## Title
A Framework to Structure Agent-Based Modeling Data for Social-Ecological Systems



Journal
Structure and Dynamics, 4(1)

Authors
Altaweel, Mark R
Alessa, Lilian N.
Kliskey, Andrew
et al.

Publication Date
2010-01-07

DOI
10.5070/SD941003307

## INTRODUCTION

Despite the growing popularity of applying agent-based models (ABMs), including the coupling of ABMs with other modeling techniques, only recently have there been efforts to standardize critical modeling aspects such as model descriptions, calibration of parameters, and verification (Grimm et al. 2006, Kettering et al. 2006, Richiardi et al. 2006). The lack of rigor in constructing and applying ABMs has likely led to some reluctance by researchers to implement these models for understanding social-ecological systems (SES). Formalization and standardization techniques to make models more manageable, distributable, testable, comparable with other models, and understandable to the general research community are a welcome development that may help to increase interest in ABMs even by those who have shown past skepticism (Gilbert and Bankes 2002, Goldspink 2002, Hales et al. 2003, Richiardi et al. 2006).

Two common ABM protocols in use for SES are Overview, Design Concepts, Details (ODD, Grimm et al. 2006) and MR POTATOHEAD (2010, Parker et al. In Press). The ODD protocol provides a general structure and framework for presenting model details, including their mathematical representation. MR POTATOHEAD, on the other hand, allows researchers to apply a template that enables the design and communication of a model for investigating land use questions. The development of such protocols, ontological structures, and standardization methods for developing models, however, may not alone be sufficient to facilitate the application of ABMs for SES. Projects also need to focus on the critical steps that are relevant for delineating data needs so that proper information can be collected in a systematic format that ultimately aids in the creation of desired models. This is potentially useful for those who are interested in ABMs but might be unsure as to what data are relevant for constructing models and for communicating the data collection process to the larger research community.

In some SES modeling, data derive from social surveys, ethnographic fieldwork, or other empirical observations that have been carried out prior to model development. In other SES cases, multiple researchers might be involved, but only part of the team is proficient in modeling and those proficient in modeling often do not fully understand the domain sciences involved in the effort. For both of these instances, data for constructing ABMs may become problematic to obtain, as certain data qualities and understanding of social-ecological interactions might be missing. Collected data often lack such characteristics as time scales, relevant behavior properties, pertinent entities involved in behavior, and descriptions properly expressing behavioral actions. In addition, collected information often does not articulate clear feedback relationships between social and ecological components. This lack of information often results in the need to recollect data or use proxy values to compensate for missing records (Watson et al. 2005, Verbug et al. 2009).

To minimize such data problems, data collection efforts should be coordinated with model development, as information from modeling may direct data collection just as data collection could inform model development. A structured method that organizes data collection to be directly applicable for modeling projects – and that can be adaptable as new information is received – enables scientists to focus efforts so that data collection is more efficient, creates a clear record of information, conserves project resources (i.e., time and money), and potentially facilitates model creation. This paper presents an attempt at constructing a standardized method to collect and organize data through an approach we call Delineate, Structure, and Gather (DSG). We first describe social-ecological systems and the problems associated with modeling such systems. We then describe how data collection can be organized and guided by DSG. We next present our case study with which we illustrate our approach, demonstrating an ongoing project in which current fieldwork applies DSG. In the discussion, we argue for the relevance of our methodology in the study of SES, including how DSG fits within a modeling cycle that addresses the stages required for developing models. We also provide suggestions on steps that can be taken to further develop DSG.

## PROBLEMS IN MODELING SOCIAL-ECOLOGICAL SYSTEMS

We define SES as having a system-based perspective that looks at the behaviors and interactions between social (e.g., economic, cultural practice) and ecological (hydrologic cycle, weather, soil characteristics, etc.) components (Berkes and Folke 2000). Scientists increasingly apply SES perspectives in their work in order to better understand the interplay between ecosystems and the societies that exist within them (Holling 2001). Because SES often involve multiple interactions between social and ecological components, with these interactions being nonlinear and difficult to predict, SES can be described as complex systems (Gunderson and Holling 2002). Understanding complex systems, specifically how these systems change, often requires a modeling approach that can be used to experiment with, and define, the structure of interactions that lead to different states in SES.

Scientists often apply ABMs, perhaps along with other modeling techniques, in their attempts to understand SES (Janssen and Ostrom 2006). However, constructing such modeling approaches could be problematic for researchers as data required for bottom-up behavior have not been traditionally collected and might be relatively slower to obtain than values for top-down techniques (McAllister et al. 2005, North and Macal 2007). In many SES projects, interdisciplinary teams of scientists collaborate with not only those from different disciplines but experts in computer modeling, who are often asked to integrate disparate data into a modeling framework. In many cases, this creates problems of communication, as many researchers use different terminology in expressing aspects of their discipline and modelers may not fully understand the data needs of the problems addressed.

Such data collection and communication problems are not unique to SES studies. Computer science and other fields have long established protocols and standardization practices that facilitate communication between project components or the larger community in order to make concepts more clear between all those involved (Holzman 1991, Benjamin et al. 1995, Odell et al. 2001). For SES, because computationally technical and non-technical individuals are often involved, protocols that communicate data structure and needs between these researchers should be understandable to all those involved.

Data collection for developing models has a greater potential to benefit projects if methodologies are developed to facilitate and make information collection more organized and efficient for technical and non-technical individuals. Therefore, our intent is that DSG tightly integrates and enhances communication between data collection efforts and model development, particularly if projects have already determined to use ABMs. This approach not only attempts to assist communication between project members, but it allows efforts to communicate to the broader research community on how data are collected and used to develop models.

## METHODS

The primary goals of DSG are to guide initial data collection, be adaptable to new information as it is observed either through modeling (i.e., initiated through validation) or empirical observation, and inform how models can be created (Figure 1). DSG, in other words, is continuously applied within a modeling cycle that requires reassessments of initial assumptions, conceptual models, and data (Thulke et al. 1999, Grimm and Railsback 2005).

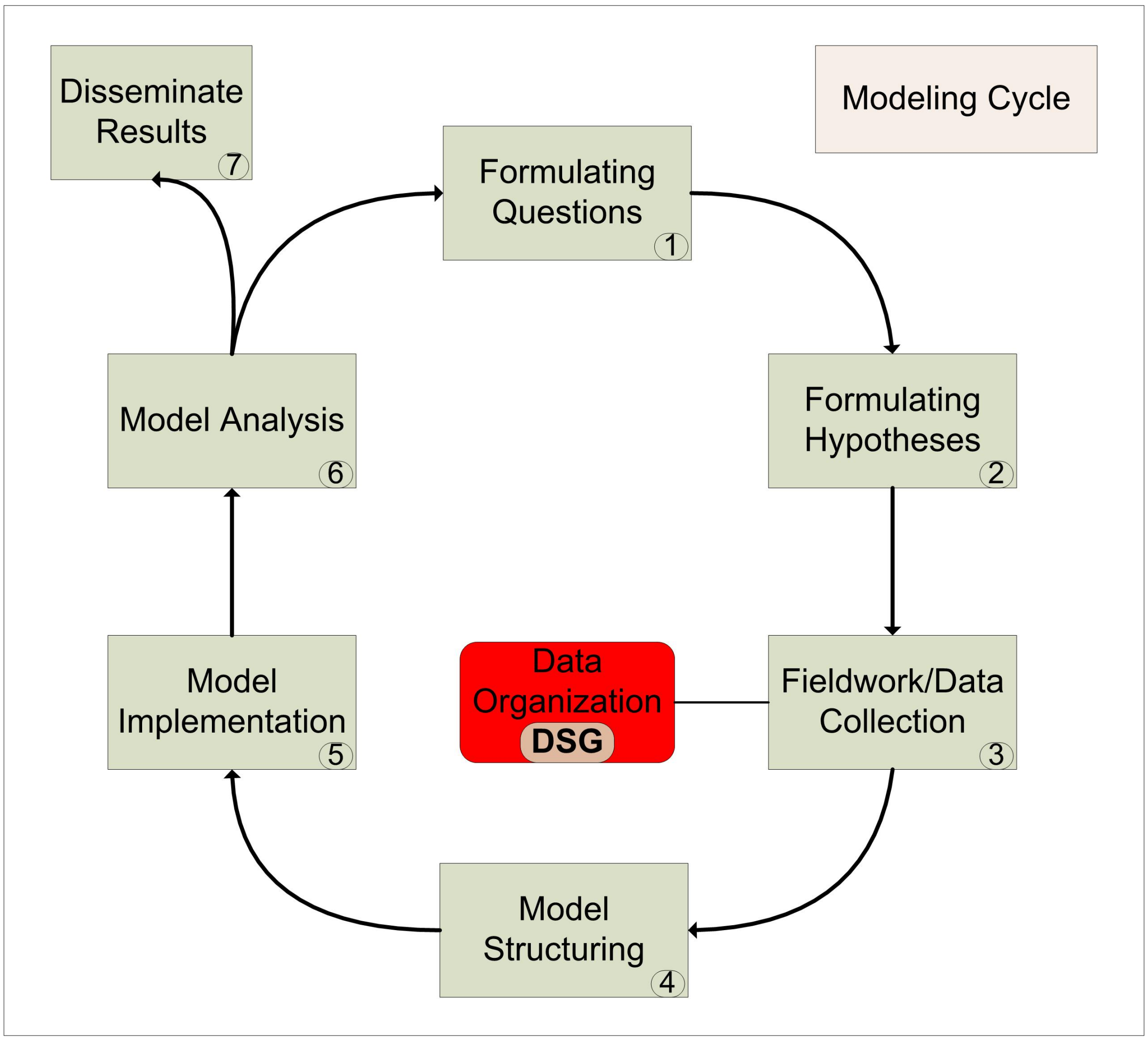


**Fig. 1**. Schema showing how DSG applies to a modeling cycle similar to that described by Grimm and Railsback (2005). DSG is applied within the "Fieldwork/Data Collection" step that incorporates data organization.

In our view, after questions and hypotheses have been developed (i.e., #1 and #2 in Figure 1), DSG becomes applicable within a fieldwork and data collection step (#3), as it informs and structures efforts. In the next step (#4), data gathered and formally organized can directly shape model structuring. The next consecutive steps are model implementation (#5), analysis (i.e., verification and validation; #6), and dissemination of potentially useful results (#7). Within this cycle, DSG is reassessed as needed based on model analysis and observations carried out, which may first force a reexamination of the initial questions and hypotheses (#1 and #2). Even if DSG is reassessed, the structure and categories within DSG remain the same. At the dissemination stage (#7), DSG should help in presenting the results to the larger research community, as it describes data collection procedures and the information used for model development.

Based on the above description, we advocate that DSG fits within a pattern-oriented modeling (POM) approach that details behavioral patterns and interactions (Grimm et al. 2005). This paradigm allows the identification of relevant behaviors, required parameters, appropriate spatial characteristics, and applicable temporal scales. These aspects form behavioral patterns that are expressed through collected data and become evident in the application of models. In addition, behavioral interactions that enable system feedbacks to be developed are indicated through collected data and linked behaviors. The following discussion demonstrate how DSG can contribute to model development and the POM approach.

**DSG Details**

Delineate (D), Structure (S), and Gather (G) is a framework that refers to a data collection method for determining data needs, organizing how those data needs fit into relevant categories, and presents how and what data are collected. These three elements are all organized within a general schema displaying the makeup of the framework (Figure 2). Data needs are delineated (i.e., D) within categories (e.g., characteristics) that are organized under a structure (i.e., S) that defines a given behavior, including its relationship to other behaviors. Gather refers to the methods applied to collect data as well as the actual information used to represent behavior in Structure. We recognize that a variety of data collection techniques can be used to create an ABM; our discussion focuses on defining data that need to be collected and its organization, rather than on the specific techniques (e.g., social survey) or questions used with research subjects. In other words, DSG does not prescribe data collection techniques, but it accommodates approaches that apply multiple methods for data gathering.

Figure 2 provides a general guide for the category descriptions that will be discussed below, showing how the three elements of DSG are applied. In addition, Appendix 1 lists the relevant categories and presents their questions. This offers an example template that can be used to collect and organize data. The categories provide data collectors with a guide for the types of information that need to be addressed during data collection, including predetermined questions that guide such efforts. Other types of questions that specifically address subjects or data sources are created as needed so that required information defined in categories are targeted. In other words, data collection populates information requirements determined in categories. Later, researchers may want to revisit certain data categories represented in DSG in order to address new research questions that may emerge (see Figure 1). After a description of the approach is provided, we will present a case study showing how our approach is applied in data recovery (Appendix 2).

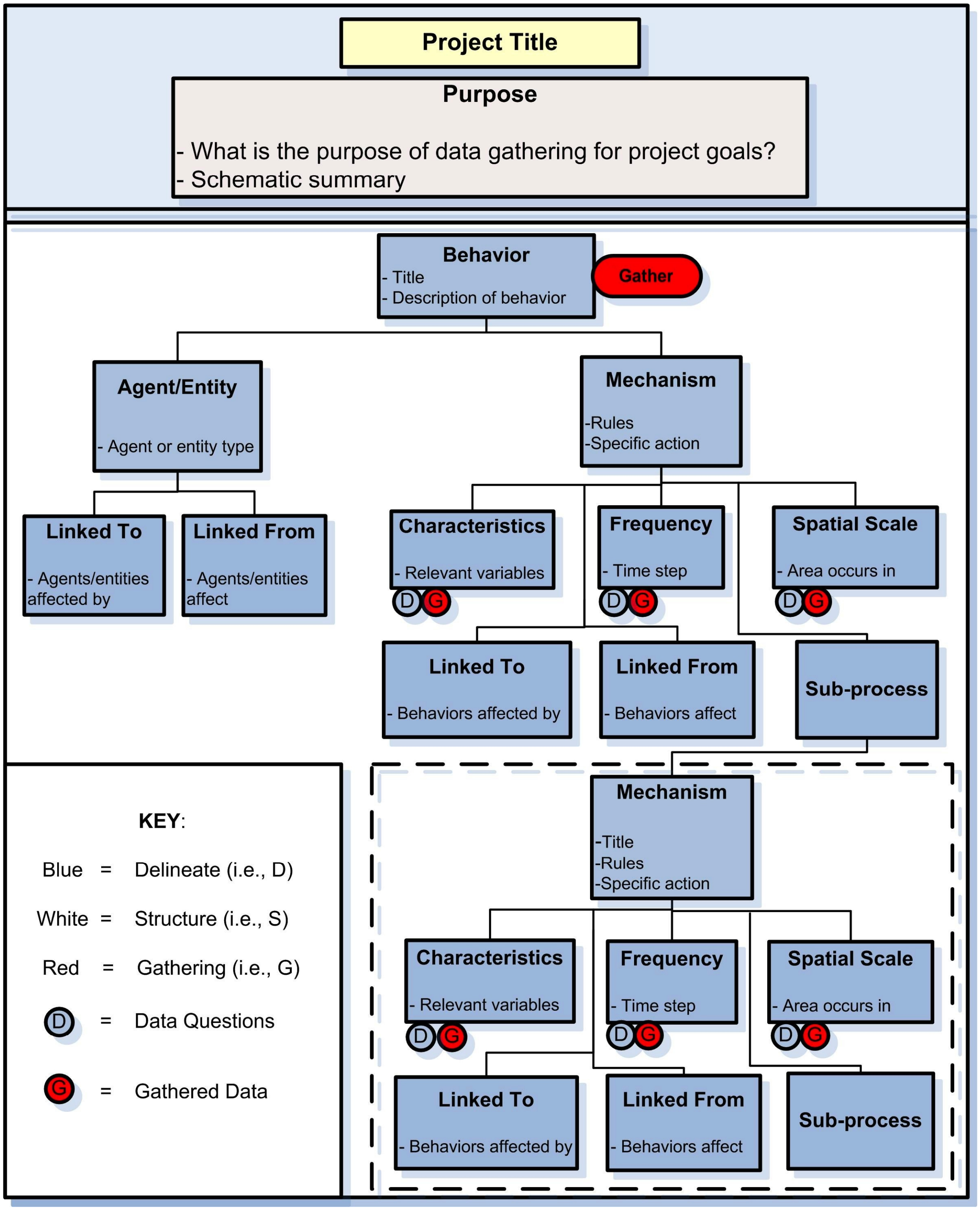


**Fig. 2**. Layout of DSG showing major categories and subcategories. In the schema, categories contain abbreviated information, which are described in detail in the text. The dashed lines indicate optional subcategories that are added if a Sub-process is defined.

*Purpose and Behavior*

After a general title (i.e., Project Title in Figure 2) is provided, the first category to be addressed is the Purpose category. Researchers indicate a description of the data collection's intent, and in this category they are specifically asked: What is the purpose of data gathering? In other words, researchers are asked how the data collection effort helps achieve project goals. To simplify the organization and future construction of models, researchers may want to include a simple schematic drawing to summarize relevant behaviors addressed. The schema may show behavioral interactions and relationships, allowing researchers to express conceptual models and behavior patterns that relate to the main research goal.

After the Purpose category, the Behavior category is addressed. Behavior and its subcategories are repeatable depending on the number of relevant behaviors studied. Behavior is a top-level category, which is then divided into two primary branches entitled Agent/Entity and Mechanism (Figure 2). Behavior refers to specific action(s) that the study is interested in, i.e., a social and/or environmental function addressed in the SES. The category can represent the entire research goal or form part of a larger set of behaviors that may interact with each other. Behaviors can initially be very basic, but as a project develops researchers may want to revise this category to define more specific sub-behaviors (i.e., see Sub-process). Examples of this are shown in Appendix 2 (i.e., Behaviors 1a & 5a). In this category, the researcher should provide a short title for the behavior, used to later organize and reference other described behaviors, and a general descriptive summary of what the behavior represents. Specific data qualities are not necessary in this level of the category, and the description should be kept relatively short. Behavior categories should be numbered according to their sequential location in a set of Behaviors (e.g., see Appendix 2). This aids in making the categories and data quickly understandable to analysts and others.

*Agent/Entity*

For the Agent/Entity branch, we define an agent as an autonomous unit (e.g., person, insect, etc.) that has the ability to act and make choices based on local knowledge (Bonabeau 2002), while entity refers to a physical unit (e.g., grass or tree) that does not have the ability to make decisions when the local environment or surroundings change. Agents and entities are the physical objects that receive and implement action for a given behavior. In other words, this category directly links agents/entities with specific Behavior categories. In the Agent/Entity level, the researcher is asked: What is the primary agent/entity instigator for the behavior being investigated? The researcher can simply record a general agent/entity category (e.g., people) that reflects the actor involved in the action described in the Behavior category. The Characteristics category should allow the differentiation of different subtypes (e.g., young and old individuals) of an agent or entity; however, descriptions defining these subtypes can also be placed in this category. This category, similar to others, is repeatable based on the different types of agents or entities relevant for a Behavior category.

Linked To/Linked From

Agents and entities can interact with other types of agents or entities. This leads to the subcategories Linked To and Linked From. The main question in the Linked To category is: What agents/entities are affected by actions of the agent/entity defined? For Linked From, the question is: What agents/entities affect the agent/entity defined? In other words, Linked To and Linked From allow a network of agent/entity relationships to be developed, which can be used to determine how agents/entities are connected and the interactions with each other. If specific data on network relationships exist, then this detailed information could be depicted here.

*Mechanism*

The Behavior category splits into another category called Mechanism. In this category, the researcher is asked: Describe the rules (e.g., if a person is hungry then he or she seeks food) or specific actions (e.g., rate of water collection) that makeup the behavior? The rules in the Mechanism category apply to factors that cause an action to occur, such as conditions or circumstances that make the behavior possible. Specific actions refer to behaviors that occur based on rates, probability, or mechanisms that are not necessarily clear. For this category, the researcher should provide details on either the rules or actions relevant for this category. This could include descriptions, specific data, schematic diagrams, and other details that help define a behavior.

Characteristics

The Mechanism subcategory has the following subcategories associated with it: Characteristics, Frequency, Spatial Scale, Linked To, Linked From, and Sub-process. The Characteristics category defines the states or qualities (e.g., age of individuals, seasonal river discharge values, probability of opinion based on social status) of factors that are relevant to a behavior's functionality; these data should be provided in descriptive or numeric form (see Appendix 2). The question asked by this category is: What are the characteristics that make up the behavior and what are their values? Characteristics can become the state variables that evolve or change in a model, but this is determined later during model construction. At this stage, all relevant behavioral characteristics should be collected and described.

Frequency

This category refers to the temporal qualities of data collection and behavior. The first question asked is: Over what period are the data collected? The next two questions deal with temporal qualities in the behavior's occurrence. Specifically, does the behavior occur at consistent time intervals? If the answer is yes, then how often does that behavior occur? The last question does not need to be answered, as many behaviors could be better described as discrete events, or behaviors that are triggered by or conditional upon other behaviors occurring, rather than occurring in consistent time intervals. However, even discrete events could be organized into intervals if such behaviors appear to occur at consistent rates. For such cases, the researcher should collect data over a defined period (e.g., rainfall over a month). In this category, the first question assists the researcher to reference the period in which the data are collected. Organizing and referencing data within a period of collection becomes useful at the model construction phase, as this provides temporal resolution for the studied behavior. If certain behavior qualities have varying temporal resolution, the researcher may want to use the Sub-process category to describe action(s) that occurs at different time scales.

Spatial Scale

For this category, the first relevant question is: Over what area does the behavior occur in (e.g., a watershed, an island)? The second question states: What are the relevant spatial data? This second question refers to specific data that represent the spatial bounds that a behavior affects (e.g., landforms represented in shapefiles). Even though many behaviors have a spatial dimension, some behaviors may not be spatially oriented; therefore, the researcher could state that this category is irrelevant. If there are multiple spatial scales relevant to different aspects of the behavior described, the researcher should consider placing some elements of the behavior within the Sub-process category. This allows the researcher to distinguish what particular spatial components are affected by actions within a behavior.

Linked To/Linked From

Linked To asks the question: What behavioral categories are affected by the behavior being described? Linked From asks: What behaviors affect the behavior being described? The Linked To and Linked From categories, in a similar manner to Linked To/Linked From in Agent/Entity, enable the linking of Behavior categories and feedbacks between behaviors to be distinguished, an aspect critical for POM and SES. Both Linked To and Linked From connect behaviors at the same behavior level. Behaviors 1 and 2, for instance, are at an equivalent behavioral level, while Behaviors 1 and 1a are not (Appendix 2). Therefore, connections between a Sub-process category and a higher-level Behavior category are not included here.

Sub-process

This category refers to a behavior that can be broken into ancillary parts. The question this category asks is: Are there any relevant sub-behaviors affecting this behavior and what are they? If there are sub-behaviors, then starting at the Mechanism level, the categories are repeated for each identifiable Sub-process. One difference in the Mechanism category applied *within* the Sub-process category is that the title is placed in the Mechanism category, since no Behavior category is defined. Although it is possible to repeat the Behavior category in order to represent sub-behaviors, the Sub-process category is intended to include more detailed information regarding a specific behavior already defined. In other words, researchers may choose to expand their information on a behavior as more details, specifically details that can be differentiated from other facts, are learned about a behavior. As an example, some behaviors may have varying time characteristics or affect different spatial scales, as described above, and specific characteristics of behavior affecting these scales might be best described in the Sub-process category.

Data Questions

In the Characteristics, Frequency, and Spatial Scale categories, Data Questions is applied, forming part of the Delineate element in DSG (see Figure 2). In the relevant categories, Data Questions are used to distinguish specific questions researchers apply to collect information from subjects or data sources. We differentiate between category questions, or questions that stay constant within categories (i.e., the questions described for the categories discussed), and Data Questions, which are mutable to accommodate specific data collection efforts. Nevertheless, the encoded category questions in some instances could directly address all relevant information needed, negating the need for creating further questions. Data may also simultaneously address the fixed questions in categories and Data Questions. Examples of Data Questions are provided in Appendix 2.

*Gather*

While the above categories each form a part of the Delineate element in DSG, the Gather element is used to populate data into the defined categories, with Structure representing the entire configuration. In conducting Gather, the researcher lists within the Mechanism category what type of data gathering he or she will conduct for specific areas relevant to the behavior studied. This information is placed under the heading Gather. If necessary, the researcher may specify which data collection methods affect which categories.

Directly coupling empirical fieldwork with ABM is occurring at a more rapid rate through many different approaches, with techniques applied to collect data and understand behaviors relevant for model construction (Janssen and Ostrom 2006). The Gather element, therefore, does not prescribe a specific technique to collect needed data. Traditional social surveys, ethnography, field observations, satellite data, and even role playing games have been shown to have merit in data collection and can be used for constructing models (Barreteau et al. 2001, Evans and Kelley 2004, Castella et al. 2005, Agar 2005, Altaweel et al. 2009). The Gather element is, however, particularly important in documenting how data are collected.

Gathered Data

Specific data that have been collected should be placed within the Gathered Data category. This category is associated with Characteristics, Frequency, and Spatial Scale, mirroring Data Questions. Data that are not easily listed, such as spatial data files, should be specifically referenced in some format within Gathered Data, allowing documentation of the relevant information for the category. Data Questions are numbered so that they match numbered data in the Gathered Data category (see Appendix 2 examples). Fixed questions in categories are listed using letters in alphabetical order so that their resulting data are easily matched with questions.

# CASE STUDY: SEWARD PENINSULA WATER USE

In order to demonstrate how DSG can be applied to SES, a case study is presented below. We recognize that one case study may not be sufficient to show that a developed methodology can address all relevant SES modeling efforts; therefore, the intention is that our attempt starts a broader dialogue with other SES modeling efforts in order to further develop DSG. The case study addresses water use in village societies of Alaska that are undergoing rapid social and ecological change to their freshwater resources (Alessa et al. 2008).

In applying our data gathering methodology, several different techniques in procuring the relevant data for the case study were conducted, including using social surveys, participant observation techniques, and downloading web-based sources. For the case study, we are interested in determining how people's perceptions of freshwater availability and quality affect community council decisions that address freshwater needs. After research questions and hypotheses were initiated, a conceptual model, with defined behaviors, was formed and placed in the Purpose category. This then led to the creation of the Behavior categories, including their subcategories, with data collected to populate the required fields.

## Case Study Details

The purpose of data gathering in this case is to delineate the behaviors that relate to how water use and perceptions of water quality and quantity affect water use decisions by community councils (Appendix 2). These decisions may have major implications on how they affect local watersheds through changes to the hydrologic system that may further force additional or alternative measures to be adopted by communities. The case study derives from previous and ongoing fieldwork conducted in villages (e.g., Alessa et al. 2008). Some of the data indicated are from a single village, while others were obtained from multiple communities. Not all data needed have been processed or collected, as Appendix 2 shows that not all questions have been answered. However, enough information has been collected to allow the creation of model behaviors for water use, perceptions of water availability/quality, and how community councils make decisions regarding water issues (Altaweel et al. 2009, Altaweel et al. 2010a). In previously developed models, both qualitative and quantitative data gathered were used to develop models as well as applied for verification and validation.

The purpose of Appendix 2 is not to show a fully completed project; rather, it is an example of sufficient data gathered that allows for model development of SES. Demonstrating how that data are organized and gathered using DSG is the intent of Appendix 2, even though further iterations of the modeling cycle are still needed. We urge readers to investigate the modeling papers cited in order to see the models that have been developed from what is presented in Appendix 2. Since this application of DSG is fairly lengthy, the descriptions, questions, and answers to questions, including the data gathered and gathering strategy applied, are listed in Appendix 2. The DSG template from Appendix 1 was initially used, with Appendix 2 reflecting the outcome of organizing data needs and collection.

In Appendix 2, there are five Behavior categories, which are: Select Water Source (1), Water Quantity/Quality Perceptions of Non-municipal Systems (2), Make Decision on Water System (3), Implement Decision on Water System (4), and Hydrologic Process (5). Two of the behaviors, Select Water Source and Hydrologic Process, had one Sub-process each (i.e., River Travel and Weather System respectively). The five main behaviors were identified prior to fieldwork, since they are generally understood. Behavior 5 represents an environmental dynamic, which interacts and directly affects Behavior 2 that deals with social perceptions of hydrologic system evolution. Behavior 5 not only incorporates hydrologic functions, but weather conditions have a direct effect on how this system evolves on a day-to-day basis. The other steps are primarily social.

**Data Collection to Modeling**

In creating the models cited (Altaweel 2009, Altaweel 2010a), DSG enabled the delineation of important details needed for most ABMs. After the behaviors focused on are identified (Behaviors 1-5), the next step is to identify the agents/entities involved. This includes individuals and hydrologic entities. Next, the Mechanism category provides the specific rule sets and descriptions that become represented in model code (e.g., conditional behavior rules become conditional if-then statements). The Characteristics category is critical for providing the specific data that instantiate models, which become model parameters and states such as agent types (i.e., α-, β-, γ-types) and distances traveled along rivers. The researcher should also decide at this stage which Characteristics are relevant to modeling goals; therefore, not all collected data are necessarily incorporated into model development. The Frequency and Spatial Scale categories provide the spatiotemporal resolution of behavior that is critical for simulations, specifically the time steps and place of actions in models. The Linked To/Linked From categories express how behaviors affect each other, which allows feedbacks in models to be developed. Finally, the Sub-process categories were developed in the case study only after the basic behaviors were expressed and a need for greater detail (e.g., travel along a river) affecting the primary behavior became apparent. These sub-behaviors, in essence, have either positive or negative relationships that shape the outcome of primary behaviors (e.g., there is a strong positive relationship between selecting a water source and its distance from a community). In summary, specific actors, spatiotemporal scales, state variables, functions, rates, and behavior interactions are identified in DSG, which all are critical characteristics of most ABMs. Many of these data or descriptions can be applied directly in models; data collected show significant behavior influences on social or ecological outcomes and allow the development of algorithms based on data relationships (e.g., how agent types affect decisions in Behavior 3).

Despite this mapping of DSG to models, we recognize that relevant behaviors may not be known or well-understood prior to data collection. Furthermore, not all data relevant for these behaviors can be collected. This is why we emphasize that DSG can be applied as an iterative process, with simple and/or generic behaviors initially identified that are then populated with more detailed information as needed based on relevant actions that are discernible through empirical observations or modeling. An example in the case study provided is the River Travel Sub-process. Initially, data collection teams were not aware of the various places that people go for collecting water; however, such actions became clear after data collection commenced. This sub-behavior, therefore, is included since it guides decisions on which water source people use along rivers. Behaviors 1-2 and 5 are modeled in Altaweel et al. (2009), while Behavior 3 is modeled in another publication (Altaweel et al. 2010a). Behaviors 1-3 apply ABMs, while Behavior 5 uses a Markov chain approach to modeling hydrologic functions. Behavior 4 is a focus area of ongoing research.

# DISCUSSION

### Significance of DSG

In the coming years, one may expect that empirical validation and application of ABMs to real-world scenarios will be enhanced (Boero and Squazzoni 2005, North and Macal 2007, Windrum et al. 2007). This is not only a welcome development, but in order to facilitate this development structured data gathering needs to be developed. Researchers have been calling for model development to be more focused on identifying relevant behavioral patterns that allow models to be more robust and useful for forecasting purposes (Grimm et al. 2005). Critical to this effort, DSG allows researchers to organize data for modeling that identifies such patterns; the DSG methodology places information into relevant categories and develops a plan for gathering needed modeling data. Interactions between behaviors are also identified and expressed within DSG.

Simply stated, DSG facilitates and organizes the transition from research questions and hypotheses (i.e., # 1 and #2 in Figure 1) to model development (#4). It is a structure that is adjustable to new observations and data modifications that are identified within a modeling cycle. DSG, therefore, does not fit into a simple linear process for developing research, but it is modular and expandable to allow for changes and broadening of behavioral understanding as new information arises.

We believe the benefit of DSG is that it provides a framework which helps structure SES modeling data, enabling collection and modeling efforts to be more efficient, establish a clear record of gathered data, conserve project resources, and more easily create models. Efficiency is enabled through a formal structure that informs data collectors not only on what data to gather but within which categories gathered data, that directly address characteristics of behavior, can be placed. Clear and organized records are useful in particular for model verification and validation, as this facilitates comparisons between inputs, outputs, and observed values of models. In fact, if a structure such as DSG is applied and presented along with a model, then other scientists can review not only the model but also the model creation process that presents how relevant behaviors and data are identified and captured. Project resources are conserved when data gathering is efficient and focuses directly on clear data needs that enhance modeling capability. Models can be more easily constructed because collected data should directly address important modeling needs, such as spatiality, temporal resolution, and behavior rules. All of the benefits outlined are particularly useful for researchers who are beginning to develop an ABM but have difficulty knowing what data are needed for developing a model. This also facilitates communication between modeling experts and domain scientists who provide needed data, as it allows these individuals to use the same language and data structure to develop needed models.

In the introduction, we mentioned that ODD (Grimm et al. 2006) and MR POTATOHEAD (2010, Parker et al. In Press) are two model protocols that can be applied to study SES. In fact, we believe DSG can facilitate data gathering that can be used to create models that are detailed using these modeling protocols; that is, our approach augments and maps to ODD and MR POTATOHEAD. One the one hand, these two protocols describe the procedures and structures of interactions in models, while DSG details the data structure and behavioral descriptions that inform and enable model creation. The ODD protocol has a series of categories that modelers fill with necessary description and details irrespective of modeling platform or language used. MR POTATOHEAD provides six general elements that model developers use to create more detailed representations of the elements present in their model. The focus of MR POTATOHEAD is specifically on land use modeling.

By combining such modeling protocols with DSG, the formulization and pattern identification developed in a model can be represented in ODD and MR POTATOHEAD, but model data and described actions can be organized using DSG. This allows analysts and reviewers of models to not only track characteristics of models, but it allows an evaluation of the actual data used (i.e., using DSG), including how they were obtained and structured, to construct such models. In other words, prior to constructing a model using modeling protocols, DSG can be used to organize and delineate needed model data.

Assessments using DSG allow the identification of the data's purpose, spatial, temporal, and behavioral characteristics, including information relevant for agent/entity characteristics. These aspects are categories that exist in such protocols as ODD, allowing much of the collected data to directly map and facilitate the development of a model's protocol. The Linked To/Linked From categories in Behavior, on the other hand, identify behavioral interactions and feedbacks that can be expressed in models and model protocols via algorithms. In other cases, where the mapping is not direct, researchers will need to decide what aspects of data collection correspond to the modeling protocol used in addition to what data are applied in developing a model. This makes DSG a potential filter that delineates which aspects of collected data are beneficial for developing models, as not all data collected and categorized might be relevant in the model creation step. In summary, after data are placed within a DSG framework, then models can be created using a modeling protocol, enabling DSG to be compatible with other modeling protocols. Further, the examples of data indicated in Appendix 2 show how spatial, temporal, agent/entity, and behavioral characteristics identify relevant behavioral patterns and interactions advocated within the POM approach (Grimm et al. 2005). DSG, therefore, fits within the POM schema, showing how information captures behavior patterns and identified interactions are formalized.

## Further Ontological Development in ABMs

Similar to ontologies formed to communicate between computer systems and software, ontological approaches are useful for communicating concepts between people (Guarino 1998). Regardless of the frameworks used, agent-based modeling needs to undergo a stage were ontological structures are developed, tested, and ultimately agreed upon. This, we believe, addresses some of the skepticism and criticism, including the lack of methodological rigor, leveled against agent-based approaches. We do caution, however, that frameworks should not be developed too strictly with regard to how they are followed, since a strict structure may hinder needed flexibility to collect data and construct models.

We recognize that DSG may not address all data organization needs for developing an ABM, including other modeling techniques that can potentially be coupled with this approach, that addresses SES. We, therefore, invite discussion on this topic in order to further develop what we perceive to be an important step in the creation of ABMs. We believe more discussions and presentations are needed in order to determine other potentially needed characteristics for DSG. The OpenABM (Janssen et al. 2008, OpenABM 2010) Consortium is, perhaps, one such venue to facilitate this, but discussions should incorporate more groups.

### Future Developments

In addition to using DSG simply as a guide to collecting and organizing SES modeling data within fieldwork efforts, we also anticipate applying DSG within developed software, similar to MR POTATOHEAD (2010), that will enable some form of automated or semi-automated mapping of DSG to data collection efforts. In fact, we have begun this process by applying some of the concepts in DSG to an existing data mining/information fusion tool that delineates data and behavior patterns for SES applications (Altaweel et al. 2010b, Altaweel et al. 2010c). The DSG framework can be useful in data mining tools because it can structure and prepare a database search using its defined categories, with results of the search presented in a format relevant for model construction. Information captured in data mining could be structured based on the type of agent/entity, temporal, spatial, and other behavioral characteristics that distinguish specific actions. The incorporation of data mining/information fusion can allow projects to create automated and semi-automated steps between the data collection (#3) and model structuring (#4) steps shown in Figure 1. This can be useful in such situations where web-based data and other continuous data feeds are used in developing or parameterizing ABMs and coupled modeling approaches. In summary, we envision that a richer integration of data mining with DSG will allow information to be prepared for the model structuring step in a more simplified, systematic, and organized manner.

## REFERENCES

**Agar M.** 2005. Agents in living color: Towards emic agent-based models. *Journal of Artificial Societies and Social Simulation* **8(1)**:4. [online] URL: http://jasss.soc.surrey.ac.uk/8/1/4.html.

**Alessa L. N., A.D. Kliskey, R. Busey, L. Hinzman, and D. White.** 2008. Freshwater vulnerabilities and resilience on the Seward Peninsula as a consequence of landscape change. *Global Environmental Change* **18**:256-270.

**Altaweel M.R., L.N. Alessa, and A.D. Kliskey.** 2009. Forecasting resilience in Arctic societies: Creating tools for assessing social-hydrological systems. *Journal of the American Water Resource Association* **45**(6):1379-1389.

**Altaweel M.R., L.N. Alessa, and A.D. Kliskey.** 2010a. Social influence and decision-making: Evaluating agent networks in village responses to change in freshwater. *Journal of Artificial Societies and Social Simulation* **13**(1):15. [online] URL: http://jasss.soc.surrey.ac.uk/13/1/15.html.

**Altaweel M., L.N. Alessa, and A.D. Kliskey.** 2010b. Visualizing situational data: Applying information fusion for detecting social-ecological events. *Social Science Computer Review* DOI: 10.1177/0894439309360837.

**Altaweel M., L.N. Alessa, A.D. Kliskey, and C. E. Bone**. 2010c. Monitoring land use: Capturing change through an information fusion approach. *Sustainability* **2**(5):1182-1203.

**Barreteau O., F. Bousquet, and J. M. Attonaty.** 2001. Role-playing games for opening the black box of multi-agent systems: Method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation* **4**(2):5. [online] URL: http://www.soc.surrey.ac.uk/JASSS/4/2/5.html.

**Benjamin P., C. Menzel, and R.J. Mayer.** 1995. Towards a method for acquiring CIM ontologies. *International Journal of Computer Integrated Manufacturing* **8**(3):225-234.

**Berkes F. and C. Folke.** 2000. Linking social and ecological systems for resilience and sustainability. In *Linking social and ecological systems: Management practices and social mechanisms for building resilience*, F. Berkes and C. Folk eds., Cambridge University Press, Cambridge, UK, pp. 1-26.

**Bjerklie D., D. Moller, L.C. Smith, and L.S. Dingman.** 2005. Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology* **309**(1-4):191-209.

**Bjerklie D.** 2007. Estimating the bankfull velocity and discharge for rivers using remotely sensed river morphology information. *Journal of Hydrology* **341**(3-4):144-155.

**Boero R. and F. Squazzoni.** 2005. Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *Journal of Artificial Societies and Social Simulation* **8**(4):6. [online] URL: http://jasss.soc.surrey.ac.uk/8/4/6.html.

**Bonabeau E.** 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* USA **99**(3):7280-7287.

**Castella J. C., T. N. Trung, and S. Boissau.** 2005. Participatory simulation of land-use changes in the northern mountains of Vietnam: The combined use of an agent-based model, a role-playing game, and a geographic information system. *Ecology and Society* **10**(1):27. [online] URL: http://www.ecologyandsociety.org/vol10/iss1/art27/.

**DigitalGlobe.** 2010. *Imagery and information company*. [online] URL: http://www.digitalglobe.com/.

**Evans T. P. and H. Kelley.** 2004. Multi-scale analysis of a household level agent-based model of landcover change. *Journal of Environmental Management* **72**(1–2):57–72.

**Gilbert, N. and S. Bankes.** 2002. Platforms and methods for agent-based modeling. *Proceedings of the National Academy of Sciences USA* **99**(3):7197-7198. [online] URL: http://www.pnas.org/content/99/suppl.3/7197.full.

**Goldspink C.** 2002. Methodological implication of complex systems approaches to sociality: Simulation as a foundation for knowledge. *Journal of Artificial Societies and Social Simulation* **5**(1):3 [online] URL: http://jasss.soc.surrey.ac.uk/5/1/3.html.

**Grimm V. and S.F. Railsback.** 2005. *Individual-based modeling and ecology*. Princeton University Press, Princeton, N.J., USA.

**Grimm V., E. Revilla, U. Berger, F. Jeltsch, W.M. Mooij, S.F. Railsback, H.H. Thulke, J. Weiner, T. Wiegand, and D.L. DeAngelis.** 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* **310**:987-991.

**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 Modeling* **198**(1-2):115-126.

**Guarino N.** 1998. *Formal ontology in information systems*. IOS Press, Amsterdam, The Netherlands.

**Gunderson L.H. and C.S. Holling.** 2002. *Panarchy: Understanding transformations in human and natural systems*. Island Press, Washington, DC, USA.

**Hales D., J. Rouchier, and B. Edmonds.** 2003. Model-to-model analysis. *Journal of Artificial Societies and Social Simulation* vol. **6**(4)5. [online] URL: http://jasss.soc.surrey.ac.uk/6/4/5.html.

**Holling C.S.** 2001. Understanding the complexity of economic, ecological, and social systems. *Ecosystems* **4**(5):390-405.

**Holzman G.J.** 1990. *Design and validation of computer protocols*. Prentice-Hall Inc., Upper Saddle River, NJ, USA.

**Janssen M. A. and E. Ostrom.** 2006. Empirically based, agent-based models. *Ecology and Society* **11**(2): 37. [online]
URL: http://www.ecologyandsociety.org/vol11/iss2/art37/.

**Janssen M.A., L.N. Alessa, M. Barton, S. Bergin, and A. Lee.** 2008. Towards a community framework for agent-based modeling. *Journal of Artificial Societies and Social Simulation* **11**(2):6. [online] URL: http://jasss.soc.surrey.ac.uk/11/2/6.html.

**Kettenring K.M., B.T. Martinez, A.M. Starfield, and W.M. Getz.** 2006. Good practices for sharing ecological models. *BioScience* **56**(1):59-64.

**McAllister R.R.J., I.J. Gordon, and C.J. Stokes.** 2005. KinModel: An agent-based model of rangeland kinship networks. In *International Congress on Modelling and Simulation (Modelling and Simulation Society of Austrailia and New Zealand, MODSIM)*, A. Zerger and R.M. Argent eds., [online]
URL: http://www.mssanz.org.au/modsim05/papers/mcallister_1.pdf, pp. 170-176

**MR POTATOHEAD.** 2010. [online]
URL:
http://mason.gmu.edu/~dparker3/MR_POTATOHEAD/MR_POTATOHEAD_OWL.htm

**North M.J. and C.M. Macal.** 2007. *Managing business complexity*: *Discovering strategic solutions with agent-based modeling and simulation*. Oxford University Press, New York, New York, USA.

**Odell J.J., H.V.D. Parunak, and B. Bauer.** 2001. Representing agent interaction protocols in UML. *Lecture Notes in Computer Science* 1957/2001:201-218.

**OpenABM.** 2010. *Open agent based modeling consortium.* [online]
URL: http://www.openabm.org/site/.

**Parker, D.C., D. G. Brown, J.G. Polhill, P.J. Deadman, and S.M Manson.** In Press. Illustrating a new 'conceptual design pattern' for agent-based models of land use via five case studies—MR POTATOHEAD framework. In *Agent-based modeling in natural resource management*, A.L. Paredes and C.H. Iglesias eds.

**Richiardi M.G., R. Leombruni, N. Saam, and M. Sonnessa.** 2006. A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation* **9**(1):15. [online] URL: http://jasss.soc.surrey.ac.uk/9/1/15.html.

**Thulke H.H., Grimm, V., M.S. Müller, C. Staubach, L. Tischendorf, C. Wissel, and F. Jeltsch.** 1999. From pattern to practice: A scaling-down strategy for spatially explicit modelling illustrated by the spread and control of rabies. *Ecological Modeling* **117**:179-202.

**United States Geological Survey (USGS).** 2010. *Available water data for the U.S.* [online] URL: http://waterdata.usgs.gov/nwis/.

**Verbug P.H., J. Steeg, A. Veldkamp, and L. Willemen.** 2009. From land cover change to land function dynamics: a major challenge to improve land characterization. *Journal of Environmental Management* **90**(3):1327-1335.

**Watson J.W., H.A. Abbass, C. Lokan, and P. Lindsay.** 2005. Software engineering for artificial life, complex systems, and agent-based distillation. *Proceedings of the 7$^{th}$ Asia-Pacific Conference on Complex Systems*. [online]
URL: http://www.complexity.org.au/ci/vol12/msid38/file.pdf.

**Western Regional Climate Center.** 2010. [online] URL: http://www.wrcc.dri.edu/.

**Windrum P., G. Fagiolo, and A. Moneta.** 2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* **10**(2):8. [online] URL: http://jasss.soc.surrey.ac.uk/10/2/8.html.

**Yang L. and N. Gilbert.** 2008. Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in Complex Systems* **11**(2):1–11. [online]
URL: http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VBS-4K606T7-3&_user=1722207&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000053990&_version=1&_urlVersion=0&_userid=1722207&md5=805c5e47a8455a7fcef7956a4d0a94ab - aff14.