DATA MANAGEMENT FOR SYSTEM DYNAMICS MODELING

Krista C. Kelly
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California

Merle McKenzie
Jet Propulsion Laboratory
Pasadena, California

ABSTRACT

Proper data management is an essential component of system dynamics modeling. The authors have developed an approach to data management, as set forth in this article.

The article first describes the modeling and data management activities from a critical path point of view. The approach to handling the data associated activities is then developed. This approach asserts the following:

(1) it is appropriate to address data related activities at each stage of the model development process, and

(2) when properly linked, a synergism exists between each model development stage and its associated data handling activity.

It is claimed that this approach, including sequenced data handling and synergism between data and modeling activities, can produce a more comprehensive and timely model.

I. INTRODUCTION

When modeling a large and complex system, examining intricate interrelationships between system components, what is the last thing that the typical analyst is interested in addressing? Data. Data and data associated activities tend to be philosophically (but erroneously) regarded as off the critical path of the modeling process.

As a product of an effort at a R&D organization to develop a system dynamics model for strategic planning, the modeling team developed an approach to data handling for system dynamics modeling. The results are reported here, including a description of the specific data handling activities implied by a critical path assessment of the modeling process. The paper commences with definitions (Section II). The four stage general model development process and the associated data handling activities are then described (Section III), followed by conclusions (Section IV).

II. DEFINITIONS

The following are definitions that will be assumed for this paper:

1. Data - quantitative information pertaining to the real system being modeled. Data can be one of two types: mental or recorded.

2. Mental Data - data which is derived from perceptions of the system held by system participants. One of the main uses of mental data is to compensate for either the lack of, or inaccuracy of, recorded data.

3. Recorded data - that data which is noted, either electronically, or through system documentation, in some form of physical medium.

4. Parameter - an element of the model whose value is derived or input exogenously.

5. Variable - an element of the model whose value is calculated as the model is executed.

6. Data availability - the degree to which some form of data exists which corresponds to a model parameter or variable.

7. Data applicability - degree to which pre-processing (aggregation, disaggregation, conversion, etc.) is required to render data suitable for use as a measure of a system parameter or variable.

II. DATA HANDLING APPROACHES AND GUIDELINES IN THE MODEL DEVELOPMENT PROCESS

Figure 1 depicts a critical path analysis representation of the end-to-end system dynamics model development process,

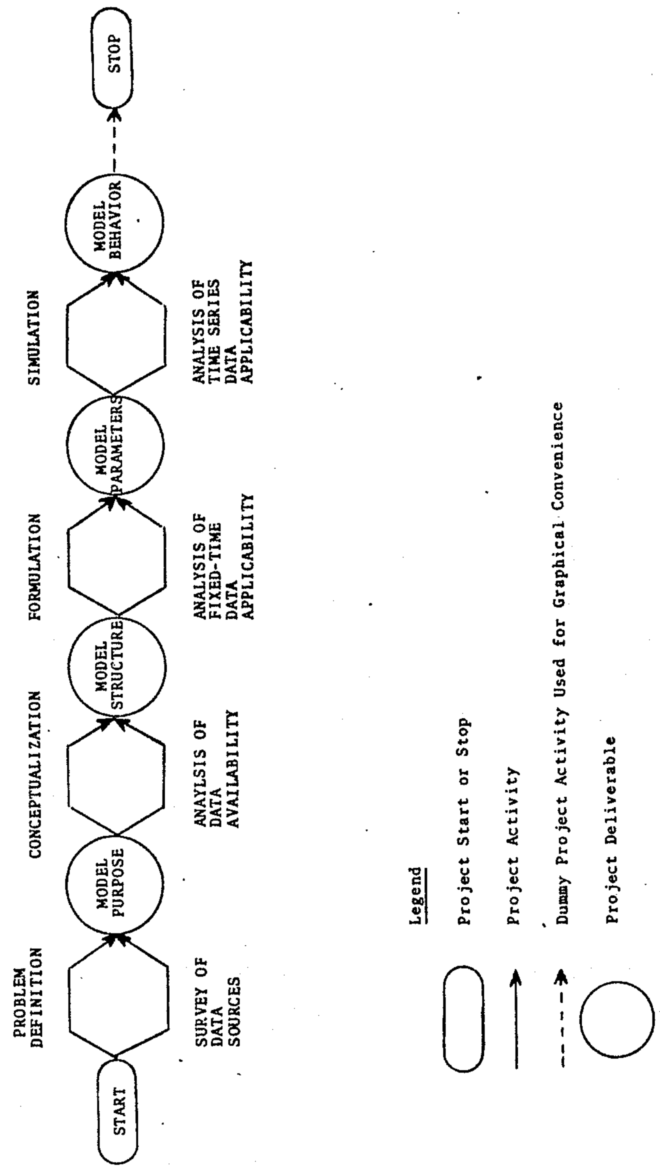

Figure 1. Critical Path Diagram of the Model Development Process Including Data Management Activities

including the proposed parallel four step data analysis process. In general, a critical path diagram is used to depict an entire project, including activities, products, and sequencing of each. As used here, Figure 1 shows the four model development products (model purpose, model structure, model parameters, and model behavior), plus the model development and data handling activities that must be accomplished for each product to be finished. The four model development activities are: problem definition, conceptualization, formulation, and simulation. The parallel data handling activities are: survey of data sources, analysis of data availability, analysis of fixed-time data applicability, and analysis of time-series data applicability. The following discussion addresses each model development stage and its concomitant data handling activity.

1. <u>Problem Definition</u>

Specification of the purpose of the model constitutes the initial model development stage. The issues of concern to the user are identified, and those variables that the user wants to measure and manage are codified through a review process.

The data handling activity which complements this initial stage of the effort is to produce a survey of potential data sources. This survey serves several purposes. Knowing what

documentation and information sources are produced, maintained, and referenced by an organization, assists the modelers in understanding the function and purpose of that organization. In the authors' case, this was of particular benefit because a baseline organizational model was being produced (Ref. 1). This survey is also useful in suggesting bounds on the model. Finally, it foreshadows future problems with data availability and thus helps to identify issues which might be more readily addressed by the model.

Potential sources of recorded data might include manual and electronic data bases, management information systems, resource and workforce reporting systems, and system documentation (organizational documentation trees, organizational roles and charters, system requirements documents, in-place system description documents, strategic and tactical program plans, technical plans, etc.).

2. <u>Conceptualization</u>

The second stage of the model development process is the conceptualization of model structure. This activity is based on extensive interviews with system participants, in which components of the system as well as relationships between components are described. In these interviews, ancillary variables or parameters are often discussed that, although not the primary metrics developed in the problem definition stage, are nonetheless quite necessary for a correct model. The ancillary measures not only affect the values of the primary system measures, but may include the ultimate leverage points of the system, or those areas to which the system behavior is sensitive.

The primary data activity corresponding to this second stage is comparing the now known model parameters and variables to the potential data sources, to determine availability of data and to determine where the data collection efforts of the next stage will be directed.

During this activity, the modeler is also checking the system structure for meaningfulness by comparing the variables and parameters included in the structure to perceptions of available data. Because the structure has been drawn from the users' understanding of the system, there should be either mental or recorded data within the system to support the postulated structure. If not, then one must pursue the question further, as follows.

If no potential sources of recorded data for a given parameter or variable exist, the advocate of the measure should

be able to supply mental data to serve to quantify the measure. If this is not possible, the advocate should refine the measure definition until it is supportable by mental or recorded data. If neither procedure is feasible, the advocate might change his mental model, and relinquish the measure. However, if it is felt that the measure is indeed a real element of the system, the advocate might take action to see that data are collected to describe the measure.

If the discontinuity persists between structure and data availability, a suggested solution is to iterate and compromise between the two. An example of how this process occurred in the authors' case is as follows. A check was made on data availability for the impact of a variable, system performance, on the generation of new system requirements. There was no data available and it was determined through further discussion that the performance variable was really a composite of two important variables, each with an independent impact on new system requirements. Thus, the structure was revised.

The check for data availability is useful in one further way: the modelers can generate a list of needed data, and keep it with them as they interview. Often a knowledgeable person will have an insight, if not into the variable and parameter values needed, then perhaps into a possible data source. As additional system parameters and variables are identified, potential data sources are added to the list.

3. Formulation

The third stage of the model development process is the formulation of equations which describe the model structure. The associated data handling activity focuses upon the analysis of data applicability for model parameters. Units and values for model parameters are developed, a process which emphasizes fixed-time data that describe the state of the system at the onset of the simulation.

Thus, data for parameters, including both exogenous inputs and quantification of feedback loops, are required in some concrete form. Jay W. Forrester advocates assigning reasonable values for most parameters, deferring intensive data gathering and analysis for those parameters to which the system is sensitive (Ref. 2). It was the experience of the working group, however, that obtaining even reasonable parameter values, using either recorded or mental data, is not a trivial task. In addition, it was found that more accurate data facilitated the ultimate approval of the model. Therefore, for both reasons, data related obstacles and solutions are discussed as follows:

Recorded Data - Obstacles

Significant obstacles to the data collection effort may exist, particularly in the case of recorded, or hard data. Data sources are not necessarily organized, and are usually not co-located. Much desirable information is not available in an automated or electronic form, or if it is, the interface may not be user friendly. In some cases, security restrictions prohibit access to sensitive data. Other obstacles may include the cost of collecting and analyzing the data to render it usable for the modeling effort.

In addition, the recorded data almost never exist in a form which may be directly mapped into model parameters. For example, information drawn from documentation (program plans, resource and schedule documents) may require aggregation or disaggregation. In his paper on parameter estimation, Alan Graham (Ref.3) describes aggregation and disaggregation at length, therefore this topic will not be further developed here.

Recorded Data - Solutions

Despite the above obstacles, recorded data are collected and used in the model formulation stage, and to support the data requirements of the model, a variety of management information tools can be applied - decision support systems, data base management systems, etc.

The administrator of such a system may be able to extract information for the model which is appropriate to the level of aggregation and the type of units of a specified parameter. In cases where data resides in flat files, specific applications software can be developed to produce the appropriate values. The authors utilized this latter approach - a small program was developed to aggregate information from flat files, utilizing sort keys already available in the data base.

If an organization has a measurable level of management tools, integration of the model with the rest of the management support system may also help the modeler in collecting data. As part of a decision support system, for example, the model data requirements could automatically be linked to the appropriate data bases. The extent to which management decides to finance such integration will also depend to some degree upon the extent of the expected future use of the model.

Mental Data - Obstacles and Solutions

To obtain mental data to quantify structural relationships, and to verify model structure, a number of people invol-

ved in the management of the system will be interviewed. This process can obviously result in inconsistent models of the same system. Thus, an important resource for obtaining mental data is a good interview technique. The modeling team found the following guidelines to be helpful in this activity:

(1) When deciding upon the units of a parameter, interview the person managing that part of the system. If a given user is not sure about a unit, follow the organizational structure charts to locate someone who can supply that mental information.

(2) When asking for a particular quantification, determine whether the respondent is weighting the response, and if so, what the impact of that weighting is on the model structure. The respondent's mental assessment may be formulated based upon a mode (most frequent occurrence), a mean (an average of all occurrences), or a temporally based (e.g., most recent occurrence) style of observation. If the response reflects a mode, it is useful to ask if there are any important exceptions to the rule. Such exceptions will quite likely lead to the discovery of another element of system structure. If the response represents a mean, it is helpful to ask for the range about the mean. The range also gives the modeler information on the possibility of extreme cases.

If there are significant exceptions to either the mean or the mode response, it may indicate a need to disaggregate the model. This is an example of the data - structure iteration process which also occurs in the second stage of model development.

(3) Another technique used for clarification and for improving accuracy is to ask a given question in a different way. Application of this technique reveals an interesting phenomenon. When a question is stated in its reciprocal or contra-positive form, often the response differs significantly from the response obtained when asked in its original form. For example, in one area of the organization, a question was asked in the form: "What is the relevant exit rate from the given level in x units per DT?" The corresponding question was also posed: "What is the average time spent in the level?" The rate supplied was not equal to the inverse of the time spent in the level. This phenomenon occurs whether dealing with one respondent, or with different respondents, and can be explained two ways. One possible explanation is that different people have different perceptions of the system. However, another possible explanation is the problem of reciprocal averages, whereby the average of reciprocals is not equal to the reciprocal of the average. If the phenomenon is caused by reciprocal averages, the solution is to use the measure most

real in the system, usually determined through review and concensus.

(4) Organizational differences in mental data will certainly occur due to the different viewpoints held by people in different organizations. These differences may be resolved via multiple reviews.

4. Simulation

The final stage is analysis of model behaviour. This step includes remaining verification tests, such as sensitivity analysis used to determine to which parameters, and to what structure, the modeled system behavior is sensitive. Stage four is characterized by the fine tuning of parameter values and structure, to determine whether the data or the structure is responsible for each effect, and whether further analysis is required.

Unlike the data required for the first three model development stages where the data is for a fixed point in time, this stage requires data as a function of time (behavioral data). The problems associated with using hard data for time-varying data analysis are manyfold those for time-fixed data. For example: within the same data set, definitions of measures can change through time; the organizing principle of the set of data can change over time; levels of aggregation of the data can change; and organizational structures and even time measures can change over time. All this change embodied in a set of data can have occurred without obvious explanation or warnings for data users.

If hard data is not available or usable for model validation this does not necessarily mean that the model has failed to reproduce the nature of the real system. As discussed above mental data can be used to perform this validation - mental data that is internalized into users' perceptions of the system. The mental data can be either of the system's past performance or of its future behavior.

Regardless of the type of data or the test viewpoint (historical or future) being used, the same basic test procedure is followed. If there is a discrepancy between model behavior and data, the manager responsible for that part of the system examines the pertinent model structure. If no structural deficiency is detected, discussions with other managers are held in which it is proposed that the model is correctly predicting system behavior. This generates both lively, and productive, discussions.

If indeed no data exists (either recorded or mental, past or future), i.e., behavioral data availability is zero, there is indeed a real problem with the model. If all the previous data handling activities were performed, it is highly unlikely that this condition will occur. If it does occur, however, the relevant portions of the proceeding three stages will need to be repeated and the model corrected.

IV. CONCLUSION

The relationships between model development stages and data handling activities as described heretofore are summarized in Table 1 below. Data handling activities are indeed on the critical path of the modeling process, and failure to accomplish them in parallel with their corresponding model development stage will impede the ability to proceed with the subsequent stage. Worse, following the natural tendency to defer data related activities may necessitate reiteration of one or more previous stages, thus delaying the project even more. Accomplishing the data handling activities synergistically with each modeling stage will, on the other hand, result in a timely and comprehensive model.

Table 1

Stage	Model Product	Model Development Activity	Data Handling Activity	Subject of Data Handling Activity
1.	Model Purpose	Problem Definition	Perform Survey of Potential Data Sources	Data Sources
2.	Model Structure	Conceptuali-ation	Assess Data Availability	Parameters & Variables
3.	Model Parameters	Formulation	Assess Data Applicability	Parameters
4.	Model Behavior	Simulation	Assess Data Applicability	Variables

REFERENCES

1. Merle McKenzie, "Systems Dynamics Modeling for Long Range Strategic Planning," paper submitted to 1983 International System Dynamics Conference, (M.I.T., July 27-29, 1983), p. 10.

2. Jay W. Forrester, Industrial Dynamics (Cambridge, Massachusetts: The M.I.T. Press, 1961), p. 171.

3. Alan K. Graham, "Parametric Estimation in System Dynamics Modeling," Jorgen Randers, ed., Elements of the System Dynamics Method, (Cambridge, Massachusets: The M.I.T. Press, 1980), pp 143-161.