But heterogeneous information standards, rapidly evolving disasters, scarce and uncertain field data make it difficult to compile meaningful information. Field operatives aren’t necessarily logistics experts or trained to assess logistics needs. If each organization locally decides on supplies to provide and routes to ship along it can eventually clog the logistical system (e.g., ports, warehouses) and create inefficiencies in the relief provided to populations. [18]

It is suggested that NGOs and UN agencies could benefit from various modeling approaches to better anticipate disasters: to shape a collaborative organization and train personnel, to understand where to optimally pre-position warehouses, stocks and vehicles in the network, and to foster cross-organization communication and awareness. Additionally during interventions, further modeling approaches could be used to conduct transport and logistics operations more efficiently: optimize shipments (e.g., routes, cargo vehicle synergies); compile and visualize information from scarce, uncertain data to support decision making; build shared situation awareness horizontally across organizations and vertically between field and management personnel. A model curator would help with identifying, comparing, selecting, and tailoring modeling approaches addressing these high-level problems. This would probably be done iteratively through conversations with involved partners. Furthermore, the curator could draw experience from other complex collaborative sociotechnical systems towards focusing the modeling effort and resources to critical areas of the system.

Architecting the transport and logistics network is a milestone for eventually coordinating relief efforts. Organizations need to know where to place their warehouses relative to those of other organizations, which goods to stockpile in advance and where, what fleet portfolio to maintain and where to pre-position it, how much personnel to hire etc. We lay out three example-modeling contributions a curator could help provide. (1) Based on data from past disasters, a statistical model can be built that visualizes disaster-stricken regions and flows of cargo across the globe. Statistical distributions enable understanding big picture issues and identifying trends. Furthermore, a curator could suggest means for combining past data with Bayesian or machine learning methods to infer a stochastic “map” of potentially high disaster locations and scenarios. This analysis could include current geo-political, climatic, geological, and economic data, for example. (2) Logistics and transport network architectures could be modeled (e.g., warehouse discrete event model, transport routes, information flows) and simulated (e.g. in an agent-based formalism). The curator could suggest using data from a past disaster as needed inputs at the downstream nodes of the network, or generating inputs arbitrarily, exogenously. (3) Alternatively, the model curator could suggest means to combine the two previously developed approaches to assess network architectures against a range of scenarios extracted from the statistical analysis (e.g. Monte Carlo simulation). The performance and feasibility of a logistical configuration in response to any potential demand profile could thereby be evaluated (Fig.3).

![Figure 3: Concept of Model Composition for Simulation](image)

However optimal and technically sound the solutions found by the approach described above, they do not address the reluctance of organizations to change their operations culture, and they do not take into account stakeholder preferences. Hence that approach alone has low chances of getting decision-makers onboard. The curator could draw from experience in modeling multi-stakeholder problems to facilitate building an interactive modeling environment that will engage stakeholders in a discussion or negotiation. A model (e.g. utility model) that quantifies benefits and costs of alternative network architectures or alternative organization architectures, with respect to each stakeholder’s preferences would allow comparing architecture alternatives along a common dimension. Appropriate interactive environments and trade space visualizations would allow exploring tradeoffs between organization and network architecture alternatives as a group and elevate cross-awareness. This requires developing stakeholder-specific cost/benefit models, especially for intangible or non-monetary attributes of utility (e.g. time for delivery, well-being of affected populations). These models are then encapsulated in a larger analysis (e.g. cost-benefit analysis, trade space exploration model) (Fig.4). In creating visualizations of the results of such models, the curator would bring advice to modelers about how to balance message clarity with transparency about underlying assumptions, logic, and approximations of the model.
VI. CONCLUSION

This paper considered the potential for model curation capabilities applied to an example use case. Model composition is a task that would benefit greatly from model curation practices: knowledge about up-to-date modeling practices, about characteristics of component models, about component model development procedures, and about model composition guidelines could be provided at much less effort by a model curator than if modelers were to research these topics from scratch, or applied piecemeal, every time a new model is built.

Building a curated model classification is a milestone for any model curation practice to occur. Its usefulness would depend on aggregated contributions from the systems engineering community, both theory and practice. More research is required to establish what metadata is necessary to structure such a classification and permit queries to be performed on it. Furthermore, model compositions performed today are mostly ad hoc. More work on model composition guidelines, especially mathematical model and simulation composition is needed.

REFERENCES