



Massachusetts Institute of Technology
Engineering Systems Division

ESD Working Paper Series

DRAFT Model Brokerage: Concepts & A Proposal

David Hartzband
Research Affiliate
Sociotechnical Systems Research Center
Massachusetts Institute of Technology
Email: dhartz@mit.edu

Subject: DRAFT Model Brokerage: Concepts & a Proposal¹

1. Definitions²

model, n. 1. A **simplified** description, especially a **mathematical** one, of a system or process, to **assist calculations** and **predictions**.
(It. modell(o), dimin. Of modo = mode)

brokerage, from broker, n. 1. an agent who buys or sells for a principal (OF, brokeor)

This is not a compositional definition.

model brokerage – the act of providing appropriate services, generally through an agent, to allow the utilization of models

2. Rationale & Introduction

Brokerage is a simple concept that is difficult to implement in any context. A mundane example will serve to show the elements of brokerage. I want to invest a modest (but important to me) amount of money, so I go to an investment manager. I expect that this person will use her expertise to:

- Locate an appropriate set of investments for me
- Provide information to me about these investments & any alternatives
- Negotiate with me about what investments to actually make, &
- Carry out a set of transactions to make the investments.

If these actions are generalized, brokerage would consist of a set of services:

- Locational
- Informational
- Dialog & negotiation
- Access
- Transactional.

Each of these services has to be projected into a context where model brokerage is provided. The building & utilization of models is ubiquitous in both science & business today. These models can range from:

- Arithmetic &/or statistical representations of financial data, to
- Finite element representation of structural behavior, to

¹ © David Hartzband, 2015

² Oxford Dictionaries, www.oxforddictionaries.com/us/definitons/american_english

- Complex mathematical representations of dynamic processes, to
- Event-based models of organizational behavior, to
- Empirical characterization of ultra-large scale data sets in healthcare & a wide variety of other segments.

Effective & efficient utilization of models is essential for facilitating high productivity in many different types of organizations. If models could be shared across technical, temporal & organizational boundaries, much higher productivity could be realized. This would also be true if models could be combined in effective ways.

3. Technical Aspects

There are two different kinds of information that will need to be addressed if model brokerage is to be provided. These are the metainformation of the specific models & the ontologies of those models. Metainformation is that information about a model that locates & specifies the form that the model is expressed in. It is similar in type, but not in content, to the metadata of a database. The ontology of the model is information about “the nature of the model’s existence”³. This ontology specifies the content & semantics of the model in a form that is concise & able to be reasoned about. Each of these types of information & the operations on the models associated with them will be discussed.

3.1 Model Metainformation

Model metainformation is mainly associated with the locational & access service aspects of model brokerage. Metainformation is generally connected with a repository of some sort. The repository serves to store & provide access to the metainformation of multiple models. As such, it is usually sharable among a group of model users (people &/or programs). The metainformation that is essential for model brokerage is the information that allows a user or program to locate a specific model, as well as the technical description of the form of the model. This information may consist of such things as the language or application that the model is written in, definitions of general object or entity definitions used in the model, user defined relationships with other models or applications etc. This information is utilized by users to locate, model & perform appropriate management (backup, restore, save, copy etc.) or functional operations (launch, edit, etc.).

3.2 Model Ontology

³ T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993.

Model ontology is associated with the formation, dialog and negotiation services of model brokerage. Model ontology has primarily been defined by the development of languages to specify relationships & semantics associated with models (*c.f.* Neches, et al., 1991). Such efforts are rooted in a specific logical representation of semantics & may not be rich enough to express an ontology in a way that is meaningful for model brokerage. A result from model theory & an analogy with previous work in multivariate analysis may provide a concise & effective way of expressing ontology.

Hartzband (1978, 1990) proposes that a multivariate probability density function can be projected into a theorem space that allows a small number of linear combinations, resulting from a linear transformation of the function's covariance matrix, to represent the majority of the information in the original distribution. There appears to be an analogy with this result in model theory that would allow a model to be represented by a (potentially) small number of well-formed formulas that would make up the ontology of the model.

3.2.1 Model Theory Assumptions

Given a formal system, F where:

1. $F = \langle A, S, P, R \rangle$, where:
 - 1.1. A is a nonempty subset of symbols the elements of which are the alphabet of F
 - 1.2. $M(A)$ is the nonempty set, the elements of which are all the finite sequences of the elements of A
 - 1.3. S is the nonempty subset of $M(A)$, the elements of which are all the well-formed sequences of $M(A)$
 - 1.4. P is a subset of the elements of S which are either self-evident or immediately inferred, &
 - 1.5. R is the set of finitary relations over S
 - 1.5.1. $R \equiv ((x, y) \mid xRy), xRx, xRy \rightarrow yRx, xRy \wedge yRx \rightarrow xRz$
2. W is a set of WFFs ($W \subset S$) in F
3. An interpretation I in which all elements of W are true is a model of F

Proposition 1. The nonredundant axioms of F which are elements of I constitute the full ontology of F

3.2.2 Multivariate Analysis Assumptions

Given:

1. X is an n by m matrix of real valued random functions x_{ij} where x_{ij} is defined for all u in some set of real numbers, then X is an expression of a multivariate probability density function

2. The Cramér-Wold Theorem:
 - 2.1. $P_X = t'x_1 + tx_2 + \dots$ where tx is a unidimensional linear combination &
 - 2.2. $t \in \mathcal{RP}$
3. $V_{x_{ij}}$, the covariance matrix of X , &
4. $Y(i)$, the principal component transformation of X

Then the principal component analysis of X will represent the mpdf as a set of linear combinations of decreasing variance as an eigenvector product. Since, in most cases, the PCA produces a small number of linear combinations that account for the majority of the variance in X ; often only two or three linear combinations can adequately represent the mpdf (in the sense of the Cramér-Wold Theorem). Thus it is possible to represent a multivariate probability density function in a small set (2 or 3) of linear combinations.

Proposition 2. In analogy with the result described in multivariate probability theory, some small subset of $S \in I$ will provide an adequate & consistent ontology of F .

3.3 Pragmatic Satisfiability (c.f. Suchenek: 1985,1986)

Given:

1. $W \subset S$ in F
2. $I \subset W$
3. $\exists(Ia, Ib \dots In \subset W)$
4. $W \vdash Ia, W \vdash Ib, \dots W \vdash In$
5. implies
6. $F \models Ia, F \models Ib, F \models In$

If each I_n is a subset of W , & any ambiguities or contradictions are partitioned into separate I_n 's, then each I_n is a true interpretation of F .

3.4 Criteria for Completeness of Interpretations

Given:

1. $M \wedge N \rightarrow M \equiv N \leftrightarrow I(M) = I(N)$ structure equivalence
2. I is complete $\leftrightarrow I \equiv F \vdash I$ for some F model completeness
3. $\forall Ia \dots \forall In \varphi \rightarrow \forall J, J \subset I, F \models J$ substructures as models

That is, if I is model-complete in F & there is some set of WFFs, J , that is a substructure of I , then J is a model of F . Robinson has shown that every theory T (& by extension every interpretation I) is model-complete if whenever M & N are models of T & $M \subset N$ where $M < N$.

4. Summary & Discussion

Given:

1. $S \subset M(A) \subset F$
2. $P \subset S$
3. $\exists(Ia, Ib, \dots In) \subset S$
4. $\exists(a1, a2, \dots an) \in Ix \in P$
5. $F \models Ta, F \models Ib, \dots F \models In$
6. $\mathcal{O} \equiv A, (Ia \in P) \cup (Ib \in P) \cup \dots (In \in P)$

That is, the ontology (\mathcal{O}) of the models (Ia, Ib, \dots) of the formal system F consists of the: symbols that make up the alphabet of F , & the union of the axioms of F that are present in the interpretations which are the models of F . In accordance with the analogy from multivariate analysis, some small number of these axioms may provide an adequate ontology for the model(s) of F .

As already remarked, both the sharing of models across technical, temporal & organizational boundaries & the potential combination of models for larger scale detailed &/or meta-analysis could much higher effectiveness & productivity in the use of models in scientific, technical & business contexts. The possibility of using a small number of statements (WFFs.) as a close to lossless representation of a complex model provides the possibility that this small set of statements could be reasoned about as an appropriate representation of the model. It also means that it would be much simpler to share this concise representation than to have to convey & reason about the larger, more complex model. Finally, the availability of this more concise representation provides the possibility of combining models, so long as the provisions of pragmatic satisfiability are preserved. Model combinations could provide the possibility of many types of analysis & meta-analysis which are not possible through the separate analysis of models.

5. Bibliography

Hartzband, D.J. 1978. Eine Logik für das Ableiten der minimalen grundlegenden Annahmen für mehrfache Modelle (A logic for the derivation of minimal basic assumptions for multiple models). Dissertation submitted in partial fulfillment for degree of Doctor Of Science. Faculty of Mathematics. University of Hamburg. Federal German Republic.

Hartzband, D.J. 1990. Characterization of multivariate probability density functions by linear combination decomposition. Group for Artificial Intelligence & Learning (GRAIL). Stanford University Distinguished Lecture.

Neches, R., R.E. Fikes, T. Finin, T. Gruber, R. Patil T. Senator & W.R. Swartout. 1992. Enabling technology for knowledge sharing. AI Magazine. 12(3):16-36

Robinson, Abraham (1963), *Introduction to model theory and to the metamathematics of algebra*, Amsterdam: North-Holland, [ISBN 978-0-7204-2222-1](#).

Suchenek, M. 1985. On the asymptotic decidability of model-theoretic forcing. Raporty Badawcze 63. Instytut Informatyki. Politechnika Warszawska.

Suchenek, M. 1986. Nonmonotonic derivations which preserve pragmatic truth. Technical Report. Computer Science Department. Wichita State University. Wichita, Kansas, USA.