# Bill of Experiments: A Tool for Scientific Knowledge Management*

Jano Moreira de Souza [1,3], Leonardo Figueiredo Cardoso [1], Alberto Sulaiman [1,2], Jonice Oliveira [1]

[1] Computer Science Department, Graduate School of Engineering, COPPE/UFRJ, PO Box 68513, CEP 21945-970, Rio de Janeiro, Brazil
{jano, cardoso, sulaiman, jonice}@cos.ufrj.br
[2] Information Science Coordination, IBICT/MCT, Rio de Janeiro, Brazil
[3] Institute of Mathematics, UFRJ, Rio de Janeiro, Brazil

**Abstract.** This paper presents the Bill of Experiments, a concept aiming at Knowledge Management of Scientific Experiments, representing knowledge about and used by scientific experiments; and a first prototype of its implementation. Experience drawn from environments for scientific and engineering software development and reuse suggests that this is a promising approach for scientific knowledge representation and reuse. In analogy with data modeling concepts, such as the classical "Bill of Materials", a recursive aggregate, we propose "Bill of Experiments". An abstract workflow is used to represent the relationship between: data, models and experiments, as well as the instantiated workflow represents a study case of relating data, models, programs, institutions and researchers. The prototype translates the workflow into production rules to be processed by the inference engine built into a relational DBMS. It will be integrated to the SPeCS environment as a tool for reuse of models, programs, data and workflow definitions.

## 1 Introduction

This paper is based on original concepts of the experience on the development of scientific software, when research in software reuse for High Energy Physics environments, HEP, in the European Organization for Nuclear Research - CERN was happening. In that occasion the aim was the development of workstations configured for development of scientific software environments, related to construction, evaluation, documentation and management of this category of systems, as described in the TABA-HEP project (SOUZA et al., 1990; SOUZA et al., 1991; WERNER et al., 1990).

The integration effort of hundreds of international scientists' work, that could be considered a technological Babel because of platforms and development methods heterogeneity, was a research problem. There was a consensus in the necessity of

* This work was partially supported by CNPq and CAPES Brazilian research agencies.

Software Engineering and Artificial Intelligence techniques, mainly in Expert Systems (SOUZA et al., 1990), that allowed the construction of an heuristic for reuse of software pieces, whose knowledge contained in the source code could be termed currently as business rules (SULAIMAN et al., 2000).

In the context of scientific experiments reuse and, in analogy with fundamentals of data modeling, specially the recursive relationship part-component (CHEN, 1976), as well as its natural unfolding in "recursion using aggregation" (RUMBAUGH et al., 1991), the classic concept of Bill Of Materials, or BOM is being used, as cited in (SULAIMAN, 1999), integrated to an inference machine based on production rules (PASSOS et al., 1995; SULAIMAN et al., 1995).

The Bill of Experiments, or BOE, prototype is the first result of a research where the focus is on Knowledge Management for Scientific Experiments. The main objective of this tool is to improve the access, sharing, reuse and innovation of scientific knowledge, helping researchers to represent, share and retrieve the knowledge on models, programs, definitions of workflow and data on scientific experiments.

The remainder of the paper has the following structure: section 2 presents the concept of Scientific Knowledge; section 3 presents the concept of Bill of Materials, BOM; section 4 describes the tool for reuse of scientific knowledge, the Bill of Experiments, BOE; section 5 shows the architecture of the BOE tool; section 6 shows examples of BOE's use; section 7 talks about future work and combines BOE with other research, such as cooperative decision (PALMA et al., 2000; PALMA et al., 2001) and the use of domain models and ontology for scientific experiments (MANGAN et al., 2001); and section 8 is the paper's conclusion.

## 2 Scientific Knowledge

We have reached the knowledge era. Research centers, not only the operational enterprises, are more and more involved with this reality. Researchers create and change information quickly and in a larger volume than before. Although the biggest part of this interchange is being made under documentary form it also has been observed a great occurrence of informal or tacit knowledge in the interactions between people. The main target of knowledge management for scientific purposes is to answer dynamically and efficiently the changes in a highly unexpected external environment, allowing researchers to store, spread and manage the knowledge of their specific areas.

In reality, reaching these goals, connecting all the activity areas of a research organization, is not an easy task, given that Knowledge Management is still relatively recent. The availability of standardized models is still limited. The scientific research management is being critically dependent of effective scientific data and information organization techniques.

The fast development in computers, communication and scientific instruments, allowed researchers to collect, generate, process and share an unprecedented amount of data. In the nineties, there was an enormous growth in computation capacity, the data storage costs had diminished and the access to fast and reliable communication networks became reality.

Paradoxically the data production cost is very high: satellites, particles´ accelerators and supercomputer centers represent sources of information generation that costs billions. Scientific data are growing in quantity and complexity, through electronic sensors´ use, simulations and experiments. Without effective retrieval, analysis, and manipulation of those data, the investment made will not bring the expected benefits for society.

Scientific data are not only becoming larger in volume, but the extraction of useful knowledge is also becoming more difficult, as the data, models, tools and experiments become more complex.

In his work, Sveiby (SVEIBY, 2001) identified two big tendencies in the concepts of Knowledge Management: the IT researchers and the human resource management researchers. While the IT group uses mainly computer science to develop information systems, document management, groupware etc, the second one works in evaluation, change and improvement of individual human competencies and behaviors.

We propose an IT approach tool based on AI and database concepts that has as the objective to integrate the researcher with the whole knowledge contained in a knowledge base about experiments, data and formal models related with his/her research area. The desired contribution is the organized capture of the specific knowledge (inherent to scientific experiments), stimulating the research community to feed the knowledge base and benefit with the cooperation, sharing and reuse of the organized knowledge.

## 3 The Bill of Materials Concept (BOM)

In the database literature, one can easily find examples to express recursive relationships in the conceptual and logical data modeling. Examples are found in (CHEN, 1976; DATE, 2000; ELMASRI et al., 1999), among others, and its natural unfolding, is the “recursion using aggregation" (RUMBAUGH et al., 1991).

One of the most common instances of it is the "Bill of Materials" or BOM, that is the relationship among items whose components, in general, can be other items, and each new item, in return, can have other components, creating the recursive relationship as shown in figure 1.

The instance of BOM found in (SULAIMAN, 1999) is an application for planning components purchase for construction or assembly of electronic components. Each BOM is an aggregate of components and each component is composed by items. A component will be ready to be delivered only when all its items are ready and instantiated. Consequently, each part instance is completely identified with the respective price, estimated period of delivery, supplier origin, and availability. Each item is done only when one part is chosen between all of the possible choices. If a part is complex, it can be expressed as a new BOM (its recursive nature), see figure 2.

In other words, Sulaiman (SULAIMAN, 1999) specifies production rules to represent component/item relationships, where each component of a BOM is a head of a rule. All the items of a component together are the body of the same rule. Every item describes a functionality (in the sense that an item can be instantiated by a part, that is physical), not a physical piece. Therefore a part is an instance of an item. A component is instantiated when each item is instantiated by one part.

In (SULAIMAN, 1999) the market segment of electronics is also considered, and the existence of a search engine to find and store in a database system the offer of parts by several suppliers, examining prices and delivery times. The transactions between the search agent and the suppliers of pieces constitute a B2B WEB system, and XML code corresponding to the rule bases of the model is shown in figure 3. The customer's interaction is based on the B2C concept and it is planned so that a the buyer can evaluate a purchase on the grounds of price, delivery times, makers, or a combination of those variables so that a better cost-benefit relationship is reached. It is assumed that a customer has the knowledge of which parts needs to be bought for the correct assembly of a component.

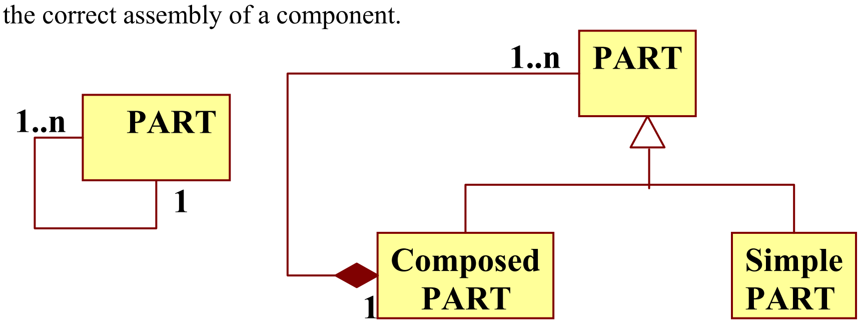



**Fig. 1.** Class diagram exhibiting the self-relationship between pieces and their components, unfolding into a recursion using aggregation

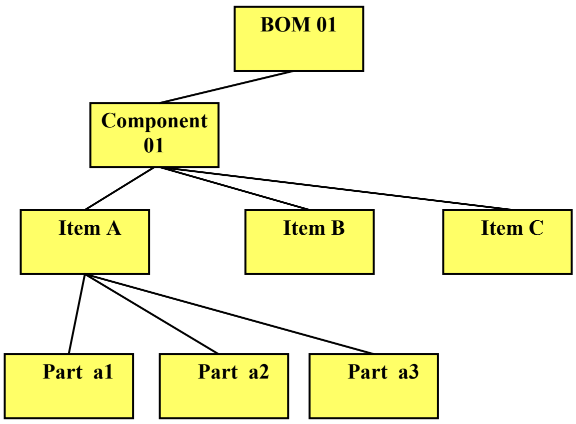


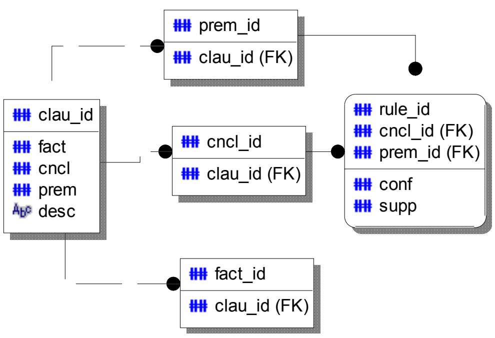

**Fig. 2.** Hierarchy of a BOM

A component is considered ready for delivery from the moment that all their items are instantiated and, consequently, each part is completely identified with the respective price, estimated time for delivery, supplier, and readiness. Figure 4 depicts the search space for the example in figure 2.

The constraints of the model are the following: a) the delivery term of a component is always, at least, equal to the largest time for delivery amongst all of the involved items; b) the smallest price of a component is of the sum of the prices of the component items, which aggregates to the component. Exploring the deductive nature of the system, the user can request explanations on prices, periods and assembly of some component.

In (SULAIMAN, 1999) one can find solution proposals for the component of better cost-benefit, with the users inserting limits of tolerance for the price variables and period, and assigning weights for this variable. The “weight” is understood as the importance of a variable in relation to the others.

**BOM Rule Base**




```
<!-- each BOM has its own knowledge
base -->
<!ELEMENT know (rule)*>
<!ATTLIST know BOM_name CDATA
#REQUIRED>
<!-- each rule has antecedents and a
consequent -->
<!ELEMENT rule ((prem)* (cncl)?)>
<!-- each rule has confidence and
support -->
<!ATTLIST rule (rule_id (#PCDATA)
         conf (#PCDATA)
         supp (#PCDATA)) >
<!-- each antecedent is an Item,
here -->
<!ELEMENT prem CDATA #REQUIRED>
<!ATTLIST prem prem_id (#PCDATA)
           Item>
```

**Fig. 3.** Relational Schema of the rule base and part of the related DTD/XML

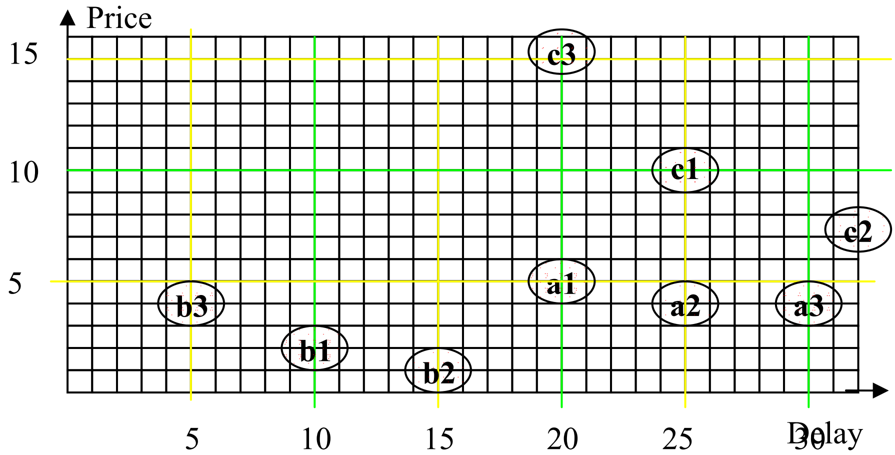
## 4 A Tool for the Reuse of the Scientific Knowledge: Bill of Experiments

In this section, we introduce the concept of Bill of Experiments (BOE), a deductive system to support the reuse of Scientific Knowledge. As exposed previously, conceptually, BOE is based on BOM. For a better understanding of this proposal, an analogy is established between the preparation of the scientific experiment and the purchase of items for the assembly of components.

The process of simulation or that of experiment application using one or more sets of scientific data can be compared to the composite item assembly, generating new relevant information for the research community. Just as several items can make up a component, the composition of data, experiences using such data and the execution of simulations can compose new knowledge.

As well as a new obtained part can serve as an item to be used in the composition of a new component, a new mass of data generated in BOE can serve as a base for the application of another mathematical model or experiment that will generate a third information and so forth. In other words, each discovery using BOE’s data is appended to the Knowledge Base and will hereafter be available to be used as given in the generation of new results.

The focus of the present work is to present the storage of scientific data and their relationships in a deductive environment. After that, we shall comment the emphasis of the present work for data can and is being generalized for the other components of a scientific environment like models, tools, workflows and others. The following paragraphs explain each term of analogy implemented in this prototype.




A <a1,20,5>,<a2,25,4>,<a3,30,4>
B <b1,10,2>,<b2,15,1>,<b3,5,4>
C <c1,25,10>,<c2,32,7>,<c3,20,15>

**Fig. 4.** Search space considering Price and delivery time

**Simple item or part:** In BOE instead of simple items, there are groups of data that can serve, as a base for the application of some model or scientific experiment. As an example, consider the value of gravity, the table of chemical elements, some formulas, among others.

**Composite item or Component**: It consists of the group of data resulting from the application of some model or scientific experiment on the simple data or on other composed data.

**Supplier:** As well as a supplier manufactures a part; the scientific data can have an origin or an owner. This way, research institutions, companies and government entities can be represented as "suppliers" of scientific datasets. In a similar way, several suppliers can manufacture the same item and several organizations can make available the same scientific data, being under the responsibility of the "buyer" to choose which "supplier" can give him/her the "item" in the way that better suits him/her, depending on the cost and time of "delivery".

**Price:** In the same way that an item of the BOM has costs, the masses of scientific data also possesses them. It may derive from its purchase, obtaining, retrieval, discovery or availability. The metric of the cost tends to be richer in BOE than in BOM, which makes the problem more complex. In BOE the metric of cost can be:

the price of the availability of a datum, payment of intellectual property for the use of information, the cost of retrieving or field research, the transmission cost of the data or the cost of a researcher's adviser. As in BOM, the smallest cost of a component consists of the sum of the smallest prices of the items, which aggregates in a component (Equation 1).

$$cost(Component) = \sum_{\forall x \in Component} \min[cost(x)] \tag{1}$$

**Time:** The analogy is direct, because just as a component is not always available for immediate use, the same can happen to scientific datasets. It may be necessary to wait for the sending of information through electronic mail, the complete execution of a model or the time of the accomplishment of an experiment in order to collect their results to feed the Knowledge Base. The availability time of a composed data is always the largest time of delivery amongst all data, plus the time for the execution of the models or experiments (Equation 2).

$$delay(Component) = \left( \max_{\forall y \in Component; y = Meta-Data} [delay(y)] \right) + \sum_{\forall z \in Component; z = Experiment, Model} delay(z) \tag{2}$$

**Centralized database:** It is assumed that an integrated base of scientific data, mathematical models, experiments and the relationships amongst them already exists, or will be provided by a environment such as SPeCS (PALMA et al., 2000; PALMA et al., 2001) with a degree of transparency.

The main objective of BOE is to provide data, models, programs, workflow definitions and data of all the above elements from scientific experiments, and if they don’t exist, to guide and help the researcher on how to obtain them, becoming a tool to improve the access, sharing and consequent innovation of the scientific knowledge. BOE is not just a database in which one can just drop information in digital form, but is also a potential resource for researchers in their search for scientific results, which will lead to new information of their interest, and in this way help the scientific production.

For each new experiment, its steps are recorded, becoming a source of reference to other registered specialists of the area; sharing innovative information and adding value. Such information tends to be adequate in terms of hypotheses, problems, models, times and prices, not just limited to a simple existence indicator. Besides that, the deductive environment allows them to obtain explanations on relevant data through a backward chaining algorithm, using an inference machine. Through the experience contained in the Knowledge Base it is possible for the researcher to discover all data and models that originated an experiment. Researchers will be able to find appropriate cost-benefit relations for collecting important data for their research and in this way to have his work less expensive and more dynamic.

At the moment, in this research, the representation of masses of scientific data, experiments, and mathematical models were limited to their data. In other words, the scientific data are not stored themselves, only the detailed description of the same ones is registered with inherent attributes (owner institution, price, delivery times, dates, validity etc).

## 5 BOE Architecture

In the present version the BOE prototype works integrated with a relational database that stores the research institutions and researchers' information, prices and delays for data, experiments, mathematical models and corresponding tools, and the workflow of the experiments. These are **instantiated** workflows, connecting instances of experiments, data, models, tools, etc. They are the *glue* that put them together to form a study case. When BOE gets integrated to SPeCS, it will be possible to represent and query abstract workflows of experiments, besides the instantiated ones.

As cited in (WAINER et al., 1996) "Whereas office work is about goals, scientific work is about data. Collecting, generating, and analyzing large amounts of heterogeneous data is the essence of such work, or at least of the components of scientific work that are more naturally the target of WFMS. Gathering and merging data from various experiments, generating data from a computer model or performing statistical analyses in the data, are among the activities that could profit from WFMS support". We can't forget that the goals (achieved results like new particles, new formulas etc) are still present, but in scientific work there are needs for more data.

There are tools in the market to make automated data acquisition, data analysis, and process control software, for example, DAQLab (AZEOTECH, 2002), Citect (CITECT, 2002) and LabView (NATIONAL INSTRUMENTS, 2002). Although such tools work with workflows and data acquisition, they are not able to plan which data will be used in a process managed when there is a hypothesis that BOE assumes, in other words, various sources of the same data generating masses of data which have costs, deadlines and differentiated qualities.

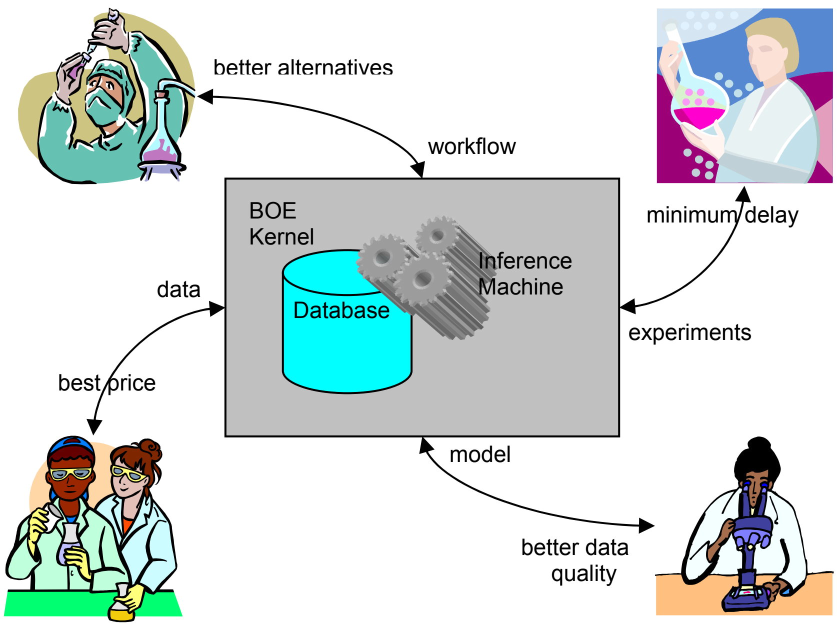



**Fig. 5.** BOE Architecture

BOE stores workflows as production rules, which allows the use of an inference machine for the attainment of detailed information on them. Besides the capacity to expose the viability of instantiation of a workflow, pointing out better alternatives for the tasks, the inference machine explains all its steps, demonstrating the reasons for making its decisions.

In this architecture, the collaboration of researchers and institutions is fundamental. Researchers and scientists are responsible for publishing their research work, with the detailed stages of their scientific workflow. The institutions and companies interested in selling or making available their data and experiments contribute with information of prices and delivery time for them. At the same time, the agility in obtaining important data for each research will improve and its cost will decrease.

Besides these capabilities, the reuse is motivated because researchers of common areas can optionally share their workflows and know other people's work in progress.

# 6 BOE in Action

In this section three examples of the use of the BOE tool are presented. The first example is useful to show the representation with production rules of two alternatives for obtaining a result in the seismic area. An imaginary scientific application is presented as a second example (a toy example). Finally, a real scientific workflow is presented as third example. For the last two cases, the results generated by the tool will be shown.

## 6.1 Inverse Problems

A mathematical model for inverse problems is used to illustrate a simulation. We defined inverse problems as those that are related with the obtaining of characteristics or parameters of mathematical models based on the reduced knowledge of the solutions of the corresponding equations.

A typical example of inverse problem in the area of seismic research is: "*Determination of data of oil reserves, based on superficial measures of reflection data or of dispersal of waves"* (ZUBELLI, 1999).

It is observed that in this case the application of different methods on two different datasets provides results scientifically similar. Therefore, this example perfectly fits in the structure of BOE, in which, the result of the application of a rule can be obtained starting from different experimental facts.

We defined the following rules for the use of BOE with this example:

```
Oil_Reserv(X):-Reflect_Data(Y),Inverse_Prob_Model(Y,X)

Oil_Reserv(X):-Disper_Waves(Y),Inverse_Prob_Model(Y,X)
```

`Oil_Reserv` stands for the scientific data of oil reserves, `Reflect_Data` stands for the superficial measures of reflection data, `Disper_Waves` are the measures of the dispersal of waves and `Inverse_Prob_Model` represents the mathematical model used in the simulation.

In this example, a researcher interested in data of petroliferous holds can obtain results through the application of different simulations, as demonstrated above. In this example, the BOE tool will make available for the user two options: one with the smallest price and another with the smallest delivery time for obtaining the needed data. The user should opt for one of the two alternatives.

### 6.2 A Toy Example

This example uses a hypothetical workflow (figure 6) with institutions, data, models and experiments. It will bring the exact notion of the functionalities of the tool, showing explanations about the calculations of smaller price and time.

We defined `INST1`, `INST2` and `INST3` as research institutions, and `D1`, `D2`, `D3`, `D4` and `D5` as scientific datasets. Furthermore we define M1 and M2 as mathematical models, and E1 and E2 as scientific experiments. The datasets `D1` and `D2`, the models `M1` and `M2` and the experiments `E1` and `E2` are property of the institutions described previously. Consequently, such items are facts in our Knowledge Base and they possess well defined costs and delays. Table 1 exemplifies the costs and periods adopted for each item in this example.

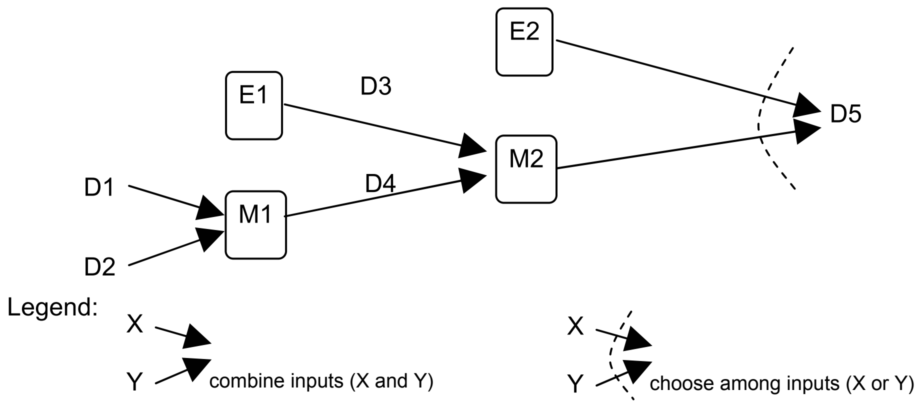



**Fig. 6.** Workflow involving Two Experiments and two Models to obtain result D5

The workflow presented in figure 6 was mapped to the following production rules:

```
D3(X):-E1(X)

D4(X):-D1(Z), D2(W), M1(Z,W,X)

D5(X):-D4(Z), D3(W), M2(Z,W,X)

D5(X):-E2(X)
```

| | INST1 | | INST2 | | INST3 | |
|---|---|---|---|---|---|---|
| | Cost | Time | Cost | Time | Cost | Time |
| D1 | $100,00 | 3 days | $100,00 | 2 days | | |
| D2 | $500,00 | 10 days | | | $400,00 | 12 days |
| M1 | $50,00 | 1 day | | | $30,00 | 2 days |
| M2 | | | $10,00 | 1 day | $12,00 | 3 days |
| E1 | $10,00 | 3 days | $12,00 | 2 days | $18,00 | 7 days |
| E2 | | | $570,00 | 10 days | | |

**Table 1.** Data for Workflow in figure 6

At the end of the inference, the tool makes available a report with all of the stages of the deductive process (summarized in figure 7, the middle portion was suppressed). In this example, the researcher is interested in dataset D5. The report details all the explanation of the deduction and the wanted answers are at the end.

**Inference Machine Tasks**

```
->** I'm beginning the proof of D5 to find the smallest pric
es and minimum delays **
->I discovered that D5 is a conclusion, I will try to prove
the existent premises.
    ->** I'm beginning the proof of D4 to find the smallest
    prices and minimum delays **
    ->I discovered that D4 is a conclusion, I will try to pr
    ove the existent premises.
        ->** I'm beginning the proof of M1 to find the small
        est prices and minimum delays **
        ->I discovered that the smallest price of M1 is US$3
        0,00 and it's available in the institution(s) INST
        3
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-> ==> E2 True
->CNCL ==> Rule: D5 :- E2 was proven with success!
->CNCL ==> I discovered that the smallest price of D5 is US$
570,00
->CNCL ==> I discovered that the minimum delay of  D5 is 10
day(s)
->I asked the user which rule he would like me to use to con
tinue...
->He chose the rule D5 :- D4, D3, M2  that generated Price =
 US$550,00 Delay = 12 day(s)
```

**Fig. 7.** Inference Report Generated by BOE for the Toy Workflow

Of course, pruning and some heuristics can be used to reduce the search space. The complexity of the algorithm should not be a problem, given the number of alternatives expected to be found in scientific workflows.

### 6.3 A Real Example

This third example demonstrates the contribution of the BOE tool for a real scientific workflow. The stages and operation of the workflow are detailed in (AILAMAKI et al., 1998) and summarized below.

The objective of the experiment is to produce daily forecasts of near-surface temperatures in cranberry bogs in Wisconsin. These forecasts give cranberry farmers advanced warning of over-night frost conditions, so they can take action to protect their vines from frost damage.

1. Around noon each day, satellite and ground-based meteorological observations are processed in the Atmospheric Sciences Department of UW, generating a 24-hour weather forecast at several heights in the atmosphere for the whole United States;
2. This US forecast is fed into a Bog Forecast Extraction program that extracts forecasts for points that are 25 meters above specified cranberry bog locations;
3. These forecasts are sent to the Soil Science Department where they are processed by CranEB, to derive a forecast for the level of the cranberry vines (canopy level);
4. Later in the day, as new weather observations become available, the initial 25m bog forecast can be updated:
   - Scaled CranEB output forecasts are compared with new observed weather conditions in a package of statistical routines.
   - Appropriate corrections to the original 25m bog forecast are determined, and CranEB is rerun. With this feedback mechanism, the canopy-level forecast is updated continuously throughout the day.
5. Text files generated by CranEB are fed into the DEVise Visualization tool that generates GIF plots of canopy temperature vs. time. These plots are then published on the Web.

The graphic representation of the workflow described above is in figure 8. The mapping for production rules generated the following result:

```
gif_graphs(X):-canopy_level_forecast(Y),
VISUALIZATION(Y,X)

canopy_level_forecast(X):-25m_bog_forecast(Y),
CRANEB(Y,X)

canopy_level_forecast(X):-25m_bog_forecast_upd(Y),
CRANEB(Y,X)

obs_level_forecast(X):-25m_bog_forecast(Y), CRANEB(Y,X)
```

```
obs_level_forecast (X):-25m_bog_forecast_upd(Y),
CRANEB(Y,X)

25m_bog_forecast_upd(X):-25m_bog_forecast(Y),
ground_based_observations (Z),
obs_level_forecast(W),
STATISTICAL_ANALYSIS(Y,Z,W,X)

25m_bog_forecast(X):-us_forecast(Y),
BOG_FORECAST_EXTRACTION (Y,X)

us_forecast(X):-satellite_observations (Y),
ground_based_observations(Z),
US_FORECAST_MODEL(Y,Z,X)
```

When we consider only the metrics ***price*** and ***time***, it is concluded that the dataset "Canopy-level Forecast" will be obtained cheaper and with shorter delivery time when the user chooses not to use the process "STATISTICAL ANALYSIS". However, with next version of BOE, implementing the metric for quality, probably, the "Canopy-level Forecast", generated starting from the successive refinement of the entrance of the process "CRANEB" ("25m Bog Forecast (upd.)"), will provide in more quality.

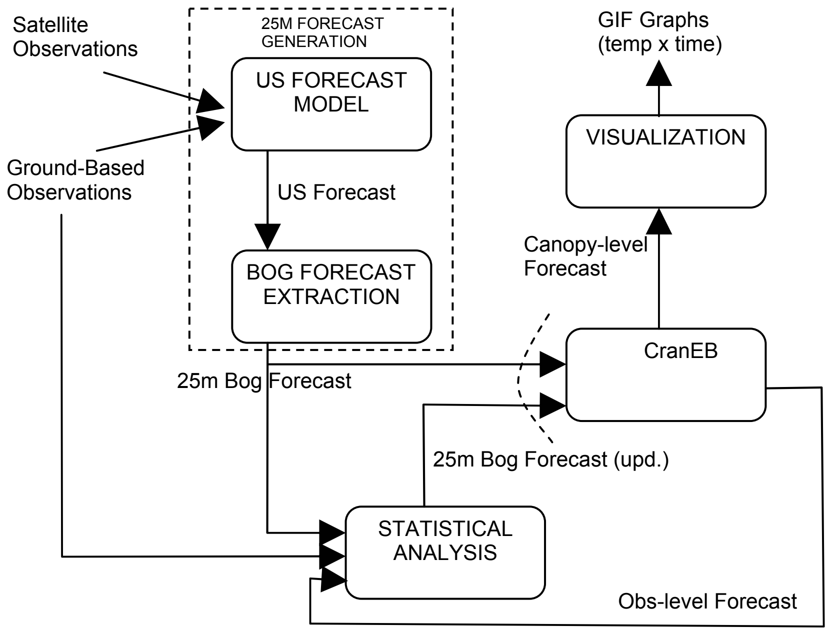



**Fig. 8.** The Cranberry Workflow (AILAMAKI et al., 1998)

The data on institutions, prices and delays for the simple data and involved models were defined arbitrarily. With the objective of finding a better alternative for

"Gif_Graphs", the BOE tool generated the inference report that is summarized in figure 9 (the middle portion was suppressed).

**Inference Machine Tasks**

```
->** I'm beginning the proof of gif_graphs to find the small
est prices and minimum delays **
->I discovered that gif_graphs is a conclusion, I will try t
o prove the existent premises.
    ->** I'm beginning the proof of canopy_level_forecast to
     find the smallest prices and minimum delays **
    ->I discovered that canopy_level_forecast is a conclusio
    n, I will try to prove the existent premises.
        ->** I'm beginning the proof of 25m_bog_forecast to
        find the smallest prices and minimum delays **
        ->I discovered that 25m_bog_forecast is a conclusion
• • • • • • • • • • • • • • • • • • • • • • • • • • • • • • • •
    e smallest prices and minimum delays **
    ->I discovered that the smallest price of VISUALIZATION
    is US$5,00 and it's available in the institution(s) IN
    ST2
    ->I discovered that the minimum delay of VISUALIZATION i
    s 1 day(s) and it's available in the institution(s) IN
    ST1, INST2
-> ==> VISUALIZATION True
->CNCL ==> Rule: gif_graphs :- canopy_level_forecast, VISUAL
IZATION was proven with success!
->CNCL ==> I discovered that the smallest price of gif_graph
s is US$232,00
->CNCL ==> I discovered that the minimum delay of  gif_graph
s is 5 day(s)
```

**Fig. 9.** Inference Report Generated by BOE for the Cranberry Workflow

## 7 Future Work

At this time the integration of BOE is being studied with other researches in development by the same group, such as cooperative decision (PALMA et al., 2000; PALMA et al., 2001) and the use of domain models and ontologies for scientific experiments (MANGAN et al., 2001).

Besides this integration, we are testing a larger number of metrics, such as quality, reliability, readiness, reusability and usability, in addition to those restrictions already available: price and time.

BOE will access distributed databases of suppliers and institutions directly, using X-Arc (PINTO et al., 2001) a XML based layer for the integration of spatial data and metadata, and make available its Knowledge Base on the Web.

Ontology bases are being created for the integration of similar data and the "***Diary of the Experience***" Tool is being developed, which will accompany all of the steps executed in an experiment, "learning" the researcher's profile automatically and customizing the data that is assumed necessary for his research.

## 8 Conclusions

This paper presented the concept and prototype of the tool Bill of Experiments (BOE). The main objective of BOE is to provide data about and for scientific experiments and, in the absence of these, to guide the researcher in obtaining it, becoming a tool to improve the access, sharing, reuse and consequent innovation of the scientific knowledge.

An abstract workflow is used to represent the relationship between data, models and experiments, as well as the instantiated workflow represents a study case relating to data, models, programs, institutions and researchers.

This study is based on an analogy with the model of Bill of Materials presented in (SULAIMAN, 1999). We showed that our analogy is valid, once the recursive nature of parts that are composed by other parts and have a cost and delay for delivery can be applied to scientific experiments.

BOE could help datasets, experiments, models, workflows, and other scientific artifacts to be reused in new experiences. The institutions or researchers should attribute a cost function for each item. We demonstrated the viability of our analogy with a prototype.

In the representation of a workflow with production rules, the interdependence amongst experiments was captured in a clearer and more precise way. Besides that, obtaining explanations in a deductive environment is much easier than when one uses some mathematical method of optimization. Even for complex computational problems, the scientific workflows tend to be simple, what reduces the concern with the complexity of the deductive algorithm.

References

AILAMAKI, A., IOANNIDIS, Y. E., LIVNY, M., 1998, "Scientific Workflow Management by Database Management". In: *Proceedings of the 10th International Conference on Scientific and Statistical Database Management*, pp. 190-199, Capri, Italy, July.

AZEOTECH, 2002, "Quick Tour of DAQLab". In: http://www.azeotech.com/quicktour.html, Accessed in 03/2002.

CHEN, P. P., 1976, "The Entity-Relationship Model - Toward a Unified View of Data", *ACM Transactions on Database Systems (TODS)*, v. 1, n. 1, pp. 9-36.

CITECT, 2002, "Citect - Product Information". In: http://www.citect.com/html/citect.htm.

DATE, C. J., 2000, *An Introduction to Database Systems*. 7th Edition, Addison-Wesley.

ELMASRI, R. A.,NAVATHE, S. B., 1999, *Fundamentals of Database Systems*. 3rd edition, Addison-Wesley Pub Co (Net).

MANGAN, M. A. S., MURTA, L. G. P., SOUZA, J. M., et al, 2001, "Uma comparação entre Modelos de Domínio e Ontologias na Gestão do Conhecimento". In: *Proceedings of Fourth Internation Symposium on Knowledge Management / Document Management*, Curitiba. Brazil, August.

NATIONAL INSTRUMENTS, 2002, "National Instruments - Measurement and Automation". In: http://www.labview.com/, Accessed in 03/2002.

PALMA, S. J. M., SOUZA, J. M., STRAUCH, J. C. M., et al, 2001, "Coordination aspects in a spatial group decision support collaborative system". In: *Proceedings of the 16th ACM SAC2001 symposium on on Applied computing*, pp. 182-186, Las Vegas, NV USA, March.

PALMA, S. J. M., STRAUCH, J. C. M., SOUZA, J. M., et al, 2000, "Specs - A Spatial Decision Support Collaborative System For Environment Design". In: *CSCWD 2000*, Hong Kong, November.

PASSOS, E., SULAIMAN, A., GARCEZ, C., et al, 1995, "A Conceptual Model for a Knowledge Base Homogeneously Stored in a Database Environment". In: *Proceedings of 12th Brazilian Symposium on Artificial Intelligence - SBIA '95*, pp. 295-301, Campinas, Brazil, October.

PINTO, G. R. B., STRAUCH, J. C. M., SOUZA, J. M., et al, 2001, "X-Arc Spatial Data Integration in the SPeCS Collaborative Design Framework". In: *CSCWD `2001also to appear in CACM*, London, Canada, July.

RUMBAUGH, J., BLAHA, M., PREMERLANI, W., et al, 1991, *Object-Oriented Modeling and Design*. 1st edition, Prentice Hall.

SOUZA, J. M.,ROCHA, A. R. C., 1990, "TABA-HEP: a Heuristic Workstation for Developing HEP Software". In: *Proceedings of the first International Workshop on Software Engineering, Artificial Intelligence and Expert*

*Systems in High Energy Physics*, pp. 109-115, Lyon Villeurbanne, France, March.

SOUZA, J. M.,ROCHA, A. R. C., 1991, "The Architecture of TABA-HEP Worstation". In: *Proceedings of the International Conference on Computing in High Energy Physics '91*, pp. 347-350, Tsukuba, Japan, March.

SULAIMAN, A., 1999, "Using Datalog Data Model and Data Mining to Solve Bill of Materials Personalization Problems". In: *International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems*, Santa Clara, California, April.

SULAIMAN, A., MATTOSO, M., SOUZA, J. M., 1995, "Um Shell de Sistema Especialista Implementado em O2, USE O2!". In: *II Workshop de Bancos de Dados não convencionais*, Rio de Janeiro, Brazil, December.

SULAIMAN, A., SOUZA, J. M., NETO, F. P., et al, 2000, "Tecnologia da Informação para a Gestão do Conhecimento: Regras de Negócios". In: *Proceedings of Third International Symposium on Knowledge Management / Document Management*, Curitiba, Brazil, November.

SVEIBY, K.-E., 2001, "What is Knowledge Management?". In: http://www.sveiby.com/articles/KnowledgeManagement.html, Accessed in 03/2002.

WAINER, J., WESKE, M., VOSSEN, G., et al, 1996, "Scientific workflow systems". In: *NSF Workshop on Workflow and Process Automation*, Athens, Georgia, USA, May.

WERNER, C. L., PIMENTA, M., VARELA, J., et al, 1990, "Fado Tagging system: Decoupling Logic from Code". In: *Proceedings of the first International Workshop on Software Engineering, Artificial Intelligence and Expert Systems in High Energy Physics*, pp. 103-108, Lyon Villeurbanne, France, March.

ZUBELLI, J. P., 1999, "An Introduction to Inverse Problems; Examples, Methods, and Questions (Minicourse)". In: *22nd Brazilian Mathematics Colloquium*, Rio de Janeiro, Brazil, July.