A Knowledge-Based System Approach for Scientific Data Analysis and the Notion of Metadata

Epaminondas Kapetanios
Research Center Karlsruhe—Technology and Environment
Institute of Applied Computer Science
Federal Republic of Germany

Ralf Kramer
Forschungszentrum Informatik (FZI)
Federal Republic of Germany

Abstract

Over the last few years, dramatic increases and advances in mass storage for both secondary and tertiary storage made possible the handling of big amounts of data (for example, satellite data, complex scientific experiments, and so on). However, to the full use of these advances, metadata for data analysis and interpretation, as well as the complexity of managing and accessing large datasets through intelligent and efficient methods, are still considered to be the main challenges to the information-science community when dealing with large databases. Scientific data must be analyzed and interpreted by metadata, which has a descriptive role for the underlying data. Metadata can be, partly, a priori definable according to the domain of discourse under consideration (for example, atmospheric chemistry) and the conceptualization of the information system to be built. It may also be extracted by using learning methods from time-series measurement and observation data. In this paper, a knowledge-based management system (KBMS) is presented for the extraction and management of metadata in order to bridge the gap between data and information. The KBMS is a component of an intelligent information system based upon a federated architecture, also including a database management system for time-series-oriented data and a visualization system.

Introduction

Trying to give a rigorous definition of metadata presents some difficulties. At first glance, the term metadata denotes data that can be found at a more abstract level describing some other set of data. At this point, let us recall the distinction between data and information given by D.C. Thichritzis and F.H. Lochovsky [1]:

> It's important to realize the distinction between data and information. Data are facts collected from observations or measurements. Information is the meaningful interpretation and correlation of data that allows one to make decisions.

According to this definition, metadata seems to be something between data and information. It can be considered a descriptive tool of the underlying data that can be used in order to reach the information level to be provided. In this sense, metadata can be a priori defined according to the already known domains of scientific discipline and experiment (information system conceptual modeling), or may be derived from the underlying data, especially in cases where a non-a priori-definable domain of discourse addressing the scientific discipline must also be considered.

Therefore, metadata is going to be used for the analysis and interpretation of scientific data aiming at the extraction of information (see Figure 1a). On the other hand, information in terms of already known knowledge concerning the domain of discourse of the scientific application can be conceptualized and used in order to steer the learning process (extraction of knowledge-metadata) from the underlying measurement and observation data (see Figure 1b).

At this point, two questions arise:

- What metadata should be captured and modeled as knowledge?
- How can it be represented and organized to support data analysis and interpretation as well as the extraction and modeling of new knowledge?

Based on a real application concerning a scientific experiment for the study of atmospheric (upper troposphere—lower stratosphere) chemistry and phenomena, we will try to illustrate the issues of metadata as knowledge and its representation and organization. The remainder of this paper is organized as follows. In the next section, a short description of the scientific experiment under consideration is given, in conjunction with two scenarios concerning the validation

1051-9173/95 $4.00 © 1995 IEEE

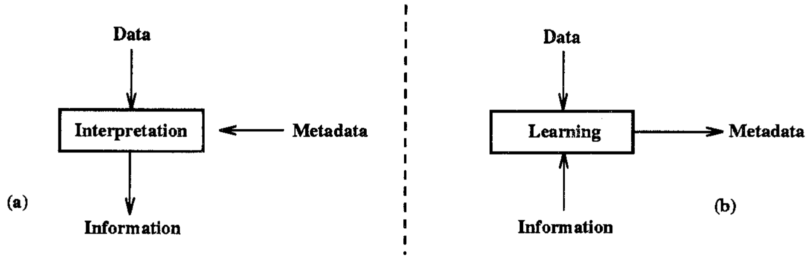

Figure 1.

activities of scientific data for the extraction of scientific results.

Based on these scenarios, a description of metadata in terms of knowledge structures is then given. Subsequently, the last section deals with the management and usage of metadata aiming at scientific datasets analysis and/or access, and not only at data management (catalogs). Consequently, we describe a knowledge-based system that enables the organization of metadata as knowledge servers, as well as the connection of various knowledge elements (pieces of metadata), in order to construct justification structures about scientific hypotheses and/or conclusions.

According to the diversity of scientific users and their needs, the main issues of the user interface under consideration (a metadata browsing and querying combination) is then presented wherein the main principles of object orientation, simplicity, and self-guiding have been taken into account. Finally, we briefly discuss those systems and approaches that we think bear the most direct relationship to our approach, and then we present the conclusions and future work.

The experiment and the scenarios

The MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) limb sounder is a remote sensing instrument that is going to be installed on various platforms (balloon, aircraft, satellite–ENVISAT mission) to undertake atmospheric (upper troposphere–lower stratosphere) data measurements ([2],[3]). Specifically, the understanding of complicated ozone-hole processes presupposes an understanding of the reactions and interactions of stratospheric trace gases (for example, O_3, ClO_x, NO_x, HCl, HNO_3, $ClONO_2$).

Raw data delivered by the instrument will be transformed into observation data by applying calibration and trace gas-retrieval algorithms. Trace gas-related observation data must be visualized and validated according to a given theory and corresponding background knowledge. The result of this validation could be a potential anomaly that is related to a subset of the source observation data or the discovery of unknown facts, which may revise the existing theory. Figure 2 outlines the general process of validating observed phenomena, and consequently, scientific laws. The following scenarios will illustrate this process.

The scenarios

The scenario of ozone depletion. We consider ozone (O_3) as a single concept in the domain of a scientific discipline. This concept is used to describe and address the ozone-related instances (observation data). Based on these instances, a significant trend of ozone concentration through a certain period of time (for example, one year) can be provided. This trend must be validated by a given theory that ozone can be depleted only if the depletion parameters have been changed. Therefore, a depletion trend that cannot be validated by a certain theory is a candidate for an anomaly. Recently (1985), the explanation based on this trend's source data of have led to the acceptance of ozone depletion processes and to a theory revision, through the introduction of heterogeneous chemistry.

The scenario of long-living gases. The group of trace gases N_2O, CFC_{12}, CFC_{11} is considered to be the taxonomy of long-living gases in the atmosphere. An insight must be given into their correlation quality. If they are not correlated, for instance, N_2O, CFC_{12}, according to the theory that dynamically dependent correlation of this trace gases group, then an anomaly has occurred. If this anomaly is not related to a subset of the underlying observation data, then the theory must be modified, which means that an unknown chemical process must be inserted into the theories. In parallel, the taxonomy of long-living gases must also be modified (in that CFC_{12} is going to be deleted from the related taxonomy).

Generally speaking, and for a better understanding of the knowledge (metadata) needed to analyze, interpret, and validate scientific data so as to provide and justify scientific results, a flow chart depicts the general process in Figure 2. A parallelogram symbolizes the piece of knowledge or metadata that is used by the corresponding activities of learning (extracting metadata) with help of a priori de-

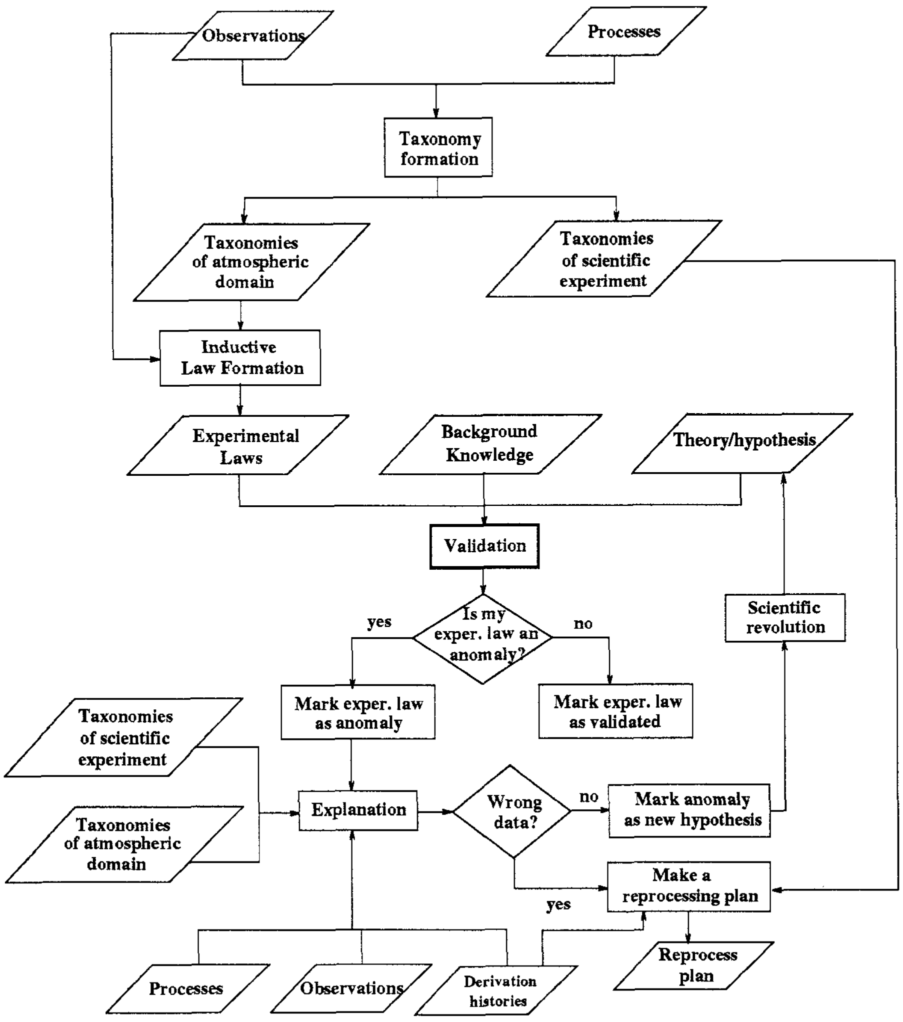

Figure 2.

fined knowledge (see also Figure 1b), and consequently, of using this metadata for analysis and validation (see also Figure 1a).

Metadata as knowledge structures

With respect to the scenarios given above, we will try to describe the metadata in terms of scientific knowledge structures. These definitions are closely related to those given in [4]. We will also distinguish between a priori defined knowledge and knowledge that may be extracted through a learning process and formed as metadata for data analysis and validation.

Measurements and observations. They are objects addressing the relevant instances of measured (or transformed into) observation data, for example, interferograms, calibrated spectra, trace gas distributions. These objects can be described by atomic or complex concepts. They enrich the semantics of the underlying scientific data and are considered to be equivalent to the database conceptual design of an information system (Figure 3). This knowledge is a priori definable and strongly related to the application domain (scientific experiment) under consideration.

Transformation processes. They are classes of objects representing atomic or complex concepts addressing the instances of transformation processes. They can take values by interactively changed parameters, versioning of specification and implementation of algorithms, and so forth, throughout the transformation processing chain of measurement data towards observations.

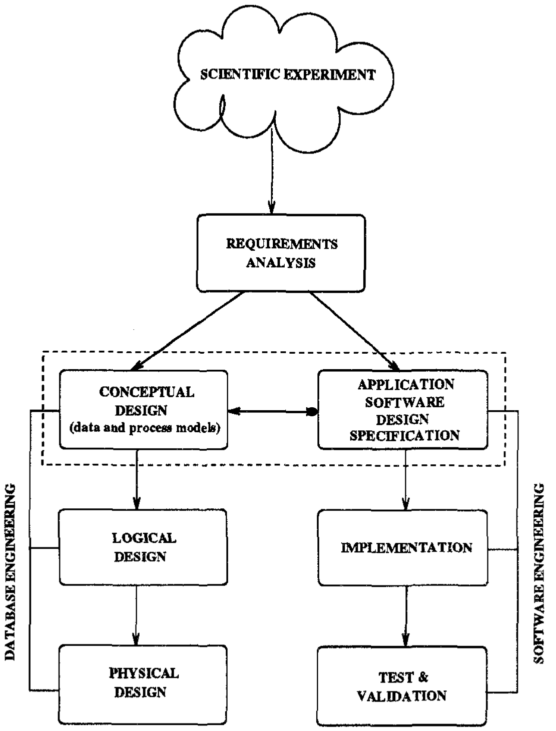

Figure 3.

The concepts of transformation processes enrich the semantics of processes and implemented algorithms, and are considered to be equivalent to the design issues of the software applications (as being addressed during the development of an information system). This knowledge is a priori definable and strongly related to the application domain (scientific experiment) under consideration.

Generation histories. They constitute the behavioral model of the information system in terms of event-condition-action triples. They are expressed by network concepts that are related to the data-derivation history of the scientific data instances [5]. Based upon these network concepts, observations or measurements from which other observations have been derived can be traced back. This knowledge is a priori definable and strongly related to the application domain (scientific experiment) under consideration.

Generation histories interrelate transformation processes with measurements and observations, in that the latter are considered to be the inputs and outputs of the transformation processes. Bringing these two knowledge structures (data and process models of the information system) into conjunction, we will be able to gain insight into the conditions under which observations have been transformed and/or generated.

Taxonomies. They are links used to organize concepts into a hierarchy or some other partial ordering [6]. We must distinguish here between taxonomies and concepts, in that the notion of a concept is primarily the notion of a data structure. Taxonomies are considered to be storing information at appropriate levels of generality and automatically making it available to more specific concepts by means of a mechanism of inheritance.

From another perspective, taxonomies provide the information to steer the extraction of new knowledge through learning methods. Considering the taxonomies that deal with the concepts of the scientific discipline (for example, trace gases), a connection point can be specified through the generation history between the concepts of the scientific experiment and discipline. This knowledge is partly a priori definable and is related to the application domain (scientific experiment and discipline) under consideration.

Experimental laws. They summarize relations among observed variables (for example, NO_y, temperature, pressure), or atomic objects (for example, trend of ozone concentration). They can be in qualitative or quantitative form and must be inductively inferenced according to the underlying data. This knowledge cannot be defined a priori. It can be regarded mainly as metadata to validate scientific data (extraction of information in terms of correct or falsified observations).

Theories/hypotheses. They represent scientific hypotheses about chemical processes in the atmosphere. They differ from experimental laws in making reference to unobservable objects or mechanisms. Scientific hypotheses are statements that belong to the empirical sciences and have this status if and only if they are falsifiable [7]. These statements are falsifiable if and only if there exists at least one potential falsifier. Thus, a logical relation exists between the scientific hypothesis and the class of potential falsifiers. This knowledge can be partly defined a priori.

Anomalies. They are experimental laws marked as potential falsifiers of a theory/hypothesis. It will be the output of a validation process of an experimental law and could demonstrably falsify a scientific hypothesis. This knowledge cannot be defined a priori.

Background knowledge. This is a set of beliefs or knowledge about the environment, aside from those that are specifically under study. It differs from theories/hypotheses or experimental laws in that the scientist holds background knowledge with relative certainty, rather than as the subject of active evaluation. Auxiliary datasets, like climatology or spectroscopic data (spectral lines of already known molecules) and/or data from other contemporary experimental campaigns, are mainly considered to form the background knowledge. This knowledge can be defined a priori.

Experiment model. They are descriptions of the environmental conditions for an experimental or observational setting, indicating the manner in which an experimental law or theory/hypothesis applies to a particular situation (specific experimental arrangement) or external conditions. Instrument and flight-related datasets give an insight of these conditions and settings.

Following these definitions, it is still difficult to distinguish between metadata and information. Therefore, we consider metadata to be data or information that is used to provide information going beyond data or to address information related source data.

Management and usage of metadata

Considering the three (3) main functions of metadata, that is, data management, data access, and data analysis, we will focus mainly on the functions of data analysis and access. Before continuing, we should mention that metadata in the form of catalogs (describing measurements and observations that have been gathered and subsequently transformed) provide rather the function of data management, especially in a distributed data management environment.

The metadata as knowledge structures, as described in the previous section, support the functions of data analysis and access. In order to illustrate this, we will also draw some parallels to the construction of a scientific paper as it is done in [8].

The knowledge structure of experimental laws and/or anomalies can be assigned to the abstract part of the scientific paper. Herein, the summary of generated datasets is presented, and the overall focus and content of the relevant datasets are mainly considered. The body of the paper can be related to the descriptions of the underlying measurement and observation data, as well as to how to transform and use them.

The description of the underlying data is addressed by the abstract concepts that describe the measurements and observations at the level of the database conceptual design (see Figure 3). The method of transforming and using data is addressed by the knowledge structures of transformation processes, generation histories, at the level of the conceptual design of the user applications (see Figure 3). Taxonomies will provide more generalized descriptions in the body of the scientific paper. The references cited in the paper are assigned the background knowledge and/or the experiment model structures. These are necessary for a full understanding of the context of the data collection and analysis.

Last but not least, the title of the paper is assigned the theory/hypothesis under consideration. The hypothesis has to be justified by accessing the relevant data sets. Thus the metadata function of data access will be provided by the corresponding structures that connect hypotheses to experimental laws and these, in turn, to the underlying data sets. We will return to this function in a later section.

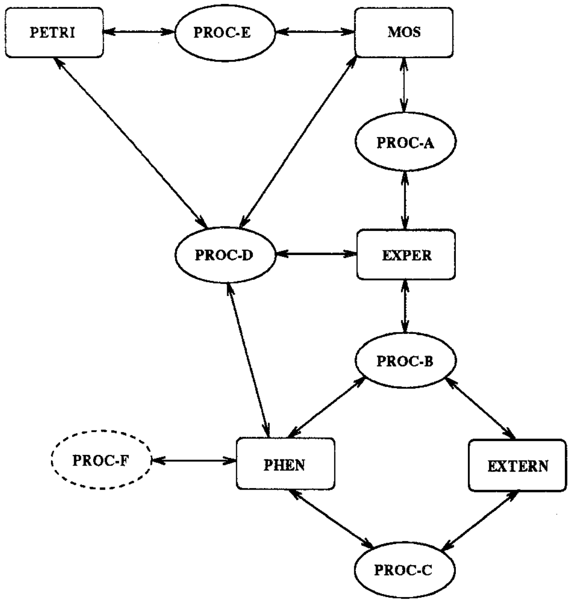

Figure 4.

Organizing metadata as knowledge servers

Metadata must be modeled as a database in its own right. Querying of metadata will have the purpose of finding which datasets are relevant to the user's interest. This mainly concerns the scientific users who are faced with the problem of data accessing according to the extracted knowledge (data analysis) provided by the system. Considering the metadata as knowledge structures to be effectively managed, a system will be favored that can supply database management facilities, as well as the accommodation of the needed knowledge-representation formalisms, as is illustrated in the following section.

At this point, we will give an insight into the metadata as knowledge structures with characteristic representation issues. We have to examine the nature of the knowledge to be addressed in order to understand what is the appropriate knowledge representation formalism to be further considered. Consequently, the appropriate knowledge servers will be determined and presented (Figure 4) which constitute the knowledge-based system for the handling of metadata.

Starting with the knowledge structures of measurements and observations, a time-series-oriented representation and addressing of data instances is mainly considered that can be sufficiently handled by a conventional relational database management system (data accessing profile is not based on complex structures). The mapping of knowledge representation to relations appears straightforward, and relational queries can naturally express queries that one would pose to the underlying representation [9].

Thus the RDBMS provides a suitable knowledge representation mechanism for this kind of knowledge structure. Catalogs that give an insight into what has been generated at the level of data instances can be sufficiently used for the data management as a function of metadata. Therefore, the RDBMS will be further considered as the knowledge server to be known as MOS (Figure 4).

The knowledge structures of transformation processes and taxonomies, as well as the concepts at the level of the conceptual database design for the measurement and observation data, push the representation formalism as provided by a conventional RDBMS to its limits ([9],[10],[11],[12]). The knowledge representation mechanism for these knowledge structures must be considered at the level of abstract conceptions that concern real-world entities and actions.

These conceptions are going to be referred as objects to be classified in categories or to change dynamically an assigned category, for instance, by modifying the taxonomy of long-living trace gases (see also scenario). The knowledge server dealing with these knowledge structures will be known as EXPER (Figure 4).

Accordingly, the modeling and derivation of the generation histories will be based on the conceptions of an extended predicate-transition network [5]. Places and transitions are also modeled as objects. Their instantiations are related to a certain algorithm version, or a certain processing path, or interactively instantiated process parameters. The corresponding knowledge server will be known as PETRI (Figure 4). The conceptions supported by the knowledge servers EXPER and PETRI are naturally represented by complex structures through which they will also be accessed. Therefore, these conceptions can be handled sufficiently by an object-oriented DBMS.

A part of the background knowledge, especially the knowledge which is referred to climatological or spectroscopic data of known molecules, will be mapped onto a separate knowledge server. On the same server, the experimental model may also be mapped and therefore, the knowledge structures that are regarded to affect the environment or conditions of the scientific experiment can be represented by the knowledge server EXTERN (Figure 4). The addressing mode of these knowledge structures must be further elaborated.

For the knowledge structures of experimental laws, anomalies, and scientific hypotheses, a more powerful representation mechanism is needed in order to accommodate conceptions such as the description of atmospheric phenomena, space-, time-, and context-dependency, incomplete data, qualitative reasoning, and so on. For this purpose, a representation formalism on the basis of a semantic network will be further considered—knowledge server PHEN (Figure 4). This decision has been taken due to the powerful expression mechanism of describing natural phenomena [13], and to the structural properties (various kinds of associations) that are provided by semantic networks.

The description of a phenomenon and/or a scientific hypothesis will be made in terms of associations which are related to certain contexts and events, whereby the context is related to the space-time association, and the event to the association of subject-predicate. These associations will constitute a propositional representation of theories/hypotheses. Although there is an equivalent representation of a semantic network in logic [14], the structural properties provided by semantic networks are more closely related to natural language representation issues [15], and can be more easily handled than logic in order to construct justification structures that will be the subject of the next subsection.

This architectural design approach provides high modularity and flexibility in terms of accommodating knowledge structures with specific knowledge representation formalisms (for instance, object-oriented and semantic network paradigms). Nevertheless, learning methods can be integrated (construction or revision of justification structures, classification, and such), whereby specific knowledge structures may be used and processed in order to steer the learning procedure.

Hence, the system also includes knowledge processors (see also Figure 4) that must be used in order to connect metadata through appropriate structures and/or modify these structures in the case of wrongly derived data sets. A short description of the knowledge processors follows.

PROC-A includes the activity of taxonomies formation and classification, PROC-B is related to the activities of inductive experimental law formation and construction of justification structures, PROC-C is related to the activity of taxonomies formation and classification, PROC-D is related to the activity of explanation process and/or revision of justification structures revision, PROC-E is related to the activity of experimental design, and finally, PROC-F is related to the activity of theory/hypothesis revision.

Connecting metadata to justification structures

Regarding the metadata functionality of data access, an appropriate structure must be provided in order to access the relevant datasets from the level of metadata in terms of extracted information about atmospheric phenomena or chemistry (scientific hypotheses). Therefore, reference structures connecting the various metadata, as defined in the previous

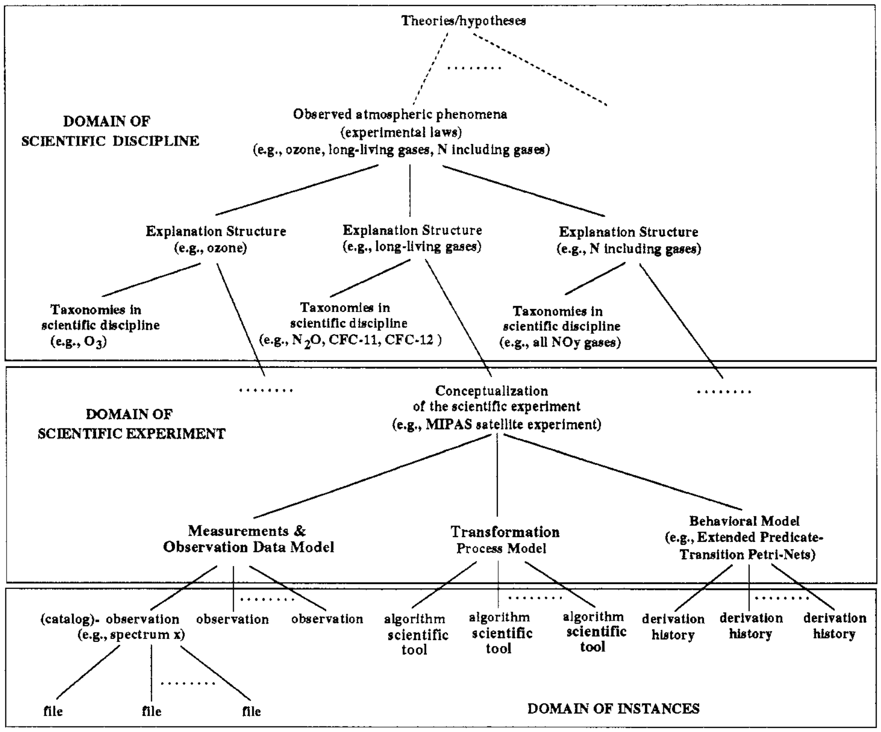
Theories/hypotheses
....
DOMAIN OF SCIENTIFIC DISCIPLINE
Observed atmospheric phenomena (experimental laws) (e.g., ozone, long-living gases, N including gases)
Explanation Structure (e.g., ozone)
Explanation Structure (e.g., long-living gases)
Explanation Structure (e.g., N including gases)
Taxonomies in scientific discipline (e.g., O_3)
Taxonomies in scientific discipline (e.g., N_2O, CFC-11, CFC-12)
Taxonomies in scientific discipline (e.g., all NOy gases)
DOMAIN OF SCIENTIFIC EXPERIMENT
Conceptualization of the scientific experiment (e.g., MIPAS satellite experiment)
Measurements & Observation Data Model
Transformation Process Model
Behavioral Model (e.g., Extended Predicate-Transition Petri-Nets)
(catalog)- observation (e.g., spectrum x)
observation
observation
algorithm scientific tool
algorithm scientific tool
algorithm scientific tool
derivation history
derivation history
derivation history
file
file
file
DOMAIN OF INSTANCES

Figure 5.

section, should enable the addressing of the underlying data instances on which a scientific result is based (justification of a scientific result). Furthermore, in case a validated anomaly occurrs, the reprocessing of wrong data instances requires an addressing facility for those data instances that are related to the anomaly specified (explanation of a phenomenon that occurred).

Definition: Justification or explanation structures are narratives that connect a theory/hypothesis to an empirical law by a chain of inferences appropriate in the field. Furthermore, they can connect empirical laws to observations related to the scientific discipline (for example, trace gases in atmospheric research), and these, in turn, to the data instances of measurements and/or observations from which high-level data have been derived.

According to this definition, justification or explanation structures can be seen as compiled knowledge [16], that is, knowledge which has been inferred and cannot be looked up at run-time. This is a critical issue when a great amount of data is going to be addressed to extract or justify knowledge. It may also be viewed as an organized knowledge that is considered to be more efficient.

Figure 5 gives the notion of justification or explanation structures. Knowledge elements from the corresponding knowledge structures representing various kinds of metadata are connected to each other in order to provide an explanation to an anomaly or justification of a scientific hypothesis.

We can distinguish mainly three layers: a) the layer of the domain instances where the time-series-oriented measurement and observation data will be managed by a RDBMS and/or the file management system of the mass storage system, b) the layer of the domain of the scientific experiment in terms of the information system's conceptualization, and c) the layer of the scientific discipline domain in terms of the scientific knowledge to be extracted and justified. The last two layers are assigned to a knowledge-based system described in the previous section.

The user interface

We distinguish mainly two categories of scientific users, that is, those who are responsible for the application programs (implementation of the various transformation processes), and those who are interested in analyzing the results and justifying scientific conclusions. For the first category, a user interface based on a query language like SQL is the most appropriate interface for the interaction between programs and data, where data independency from programs is provided. For the second category, an object-oriented interface is considered to be the most suitable interface, because scientists are very reluctant to invest time in order to learn new languages. Therefore, a combination of browsing and querying will be provided.

Using and accessing metadata necessitate a suitable user interface through which the appropriate knowledge elements as pieces of metadata can be addressed. The knowledge elements are considered to be objects that can be used independently of any particular physical or system organization. Addressing the relevant data sets will be achieved by choosing an object and the operators provided by the implemented methods assigned to the class to which the object belongs. System-dependent manipulation operators (for instance, SQL statements for accessing of the relevant datasets) will be encapsulated by these methods.

The user should be completely guided by the interface. Browsing facilities are necessary for navigation purposes. Furthermore, a hypertext-based explanation of the metadata will enhance the understanding of the underlying data and metadata. The observation of atmospheric phenomena through visualization systems is considered to be essential in order to provide an overall understanding of occurring events within the field of atmospheric research. Therefore, an information-oriented (combination of atmospheric parameters over time and geolocation coordinates on a global scale) activation of visualization systems will be addressed by the user interface, too. These issues lead to a multimedia user interface (Figure 6).

Related work

In existing systems that deal with knowledge discovery in databases ([17],[18]), the problem of efficient accessing of the source data has not been elaborated. It has been mainly concentrated on the direction of extracting knowledge and not on accessing tera(peta)bytes of source data by using the extracting knowledge for a query formulation.

In the field of scientific discovery, research on machine discovery is focused on for computational understanding of the processes that underlie scientific behavior [19]. It has been focused on empirical (quantitative and/or qualitative) discovery, taxonomy and theory formation, and generally, on the integration of manageable components in order to provide a framework for empirical discovery. But the problem remains of addressing terabytes of data through appropriate explanation structures.

Similar approaches have also been taken in ([20],[21], [22]). In [20], an intelligent multistrategy assistant for knowledge discovery from facts (INLEN) is described and illustrated by an exploratory application. It integrates a database, a knowledge base, and machine learning methods within a uniform-oriented framework. In [21], a model-based and incremental knowledge engineering is presented based on a semiformal representation that serves as a basis for communication between the knowledge engineer and the expert, through which the expert is integrated in the knowledge engineering process.

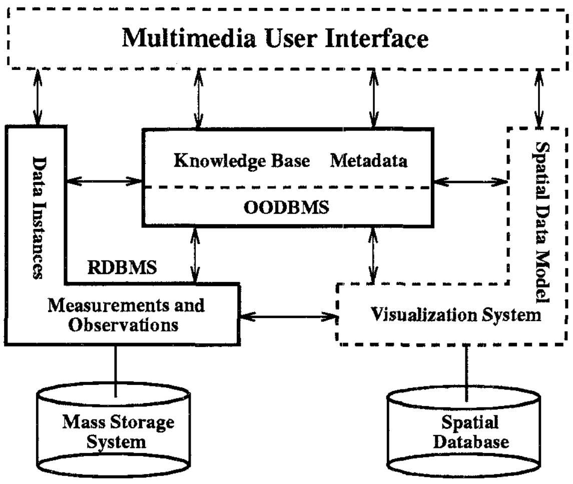

Figure 6.

In [22], the problems that arise in representing and analyzing knowledge about metabolism are described. It has given emphasis on the limits of existing techniques for qualitative reasoning, knowledge representation, machine learning, and on the challenges of building databases and knowledge bases that describe the structures and functions of engineered systems.

In the field of scientific information and database systems, the problem of data analysis has not been addressed sufficiently. It has been concentrated on methods providing a DBMS-style access to tera- or petabytes of data ([23],[24],[25]), and on the development of appropriate scientific user interfaces for accessing and extracting the requested data products ([26],[27],[28],[29]).

In particular, an Experiment Management System (EMS) ([27],[30],[31]) is currently being built to provide an integrated environment for the design and execution of scientific experiments. Within the Sequoia 2000 project [24], the visualization project Tioga addresses the question of data derivations in terms of "boxes and arrows" and is closely integrated with the underlying DBMS (Postgres). In Gaea [29], metadata has been addressed in terms of data derivation networks (Petri-Nets).

Some efforts have also been undertaken to determine which kinds of metadata must be captured for a better annotation of generated products [32], but there is no reference of how metadata are going to be managed and organized to provide explanations related to the observation data instances.

Conclusion and future work

A knowledge-based system has been presented that supports the metadata functionalities of scientific datasets analysis and/or access. Metadata will be organized as knowledge servers and will be managed by a system that provides database management facilities and the needed representation mechanisms (abstract conceptions, complex structures, inheritance, semantic networks, and so forth.). In this sense, metadata will constitute the declarative knowledge of the knowledge-based system.

Metadata as knowledge elements can also be connected to each other to provide explanations or justifications of scientific hypotheses and/or conclusions through the resultant structures. These constitute the compiled procedural knowledge of the knowledge-based system. The construction and/or modification of the justification structures will be achieved by the corresponding processor agents. Hence, a more efficient inference mechanism will be provided that can support explanations by addressing the source datasets in giga- or terabytes (in our case 1–2 terabytes of underlying datasets to be addressed by few gigabytes of metadata).

Currently, the system component for gathering and transforming scientific data is being developed on the basis of a relational database management system (time-series-related data). We have also started with the implementation of the KBMS, which will be built upon the basis of an object-oriented database management system (OODBMS). The knowledge server PETRI already exists as a prototype implemented in cooperation with the Forschungszentrum Informatik (FZI) at Karlsruhe, Germany. This mainly deals with capturing the process-related parameters as they can be initialized interactively by the scientists and thus, the derivation history (coupling of data and processes) can be addressed.

Acknowledgments

We are grateful to Prof. Dr. P.C. Lockemann for his comments on an early version of this paper, which helped to clarify our ideas and significantly improved the presentation of the final paper.

We also want to express our thanks to the scientific employees Mr. Hermann Oelhaf and Mr. Olaf Trieschmann of the Institute of Meteorology and Climate Research (IMK), at the Research Center Karlsruhe—Technology and Environment, for their time spent in fascinating discussions that we had.

References

[1] D.C. Tsichritzis and F.H. Lochovsky, *Data Base Management Systems*, Academic Press, 1977.

[2] H. Fischer, "Remote Sensing of Atmospheric Trace Constituents Using Fourier Transform Spectroscopy," *Ber. Bunsenges. Phys. Chem.*, 1991, (presented at Bunsen Meeting, Schliersee).

[3] H. Fischer, "Ozonveraenderungen in der Stratosphaere: Dynamische und chemische Prozesse," *KfK Nachrichten*, Vol. 26, No. 2, 1994, pp. 61–66.

[4] J. Schrager, ed., "Computational Models of Scientific Discovery and Theory Formation," in *Computational Approaches to Scientific Discovery*, San Mateo, CA, 1990, pp. 1–25.

[5] G. v. Bueltzingsloewen, R. Kramer, and E. Kapetanios, "On Modeling and Controlling Data Derivation in a Scientific Information System," FZI Report 3/94, Forschungszentrum Informatik (FZI), Karlsruhe, Germany, Apr. 1994.

[6] J.F. Sowa, ed., "Principles of Semantic Networks—Explorations in the Representation of Knowledge," in *Understanding Subsumption and Taxonomy: A Framework for Progress*, Morgan Kaufmann Publishers Inc., 1991, pp. 45–94.

[7] K.R. Popper, *Realism and the Aim of Science*, Hutchinson and Co. Ltd, 1983.

[8] D.E. Strebel, B.W. Meeson, and J.B. Frithsen, *Metadata Standards and Concepts for Interdisciplinary Scientific Data Systems* (in preparation).

[9] P.C. Lockemann, H.-H. Nagel, and I.M. Walter, "Databases for Knowledge Bases: Empirical Study of a Knowledge Base Management System for a Semantic Network," *Data and Knowledge Engineering*, Vol. 7, 1991, North-Holland, pp. 115–154.

[10] A. Borgida et al., *The Software Development Environment as a Knowledge Base Management System*, in J. Schmidt and C. Thanos, ed., *Foundations of Knowledge Base Management*, Springer Verlag, 1989, pp. 411–439.

[11] J. Mylopoulos et al., "Telos: Representing Knowledge about Information Systems," *ACM Transactions on Information Systems*, Vol. 8, No. 4, Oct. 1990, pp. 325–362.

[12] M. Jarke, "DAIDA—Conceptual Modeling and Knowledge-based Support of Information Systems Development Processes," *Technique et Science Informatiques*,Vol. 9, 1990.

[13] M.G. Wessells, "Kognitive Psychologie," in *Wissen und Repraesentation*, UTB fuer Wissenschaften, 1994, pp. 249–293. (Translated from the original: *Cognitive Psychology*, J. Gerstenmaier, 1982.)

[14] L.K. Schubert, "Semantic Nets Are in the Eye of the Beholder," in J.F. Sowa, ed., *Principles of Semantic Networks*, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1991, pp. 95–108.

[15] J.F. Sowa, "Toward the Expressive Power of Natural Language," in J.F. Sowa, ed., *Principles of Semantic Networks*, Morgan Kaufmann, San Mateo, CA, 1991, pp. 157–190.

[16] R. Omar, "Artificial Intelligence through Logic?" *AI Communications*, Vol. 7, Nos. 3–4, Sept.–Dec. 1994, pp. 161–174.

[17] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge Discovery in Databases: An Overview," *AI Magazine*, Vol. 13, No. 3, 1992, pp. 57–70.

[18] C. Matheus, P. Chan, and G. Piatetsky-Shapiro, "Systems for Knowledge Discovery in Databases," *IEEE Trans. Knowledge and Data Engineering*, Vol. 5, No. 6, 1993, pp. 903–913.

[19] B. Nordhausen and P. Langley, "An Integrated Framework for Empirical Discovery," *Machine Learning*, Vol. 12, 1993 pp. 17–47.

[20] R.S. Michalski, L. Kerschberg, and K.A. Kaufman, "Mining for Knowledge in Databases: The INLEN Architecture, Initial Implementation and First Results," *J. Intelligent Information Systems*, Vol. 1, 1992, pp. 85–113.

[21] S. Neubert, "Model Construction in MIKE (Model Based and Incremental Knowledge Engineering)," Bericht 277, Institut fuer Angewandte Informatik und formale Beschreibungsverfahren, Univ. Karlsruhe, June 1993.

[22] P.D. Karp and M.L. Mavrovouniotis, "Representing, Analysing, and Synthesing Biochemical Pathways," *IEEE Expert*, Apr. 1994, pp. 11–21.

[23] M. Stonebraker and J. Dozier, "Large Capacity Object Servers to Support Global Change Research," Techn. Report, Univ. California, Berkeley, Sept. 1991.

[24] M. Stonebraker, J. Frew, and J. Dozier, "The Sequoia 2000 Architecture and Implementation Strategy," Techn. Report, Univ. California, Berkeley, 1993.

[25] IEEE Technical Committee on Mass Storage Systems, IEEE Symp. Mass Storage Systems, Annecy/France, June 1994.

[26] W. Campbell et al., "Adding Intelligence to Scientific Data Management," *Computers in Physics*, Vol. 3, No. 3, May/June 1989, pp. 26–32.

[27] Y.E. Ioannidis, M. Livny, and E.M. Haber, "Graphical User Interfaces for the Management of Scientific Experiments and Data," *SIGMOD Record*, Vol. 21, No. 1, Mar. 1992, pp. 47–53.

[28] M. Stonebraker et al., "Tioga: Providing Data Management Support for Scientific Visualization Applications," Techn. Report 92/20, Univ. of California, Berkeley, CA, 1992.

[29] N. I. Hachem, M. A. Gennert, and M. O. Ward, "An Overview of the Gaea Project," *Bulletin of the Technical Committee on Data Engineering*, Vol. 6, No. 1, 1993, pp. 29–32.

[30] Y. E. Ioannidis et al., "Desktop Experiment Management," *IEEE Data Engineering*, Vol. 16, No. 1, Mar. 1993, pp.19–23.

[31] J.L. Wiener and Y.E. Ioannidis, "A Moose and a Fox Can Aid Scientists with Data Management Problems," *Proc. 4th Int'l Workshop on Database Programming Languages*, New York, Aug. 1993.

[32] F. Bretherton, "Reference Model for Metadata: A Strawman," Tech. Report, Univ. Wisconsin, 1994.