P I Assist

Document Based Archiving
and
Knowledge Management

NTT SYSTEMS INC.,
86 Dunblaine Avenue.,
Toronto, ON
CANADA
M5M 2S1
+1 (416)575-7786
(Voice and Fax)

http://www.NTT.ca
Marc.Grushcow@NTT.ca

2004-03-31

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

Table of Figures

1. Introduction

1.1 The Problem

PI Assist is a document-based Archiving and Knowledge Management System for scientific research studies. It allows for the organization and inter-relation of all the information that goes into and comes out of a research project, all in a single document.

We began work on *PI Assist* in response to a situation that developed at DRDC – Toronto[1]. *NTT Systems* had been working with DRDC since 1980 on a research program that initially focused on assessing the impact of sleep deprivation on performance. It expanded to deal with a greater variety of stressors and then to team tasks. As we approached the program's twenty-fifth anniversary, we investigated creating a CD containing a compendium of data and publications that derived from the research.

The good news was that we could locate, read, and parse all of the data. The bad news was that, particularly for the early experiments, we could not always identify sufficient experimental design information to be able to interpret the data.

In other words: *Our ability to interpret, share, and reuse data depends on our having knowledge of the context in which it was collected.*

Since we could no longer specify the design and manipulations that were in effect when the data were collected, we couldn't use them for comparative and retrospective studies. The *data's* value had depreciated to near zero.

This probably isn't an unusual situation in labs around the world. Different researchers document their work to different levels of detail and completeness. The reason we have better success interpreting more recent studies is not because researchers are more careful about documenting their work now than they used to be. It is because many of those who worked on the recent studies, particularly support staff, are still with the organization. They can remember the details of the studies, or know where to find the notebooks, and so on. This brings us to a realization:

Data persists, Knowledge decays.

But how does this affect the various parties?

1.2 The Researcher's View

When we are motivated to revisit an experiment, we are immediately faced with the issue of locating and recovering a mass of contextual information not recorded in the publications, such as:

[1] Defence Research and Development Canada, Toronto laboratory. http://www.toronto.drdc-rddc.gc.ca/home_e.html previously known as DCIEM, the Defense and Civil Institute of Environmental Medicine.

- the references that informed the work.
- the overall design (and possibly earlier abandoned approaches), including factors and hypotheses.
- how the data was manipulated prior to detailed statistical analyses including methods for handling missing values.
- anomalies that occurred during the experiment that may have affected the data.
- information required to replicate and defend the statistical results.

Since designing and running an experiment are highly collaborative activities, our ability to answer these questions depends on information provided by several groups, including collaborators and research staff. How does the Principal Investigator (PI) ensure that this is provided, stored and retained?

One might ask *why* the PI cares. There are a number of obvious reasons including professionalism, the need to defend results during peer reviews, and lab policies. These all fall into the class of things that *ought* to be done—that is, obligations to others and to the institutions for which we work.

However, there are many other reasons to care that are nearer and dearer to the scientific heart of a researcher. Efficient organization and retrieval of information will make the PI much more personally productive. Controlling the data and being able to track how it was processed and analyzed simply makes answering questions easier and faster. Retention of early design decisions, literature search results, and the overall and detailed final design all support the time efficient development of follow-on research. This is all the more true the longer the gap between the original and new work.

For the researcher, information about past experiments is necessary since it supports new research. It does this by providing useful data and analyses, and more broadly, experimental designs, hypotheses, theories, and conclusions. Note that the usual source for all but the first is normally personal notes, collective memory, and the published scientific literature.

Even when follow-up work is being done by the same researcher, it can save much time and re-treading of old ground to have the experimental design, its connections to the hypotheses and the actual data and conclusion all there in one place, along with notes about future work—all ready to be examined, manipulated or re-worked. This is particularly true if the researcher is reviving an old project that was completed ten years ago. And, of course, this is all even more pertinent if the researcher is doing a follow-up study of another PI's experiment.

In addition to the benefits of simply having one's contextual information ready-at-hand, there are further benefits when this information is pre-structured in a way that reflects scientific procedures. Such pre-structuring makes entering the data a natural act, and provides the pleasant side-benefit of helping to train new support staff and assistant researchers on the essentials of conducting good research, since they have a structured template for the kind of information they are expected to collect and maintain. There is a placeholder already there for the experimental design elements, for hypotheses, for references, for independent variables and dependent measures, and so on. Since all this is structured to reflect how a scientist works, the resulting document is not just an efficient information manager, but an integral part of the entire process of conceiving, designing and carrying out the study.

1.3 The Organization's View

Some of the PI's concerns are directly shared by the organization for which he or she works. For example, if data retention is mandated, a lab may be obliged to put appropriate procedures for this in place. However, the auditing requirement will usually be for the ability to physically or electronically deliver the data on demand; not necessarily to be able to interpret it in a way that supports current and future research. Auditing rules are rarely, if ever, designed to benefit an organization's ongoing business. Similarly, since a research organization's reputation is based on the work of its researchers, labs frequently put procedures in place that assist the analytical and publishing processes. But again, these deal with the last experiment, not the next one.

In terms of ongoing work, an organization is concerned with funding the activities of a number of researchers, many of whom share lab facilities, staff, and other resources. The more efficiently it can do this, the better. This means establishing an infrastructure that can provide services ranging from finding suitable subjects, to configuring computers and data acquisition systems, overseeing experimental sessions, collecting and cleaning raw experimental data, and so on. Typically, research assistants and related support staff do most of this work. Frequently, they design the procedures themselves and effectively create and manage the organization's "experiment running business." Efficiency is a measure of how well they organize and systematize their work. These procedures evolve over time as a result of experience and changing experimental needs, representing a significant body of knowledge that is often systematized in very informal ways including documented and undocumented procedures, mentoring, so on.

Research assistants learn from the last experiment and apply their experience to the next. Like the researchers themselves, their focus is forward. They can tell you what their data cleaning procedures are now, and likely have that information documented. But the documentation itself has evolved, and finding out what the procedures were four experiments ago might be problematic. This has serious implications for data reusability.

While research managers want to provide good research environments for ongoing work, they also have broader organizational concerns. What are the products of the research? They clearly go beyond the publications. Where is the record of all the science that never made it to publication? When researchers collaborate on an experiment, the overall design embodies all of their hypotheses. When the work is completed, a number of papers are published, each dealing with a specific aspect of the experiment. What captures the knowledge about the experiment in its totality? This includes the things that didn't work, as well as the serendipitous findings.

For research managers, information about past experiments including the data, their design, the history of their execution, and so on, all represent assets that the organization builds up over time. There is a huge investment in knowledge, both scientific and operational. How is this knowledge to be organized, protected, managed, and made available to support future work? What keeps it from "depreciating?" How does it survive personnel transitions?

1.4 The Grantor's View

Granting agencies share many concerns with the organizations that they fund. However, their focus is less on the operational side, their time scale longer, and their viewpoint crosses organizational boundaries. Historically, although they fund the research, they receive only a part of the knowledge that is created. This is typically represented by the "final report" that presents the findings. The main impact of this is that although they fund similar work in multiple organizations, there is no mechanism for each project to benefit from the detailed knowledge that each of the others generates. This effectively wastes resources due to replication, loss of synergy, and the inability to identify best practices.

For the funding agency, information about past experiments (not just results) represents knowledge that is paid for but stays with the receiving organization. This is particularly an issue where public funds are involved. How can the information be systematized and shared?

1.5 Synthesis

None of the above is meant to imply that researchers aren't concerned about the scientific or financial impact of how they do their work; or, that their organizations are only concerned with administrative procedures; or, that funders are only interested in costs and deliverables. Each of these actors is operating in their own time scale and is driven by their own priorities.

Everyone would benefit if we could have a collection of experiment descriptions, each of which:

- identified the individuals and organizations involved in the work.
- contained an overview of the experimental goals and the references that informed the new work.
- provided a detailed description of hypotheses, factors and dependent variables.
- described the subject pool in terms of relevant demographics.
- specified the experimental protocol, facilities used, task descriptions, and run histories.
- identified and described the data acquisition environment.
- allowed direct access to a variety of supporting files including raw and cleaned data, statistical summaries, and related electronic documents.
- contained full descriptions of data cleaning procedures.
- included annotated analytical results.
- gave conclusions, dealing with each of the hypotheses and the experiment as a whole.
- could be represented in a format that maximized portability and information sharing.
- could be created without adding work at any level that did not have real benefit.

Investigators would benefit from a wealth of richly documented data for retrospective analysis, and detailed experimental designs that could inform new studies.

Operational and procedural information, collected in the course of configuring, running and analyzing experiments, would meet the organization's need for capturing the knowledge that describes how the work was done.

On a more global level, the experiment descriptions in their entirety would fully document the studies conducted—something a final paper can never do—and would truly embody the organization's intellectual assets. The contextual information that they contain would provide the structure and detail needed to interpret the underlying data. Funding agencies would receive not only broad descriptions and conclusions, but also detailed, structured, and portable information that could be shared by their research organizations.

But how do we get there?

1.6 A Modest Approach

The situation we have described is commonly referred to as a *Knowledge Management* problem (note the capitals). However, if we limit our goals, we can provide a reasonable solution at a reasonable cost, with significant benefits to researchers, their organizations and their funding agencies.

Our goal in designing the current version of *PI Assist* was to produce an *archive* for a study that:

1) describes the experiment including
    a. goals and design,
    b. resources employed including their configuration,
    c. data collected along with a description of their formats,
    d. data cleaning and analysis procedures,
    e. results,
    f. conclusions.
2) is portable and easily accessible through Internet technologies.
3) can be interpreted using standard tools.
4) is produced over the course of creating, running and interpreting the experiment.
5) should not require significant extra effort on the part of those involved.
6) provides significant benefits for those who create the archive, as well as others.

1.7 The Solution

To reach these goals, we realized that we had to focus less on the data itself, which is easy to organize, and more on the metadata—information about the data. This meant designing a rich and highly structured document template for the metadata, one that reflects the way researchers actually do science. In a sense, the scientific method itself would need to be embedded in the structure of the template, to guide authors through the document creation process, and make it part and parcel of the process of doing their work.

To maximize support for sharing experiment descriptions between collaborators and institutions, we decided to use XML, the industry standard for structured documents and data sharing.

While the document template reflects the *process* of doing science, the final paper represents the *communication* of the result. While this communicated result may have a somewhat different form than

the study itself, the connection between the two is still important. In addition, while each individual researcher may understand the information they are entering at the time they enter it, the entire sum of all the information in the document still needs to be communicated in a readable way, both to external researchers and to the various individuals, each working on their own piece of the same study. To this end, we recognized the need for an automatically generated manuscript-like view of the structured metadata. This would provide, not a final paper, but a way of communicating the information in the document *in a readable way*, and in a way that might provide a starting point for the creation of a final paper.

In addition, we realized that by combining experiment description with data organization, within a single document, a research project gets archiving and document version control "for free". At certain intervals throughout a project, the ongoing study documents can be archived away to storage, and at the end of the project, we have not only a complete record of the study, but a record of what stage it was in throughout its life cycle, from conception to publication.

1.8 Benefits

The *PI Assist* document template solicits and organizes critical operational and scientific knowledge, thus capturing context, intent, process and results.

Because there will always be information that it is not practical to physically store in the study document, *PI Assist* has the ability to reference external data files, papers, and so on, so they may all be retrieved and viewed from the study document, either directly within *PI Assist* or in a standard Web browser.

Both data and metadata can be easily transferred to archival storage.

Because the document is in XML, any number of reports could in theory be created, and different information could be extracted for different audiences. The information could be reorganized and reformatted for different media and publication styles. While there is currently no built-in support for this level of customizability, the use of XML ensures that this feature can be added in a future version of the software with little change to what is already there. Thus XML gives the software the room it needs to grow in capability and flexibility in the future.

1.9 The Program

We will take a very brief look at the main aspects of *PI Assist* in this section—for a more detailed look, see section 2 for a summary of all the features, and section 3 for a brief walk through some selected areas of the program, to give you a feel for how to use it. For more detailed information, see the Online Help facility that comes with the program.

The main *PI Assist* window (Figure 1) has three areas, two of which are Views of your Study document (one for navigating and the other for presentation), and the other is an editor for entering and editing the data contained in the document:

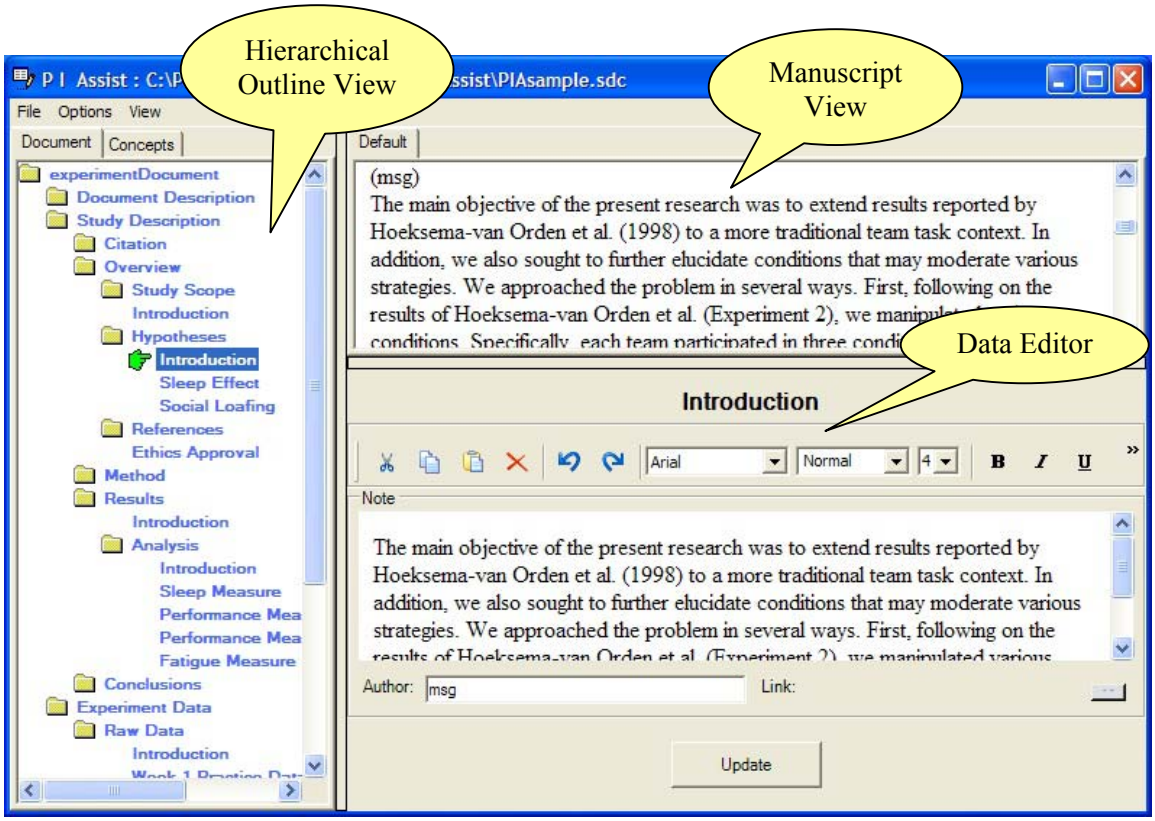

Figure 1. *The Main Window: Hierarchy/Manuscript Views and Data Editor.*

Outline View: Positioned at the left of the main window, this is a standard hierarchical outline view of the document's structure, consisting of main headings, subheading, and so on. Each heading can be expanded to view its internal structure, or collapsed to hide it. As you navigate down into this hierarchy, you will reach headings that do not expand—these are Data Headings, which contain the actual data. None of the actual data is presented for viewing in the outline view, however—only the headings appear here.

Manuscript View: This view (in the upper-right of the window) presents the information contained in the document in a readable manuscript form. It is not intended to be a final manuscript or paper, rather it is simply a view of your study presented in roughly that form. It may, however, be saved to HTML and opened in a word processor to serve as the starting point for a final paper.

Data Editor: Whenever a Data Heading is selected in the Outline View, its corresponding editor appears in the lower right of the main window. There are many different data editors, each one tailored to the kind of data expected at the corresponding heading.

Concept View: This is an alternative to the Hierarchy View, also displayed on the left-hand side of the window. Choose between the two by clicking on the corresponding tabs at the top of the View. This View relates to the Study Document and the other Views in much the same way as the Hierarchy View, except that it is flat (non-hierarchical) and organized alphabetically. It thus functions has a handy index.

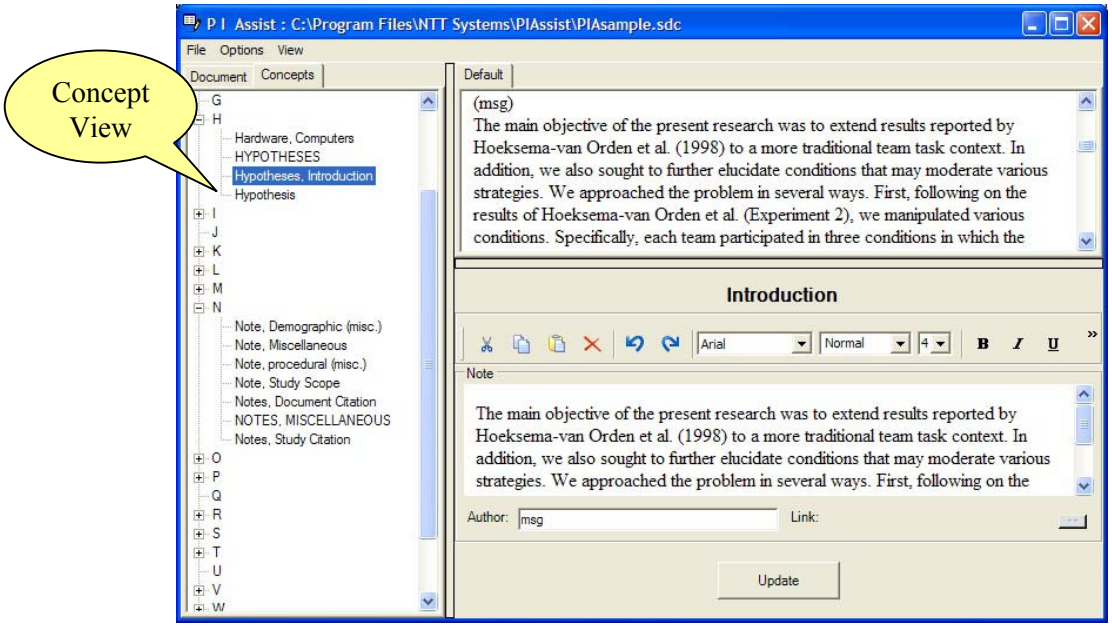

Figure 2. *The Concept View.*

2. Summary of Features

2.1 Main Document Sections

The four main headings at the top level in the hierarchy are as follows:

Document Description: This provides an overall description of the Study, its authors, how to cite it, and so on. This heading is not meant to contain any detailed information about the Study.

Study Description: This provides the description of the Study itself, including its authors, how to cite it, an abstract, the Study's method, results and conclusions. This heading is not meant to contain any of the actual experimental data.

Experiment Data: This provides the actual data, with documentation, that resulted from the experiments. There are separate sections for the data itself, documentation on missing data and the process and result of cleaning the data (preparing it for analysis). This section is not meant to contain any information on analysis or conclusions.

Other Material: Miscellaneous other documentation relevant to the Study, such as a record of experiment logs, references and other notes.

2.2 Detailed Overview

Following is a more detailed summary of each area in *PI Assist.* Note that throughout these sections, there are additional "Notes" areas, not listed below, that allow the insertion of textual notes, along with the author's name and a link to an external file.

Document Description:

Citation: How to cite this study document (as opposed to the study itself).

Study Description

Citation: How to cite the study, and related information. No detailed information about the study is intended to go here.

Overview: A broad overview of the study, including an abstract, statements of the experimental hypotheses and relevant bibliographic references. The hypotheses are the extent to which the method is explained under this heading. More detailed information on method goes under the next heading.

Study Scope: Broad overview of the entire study.

Abstract: Summary of the study, such as would appear in the abstract of a paper.

Subject Information: Topic words and keywords for searching purposes.

Summary Data Description: A summary of the data collected in the study. Since more than one type of data can be collected, you have the ability to add multiple Summary Data Descriptions. The Editor has four tabbed subsections:

General: Specifies the time period covered by the data and when it was collected, the geographic area (if any) covered, and the units used.

Clusion: Specifies populations included and excluded from the data.

Data Types: Specifies the type of data collected (clinical, survey, experimental, etc.).

Hypotheses: A statement of each hypothesis being tested in the study.

References: A complete list of references cited in and relevant to the study.

Ethics Approval: Information, with link to an external ethics file, on the ethics approval status of the study (if relevant).

Method: A detailed explanation of the method used, including descriptions of any experimental apparatus, a statement of overall experimental design (independent variables and such), and a precise description of the experimental procedures. No data or results are meant to go here.

Subjects: A list of subjects (or other data sources), with a description of demographics and exclusions (types of individuals or data sources excluded from the study). The Editor has three tabbed subsections:

Subjects: a database list of the individual subjects.

Demographics: a description of the subjects' demographics.

Exclusion: a description of subjects or subject types excluded from the study.

Apparatus: Lists and descriptions of equipment used in the study.

Facilities: Facilities used. Any equipment that does not fit into the more specific categories below may be described here.

Computers and Software: Computers used, with a list of software for each.

Data Acquisition: Data Acquisitions devices, with descriptions and specifications.

Design: A specification of the experimental design.

Independent Variables: description of the experimentally manipulated (independent) variables or subject groups.

Within Factors: describes manipulated variables (factors) within an experimental group of subjects.

Between Factors: describes factors manipulated by differences between groups of subjects.

References: bibliographic references relevant to the experimental design.

Results: A detailed statement of all the results of the experiment. The actual data from the experiment is not placed here—only analysis of the data belongs here. Conclusions drawn from these results are held back for the next heading.

Analysis – dependent measures: description and analysis, with links to relevant external documents, of the dependent variables (measures)—the variables which our independent variables were designed to (possibly) influence in some way, as indicated in the hypotheses.

Conclusions: The experimenters' conclusions, given the hypotheses, data and experimental results, are recorded here.

Experiment Data: This section simply provides the data—it does not analyze it nor does it draw conclusions. However, it does deals with preparing the data for analysis. Normally, the actual data will be in an external file and will only be linked to, not actually pasted into the *PI Assist* document itself.

Raw Data: a link to the raw data itself, with no massaging or modification.

Missing Data: a description, with link to relevant external file, of the missing data.

Data Cleaning: a description of any data cleaning procedures that were applied to prepare the data for analysis.

Clean Data: a link to the actual cleaned data that was used for analysis.

Other Material: Various other miscellaneous information and notes.
Experiment Logs: Daily (or otherwise) logs, with dates and authors noted.
Operational Considerations
References

3. A Brief Tour

3.1 Getting Started

PI Assist comes with a sample Study Document. We'll walk you through enough of it to give you a sense for what the software can do and how to use it. You can explore the remainder of the sample Study yourself, using the more complete information about *PI Assist*'s workings given in the Online Help that comes with the program.

This sample Study is based on an actual research project, but with many of the details removed and replaced with more generic descriptions. The file thus does not in its current state represent any real research project, although it is loosely based on one—its purpose is solely to help you get acquainted with the uses and capabilities of *PI Assist*.

Start *PI Assist* by choosing "Start→Programs→PI Assist" in the Taskbar. This will bring up *PI Assist*'s Main Window:

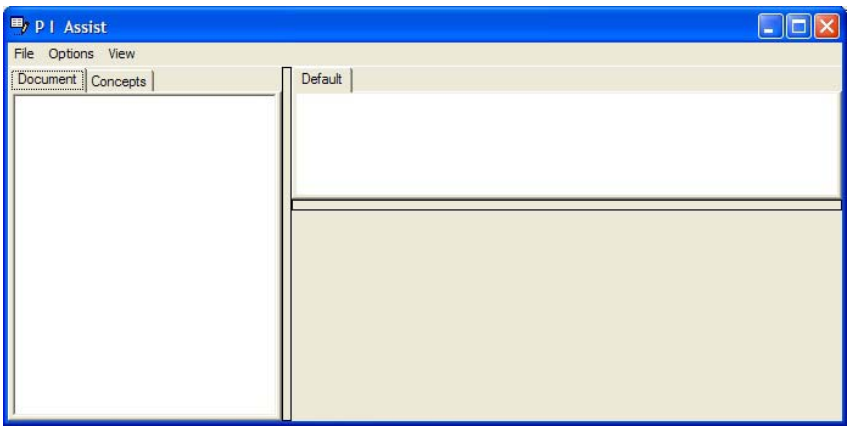

Figure 3. *P I Assist's Main Window*

Choose "File→Open" in the Main Menu:

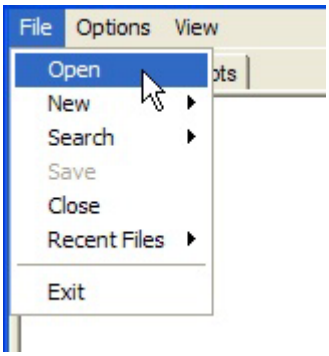

Figure 4. *Opening an Existing Document*

This will bring up the *Open Dialog* box. Select "PIAsample.sdc" to open the sample study:

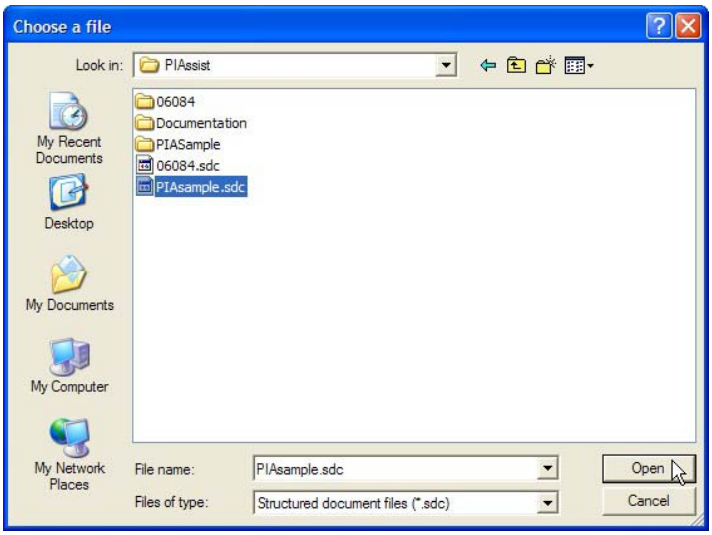

Figure 5. *Selecting a Document to Open*

The Main Window now displays the sample document, consisting of a Hierarchical Outline view of the Study Document on the left, and a more detailed and readable Manuscript View in the top right:

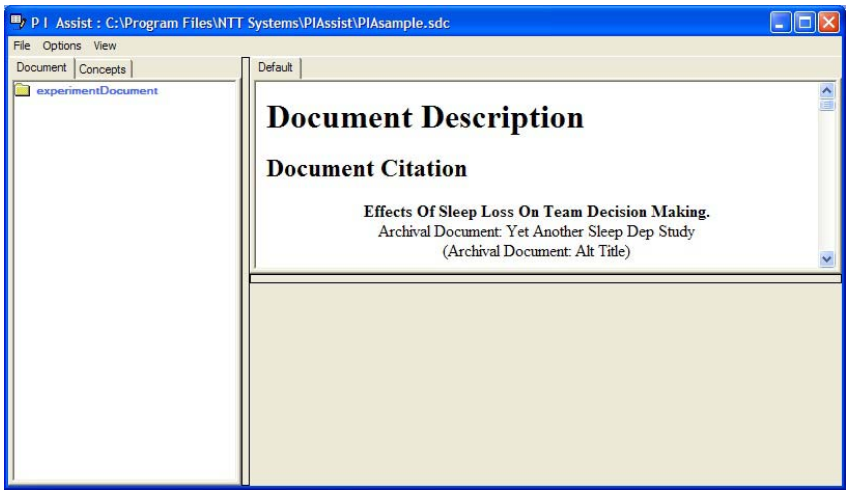

Figure 6. *Initial View of an Existing Document*

3.2 The Four Views

The Manuscript View is not meant to be a publishable manuscript or final paper for the Study, although it presents itself in roughly that form. It can be saved as HTML and opened in a Word processor, possibly as a starting point for the editing of a final paper.

The Hierarchical Outline View currently has only one heading showing: “experimentDocument”. Double-click it to open it and reveal the subheadings that lie within. Some headings in the Outline View will have a picture of a folder on the left-hand side. These "Folder Headings" can be opened to reveal further subheadings (as opposed to containing actual data). A Folder Heading is opened by double-clicking on it, and is closed in the same way. A heading without a folder is a "Data Heading", meaning that actual information is stored within it.

The currently selected heading always has a hand icon () pointing to it. Any Data or Folder Heading that has already had information added to it is displayed in blue. Headings that have never had any information added are in black. Just remember that “black” means “empty”.

Navigate to “experimentDocument→Experiment Data→Missing Data→Introduction”. The currently selected item is now an "Introduction" subheading within "Missing Data", which in turn is within "Experiment Data". The "File Description" heading under "Missing Data" is empty--no information has been entered for it yet. All the other headings are in blue, so they have had information added already (see Figure 7).

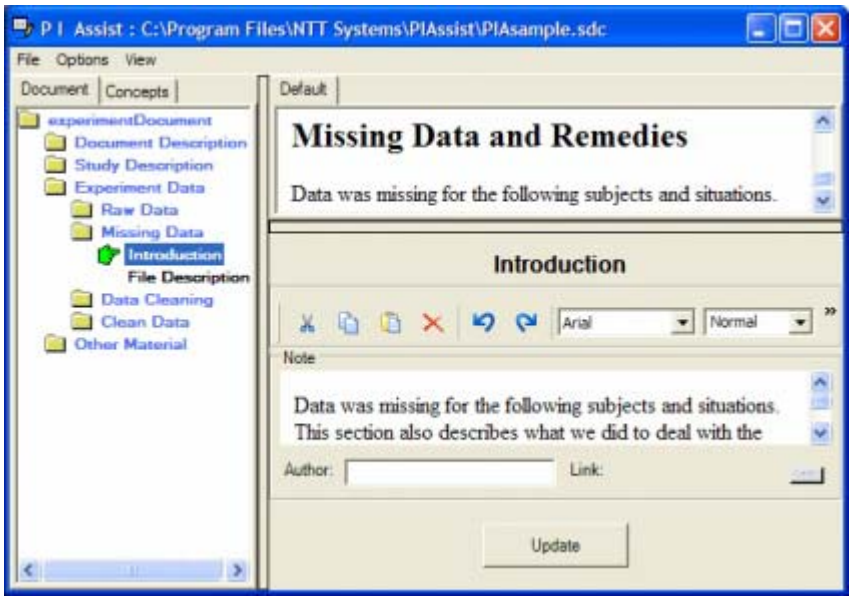

Figure 7. *Initial View of an Existing Document*

The Data Editor corresponding to a Data Heading will always be displayed in the lower right of the *PI Assist* window (if a Folder Heading is selected, this portion of the window will be blank). The Data Editor allows you to add or modify the data stored under a Data Heading. There are numerous different types of Data Editors, as they vary with the kind of data being edited—from simple text editors to more complex database entry controls and so on. The editor in Figure 7 is one of the simplest, as it is not much more than a text box, with the usual formatting tools, such as you might find in a word processor. There is, in addition, a text box for "Author" (the name of the person entering this information) and a "Link" button that allows you to link to an external file related to this heading.

Clicking on "Link" (Link: ...) will pop up a file dialog window, allowing you to indicate an external file that is relevant to the data you are entering. Once you have entered your link, it will appear to the left of the Link button, hyperlinked, as in "Link: example.txt ...". Experiment with this in the Introduction Editor, if you like. The external file can be changed by clicking on the Link button again, and specifying a different file, or clicking "Cancel" to just delete the existing link.

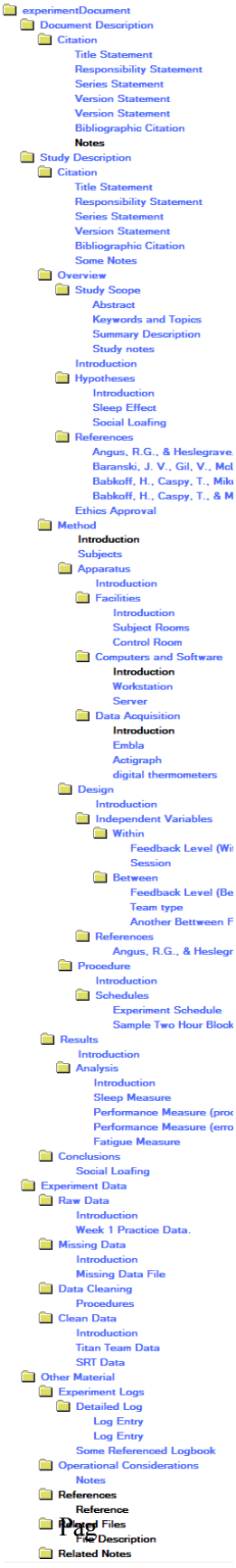

Figure 8. *The Hierarchy.*

Make sure when you use "Link" that you understand exactly how it works. When you click the button you will be given a choice between opening an existing external document and creating a new one. If you open an existing one, it will be *copied* into your Study Document space, and it is this copy that is referred to henceforth in your document, no the original. If you create a new file, then the copy in your Study Document space will be the only copy of the file there is.

If you like, type something into the main textbox in the Introduction editor—type anything you like, such as "Introductory comments on method go here." Then, click the "Update" button at the bottom of the window to save this change into your open *PI Assist* document. This *only* saves your changes within the opened document, you must still select "File→Save" from the Main Menu to save the modified open document out to disk. Every Data Editor has an Update button at the bottom, but in all other respects Data Editors can vary one from the other as greatly as the different kinds of data they are meant to handle.

We will look at some of the other common editors later on (but see the Online Help that comes with *PI Assist* for complete information about all the Data Editors).

The Alphabetical Concept View can be chosen by clicking on the "Concepts" tab in the upper left of the window. This is a non-hierarchical View that operates much like the Outline View, except that it has a flat, alphabetical structure. It initially displays one Alphabetical Heading for each letter of the alphabet.

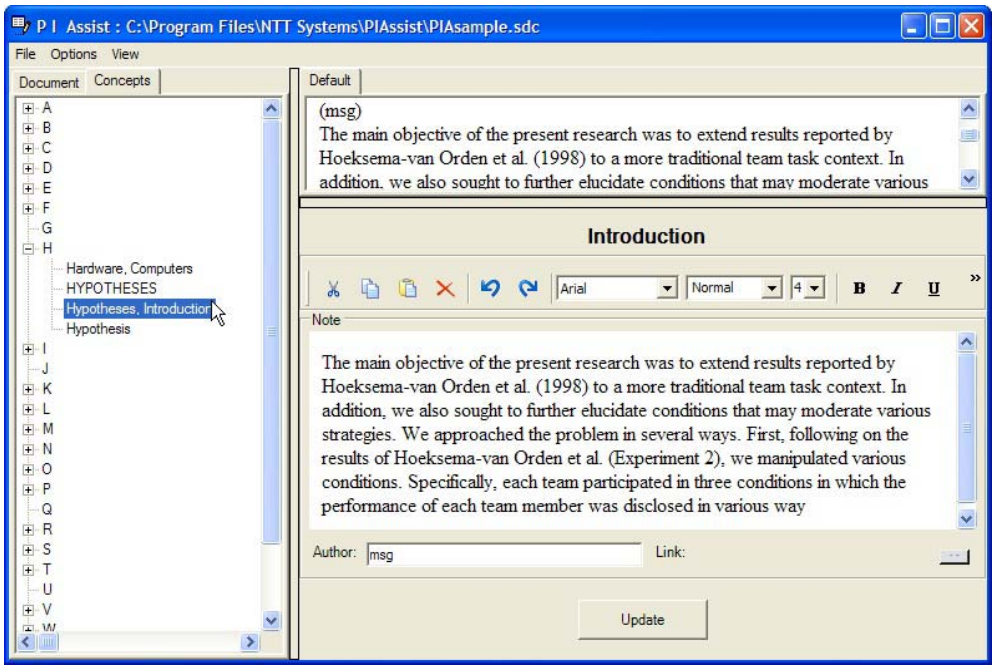

Figure 9. *Selecting a Concept in the Concept View.*

Each Alphabetical Heading can be opened by clicking on the "+" to its left, which will display the Concept Headings for that letter. The "+" will change to a "-", which may be clicked to close the heading again. Click the "H" heading (Figure 9) to reveal four Concepts beginning with the letter "H", each corresponding to a heading in the Outline View. Now select one these, "Hypotheses, Introduction", to bring up the same Manuscript View and Data Editor as the corresponding heading in the Outline View. Note that one of the four Concepts is in All-Capitals; this indicates that the Concept corresponds to a Folder in the Outline View, rather than a Data Heading.

To navigate immediately to the Outline View heading that corresponds to "Hypotheses, Introduction", just double-click on the heading and you are there (Figure 10). Note that the Manuscript View and Data Editor have remained unchanged. You are in the exact same place in the Study Document after double-clicking, having simply moved from the Alphabetical Concept View to the Hierarchical Outline View. You can also switch views simply by clicking on the appropriate tab in the upper left of the View.

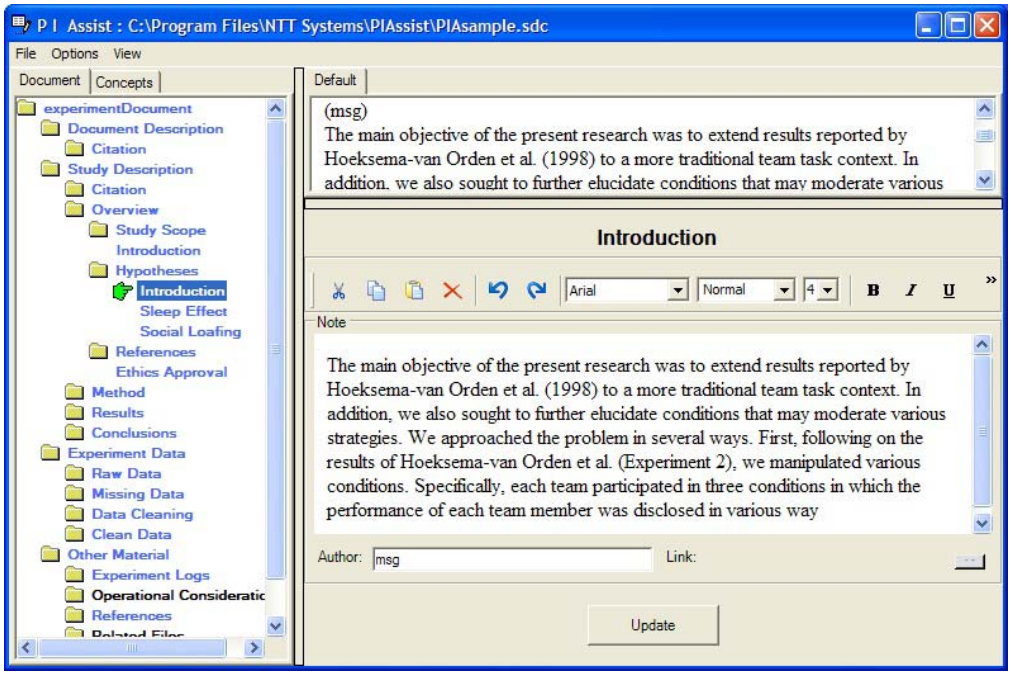

Figure 10. *Navigating From the Concept View to the Outline View.*

3.3 Adding and Deleting Data Headings

Next, let's take a look at how data items such as references are added to the Outline View. Open "experimentDocument→Study Description→Overview→References" to see the list of references for the study.

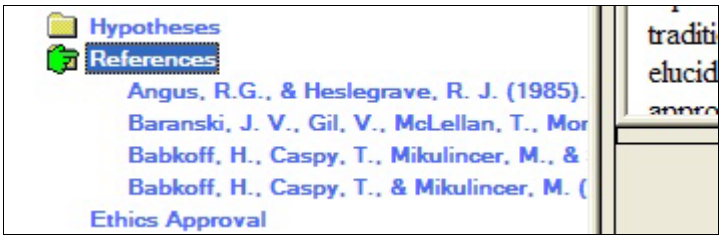

Figure 11. *References in the Outline View.*

Note that in the Outline View, the References are not displayed with a generic label like "Reference", but with the actual text of the individual references (however, the generic "Reference" label *will* appear if there have not yet been any references added).

Select "experimentDocument→Study Description→Overview→References→<Reference>" to view the Reference Editor for one of the References, where <Reference> is the text of the reference. The Reference Editor provides a textbox for entering the reference and a Link button.

To add a new Reference, right-click on any existing Reference (or the empty "Reference" heading if this is the first reference to be added), and choose "Add". Or right-click an unwanted reference and choose "Delete" to get rid of a reference.

Some Data Headings, like References, allow you to add as many headings of that type as you like. Others allow you to add only one. For instance, navigate to "experimentDocument→Study Description→Citation→Series Statement". You will notice that you have clicked on a black data-less heading, and that therefore no Data Editor has appeared in the lower right of the screen—telling you that there is nothing here. Try adding a Series Statement at this point by right-clicking on "Series Statement" and choosing "Add" (note that "Delete" is grayed out, since there is no data yet that can be deleted).

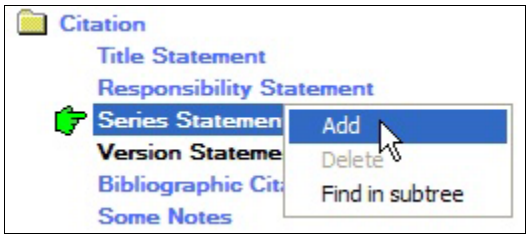

Figure 12. *Adding a New Data Item.*

Try adding some information into these fields (it does not matter what you enter, since we are going to delete the information momentarily, anyway). Now click Update, and note that the

Series Statement heading in the Outline View now appears in blue, since it has data in it. Now right-click on the heading again, and you will see that you now can Delete the Series Statement, and it is "Add" that is grayed out. Go ahead and "Delete" the information that you just entered.

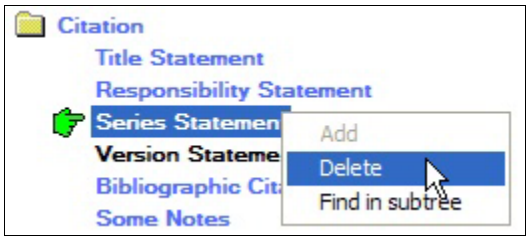

Figure 13. *Deleting a Data Item.*

You will now see that "Series Statement" has reverted to black. Note that we were unable to add additional Series Statements, since this type of data is not suited to multiple instances.

3.4 Subjects

Next, let's take a look at the "Subjects" Data heading, which stores information about the experimental subjects. Choose "experimentDocument→Study Description→Method→Subjects" to bring up the Subjects Editor.

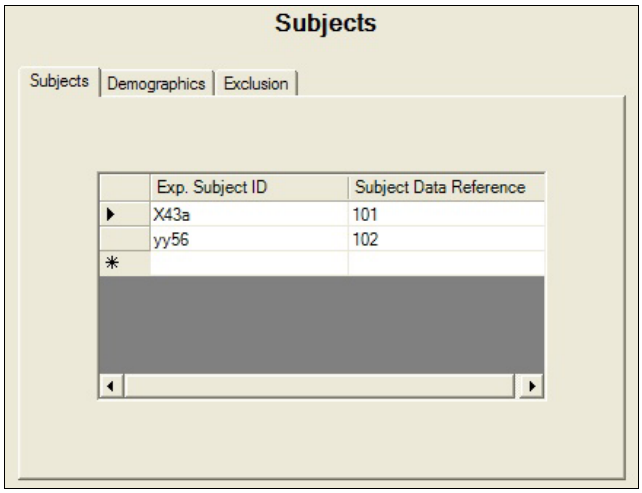

Figure 14. *Subjects Editor—Using the Database Control.*

This is a tabbed editor, with three separate sections. The first section—which should be open right now—is for the actual list of subjects, and it is this section we will now look at in more detail. The remaining

two sections are for a description of the demographics of those subjects included in the study, and a description of those types that were excluded—these last two sections are straightforward Notes Editors, so we will not examine them further here.

The Database Control used to enter the subject information is standard, and you may have used similar controls in other programs. You can see that each line is one record of data—in this case a person—with columns for the fields within each record—in this case, the Name and Affiliation and so on. To enter a new record, click in the last line, which is a special line reserved for entering new records (and always tagged by an asterisk "*" on the left). Select a line (*i.e.,* a record) by clicking on it. The currently selected record is tagged on the left, with an arrow "▸" symbol. This selected line may be modified by simply typing in the fields. While the data entry record in the last line is being modified to create a new record, the "*" will temporarily change to "▸".

Again, as with all Data Editors, there is an Update button at the bottom to save your changes into the open document.

3.5 Factor Editors

Next, we'll take a look at the Within Factor Editor. Open up "experimentDocument→Study Description→Method→Design→Independent Variables→Within". You will see a number of Factors. Choose one of them to see the Within Factor Editor. Note that there can be more than one Within Factor. If you right-click on "Factor", you can choose to "Delete" an existing one, or "Add" a new one. Numerous Data Headings work like this, while others allow only one heading of that type to exist.

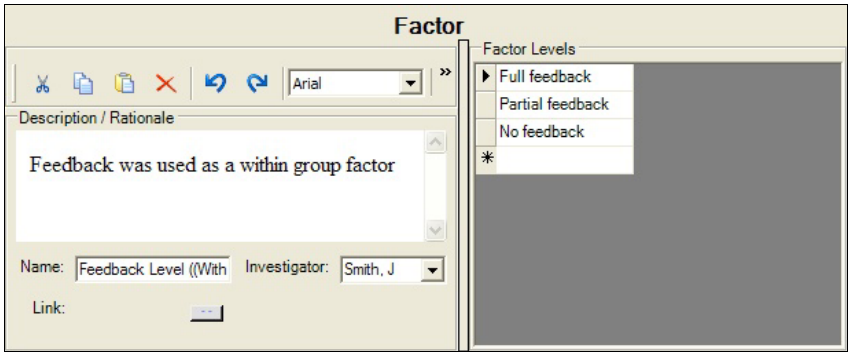

Figure 15. *Within Factor Editor.*

PI Assist has an area for both Independent Variables, or Factors (the variables manipulated by the experimenter) and Dependent Variables, or Measures (the measurements taken which may, as

hypothesized, be affected by the manipulated variables). The Dependent Measures are, naturally, in the Results section, while the Independent Variables are here, in the Method section of *PI Assist*.

There are two kinds of Independent Variables: those directly manipulated by the experimenter within each group of subjects—the "within group" factors—and those indirectly manipulated by virtue of how the subjects are grouped in the first place—the "between groups" factors.

As you can see, the Within Factor Editor has a textbox in which the factor is described, another for the name of the factor, and one for the name of the investigator. It also has a Link button for an external file. In addition, it has a database control for entering the various levels at which this factor is measured. For example, here we see the "feedback level" factor, which tells how much feedback was allowed, measured in three gradations: "no feedback", "partial feedback" or "full feedback".

Note that the Between Factor Editor is very similar to the Within Factor Editor.

3.6 Summary Data Description

Next, we will look at the Summary Data Description Editor. Choose "experimentDocument→Study Description→Overview→Scope→Summary Data Description" to see this Editor. Here, you will enter a summary of the kind of data that was (or will be) collected during the study. Since more than one type of data can be collected, you have the ability to add additional Summary Data Descriptions.

This Editor has four tabbed sections. Click on "General" to see the first section, if it is not already showing. This is where you specify the time period covered by the data and when it was collected, the geographic area (if any) covered, and the units of measure used.

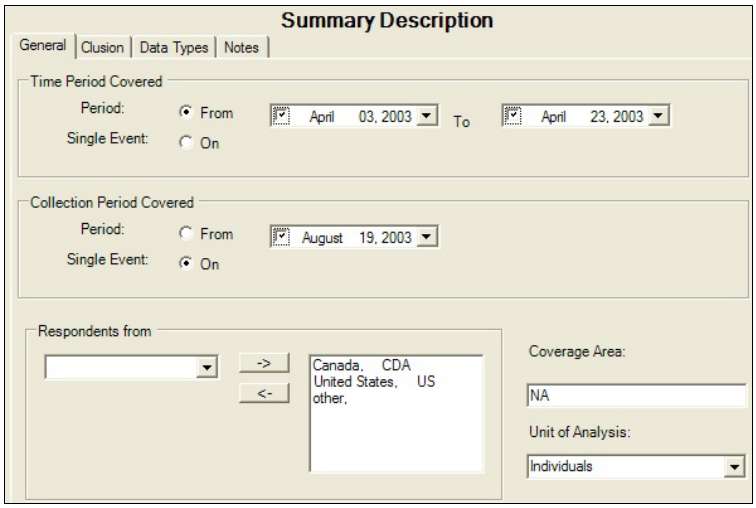

Figure 16. *Summary Data Description Editor—General Section.*

There are a total of three individual dates to enter (two for a date range and the other for a single event). Each date can be simply entered as text, if you wish, using most popular date formats. But to avoid entering an unrecognizable format, you may prefer to click on the "x" button to the right of the date, which will provide a pop-up window with a drop-down menu. Click on this menu and you will get a calendar for the month of the date currently entered, or for today's date if no date has been entered yet.

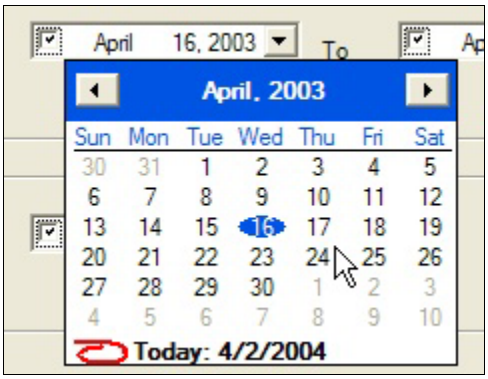

Figure 17. *Selecting a Date with the Calendar.*

The right and left arrows at the top of the calendar allow you to navigate to a different month. Or, you may choose a month by clicking on the name of the month and choosing from the resulting drop-down menu:

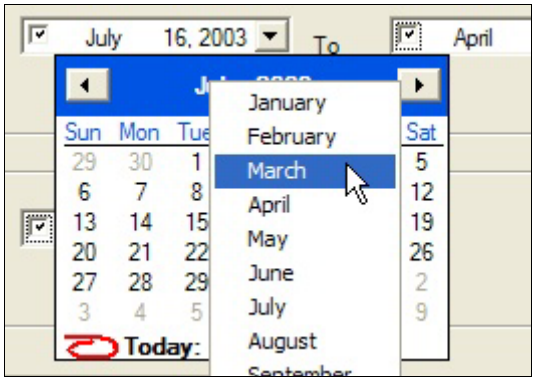

Figure 18. *Selecting a Month in the Calendar.*

To navigate the calendar by year, click on the year at the top of the calendar, and up and down buttons will be provided to adjust the year.

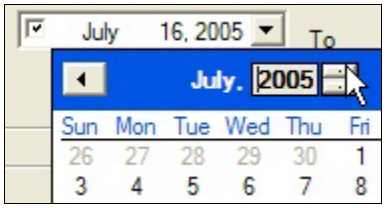

Figure 19. *Selecting a Year in the Calendar.*

The lower part of the editor provides a combo-box for entering the various geographic areas from which the data was collected (labeled "Respondents from"). Choose those areas that apply to your data with the drop-down menu, or just type in the area if it is not in the list. Click the right arrow "→" to move this selected area into the list of areas. The general area covered by all these individual areas can be entered in the "Coverage Area" textbox. Note that if your data is coming from various sources, and there is another way to split these into natural groups, then your "areas" need not necessarily be geographical in nature.

Finally, there is a combo-box that allows you to enter the units in which your data is expressed.

That's it for the "General" section. Now click on "Clusion" to see the next section, which is simply two textboxes, one for describing segments of the population included in the study, and the other to describe segments excluded.

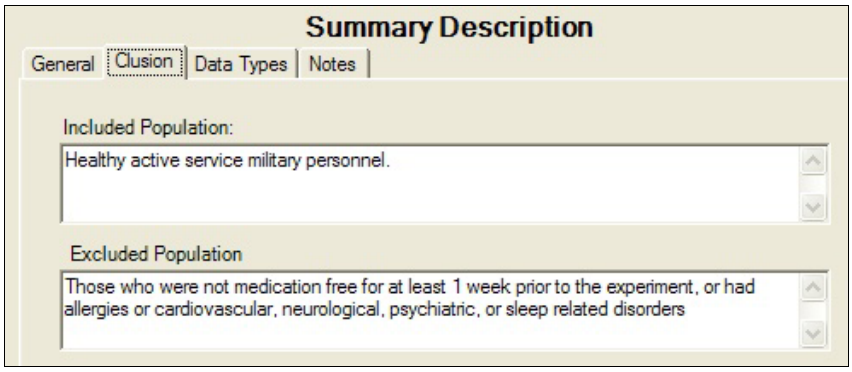

Figure 20. *Summary Data Description Editor—Clusion Section.*

The third section is for "Data Types". It provides a simple checklist, on the left side of the window, which allows you to check all the popular data types that apply to your data. An additional textbox to the right allows you to describe other data types that might not fit those listed on the left.

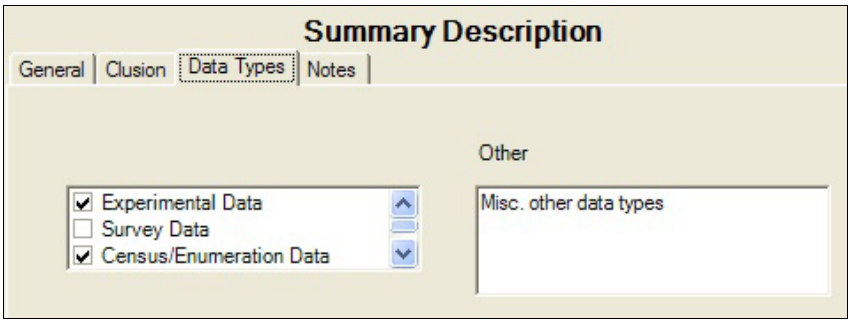

Figure 21. *Summary Data Description Editor—Data Types Section.*

The fourth section of the Summary Data Description Editor is a standard Notes Editor.

3.7 Find

A *PI Assist* document can be searched for keywords using the "Find in subtree…" feature. Right-click on any heading in the Hierarchy View to search everything under that heading. For instance, in Figure 22 the "Overview" heading is right-clicked to search everything within it for the keyword "performance".

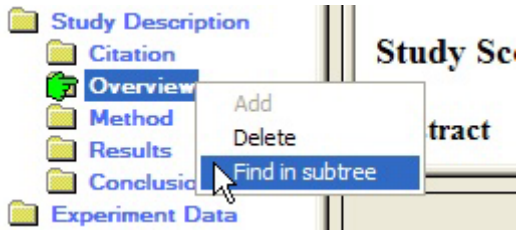

Figure 22. *Find in Subtree…*

You will then be asked to enter the string of characters to search for. In Figure 23, the keyword "performance" is being searched for. Click the "Next" button to begin the search.

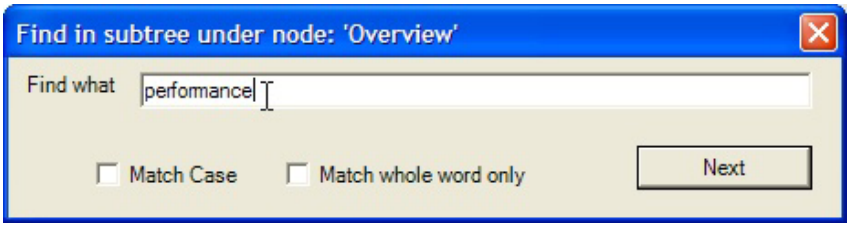

Figure 23. *Searching for a Keyword.*

Figure 24 shows first result of the above search… the keyword "performance" is highlighted right in the Data Editor where it first occurs.

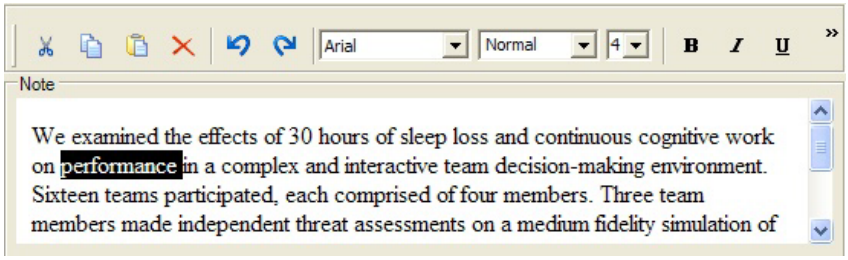

Figure 24. *Finding the Keyword.*

Click again on the "Next" button (Figure 23) to find the next occurrence of the keyword, and so on to find all further occurrences. The "Next" button will be grayed out when there are no further occurrences of the search string.

4. The *PI Assist* Database Search Facility

This section documents the Database Search Facility, a feature not yet available in any released version of *PI Assist,* but which is currently under development. It is completely separate from the existing *PI Assist* "Find…" feature, discussed earlier, which searches within the open *PI Assist* document. The Database Facility searches through multiple archived *PI Assist* documents.

To bring up the Database Search Facility, select "Search→Experiment Description".

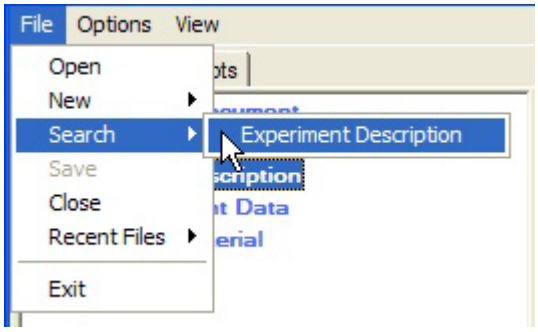

Figure 25. *Invoking the Database Search Facility.*

The Search Facility is a modified version of the regular *PI Assist* window. On the left is the Hierarchy View; on the bottom right is the Data Editor, if any (at first, this will be blank, since no Data Heading in the Hierarchy View has yet been selected).

The major difference from the usual *PI Assist* window can be seen in the top right of the window, which is no longer the Manuscript View, but is now dedicated to listing various options for the search terms.

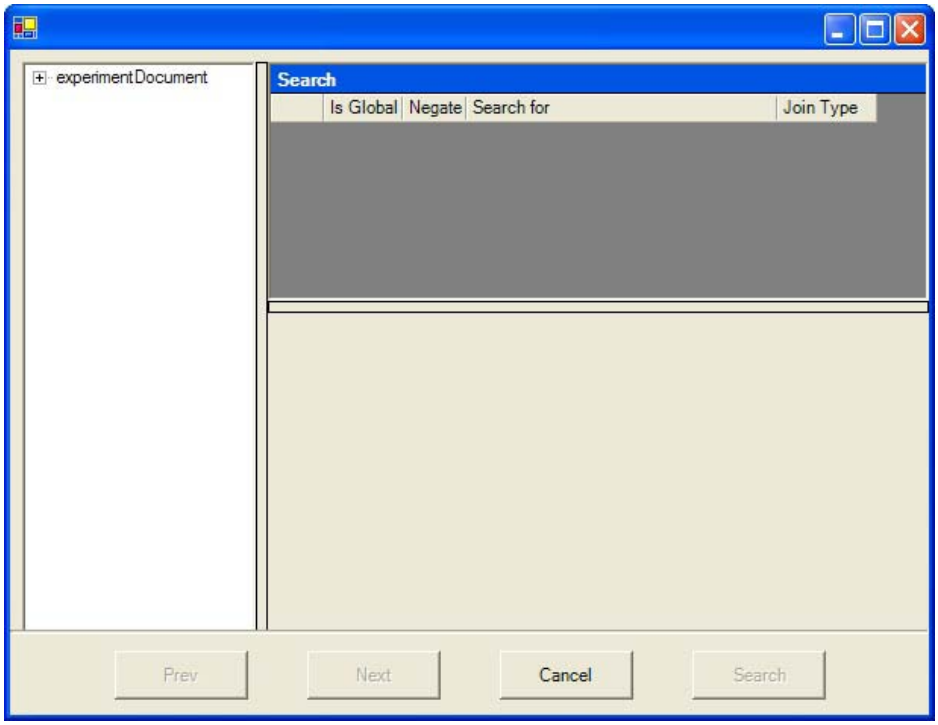

Figure 26. *The Database Search Form.*

To search for a string, first navigate to an appropriate Data Editor in the Hierarchy View. In Figure 28, we have clicked "experimentDocument→Study Description→Overview→Introduction", bringing up the corresponding "Introduction" Editor.

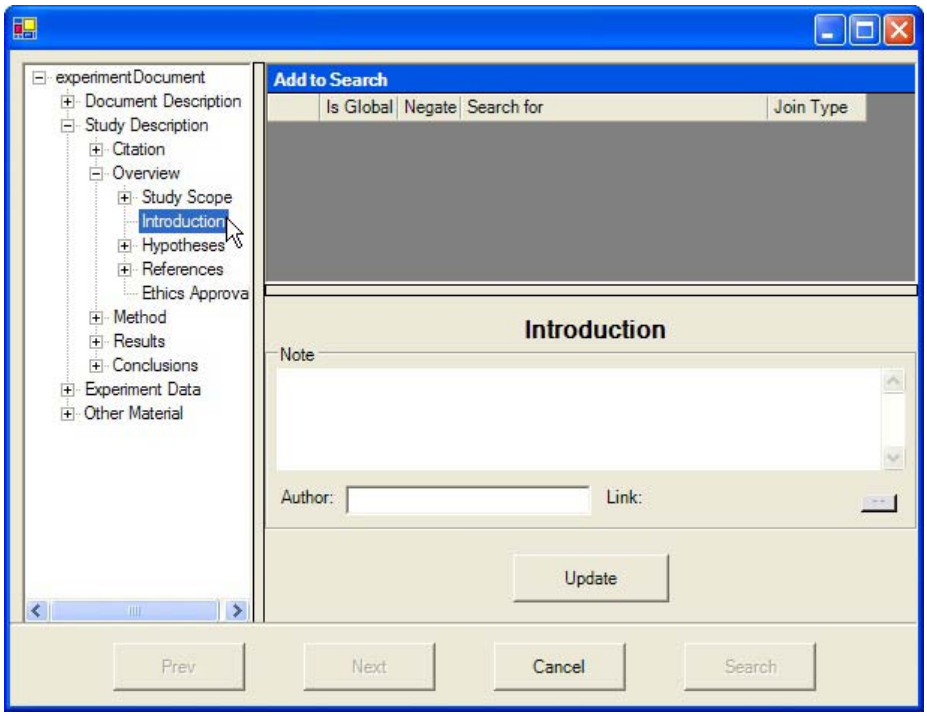

Figure 27. *Searching a Data Heading.*

These Data Editors are almost exactly like the regular *PI Assist* Data Editors, except in a few places where some modification may have been necessary for the purposes of searching. For instance, where a single date may appear in a *PI Assist* Data Editor, its corresponding Search Facility Data Editor may have a "From" and a "To" date, allowing the user to search for a range of dates.

To search for the string "performance" in the Notes section of the Overview Introduction, just type "performance" into the Notes textbox and click "Update".

One new search term will be added to your search (which has yet to be performed, as you are still building it—searches are never actually performed until you click "Search"). Options for the new search term appear in the upper right of the window.

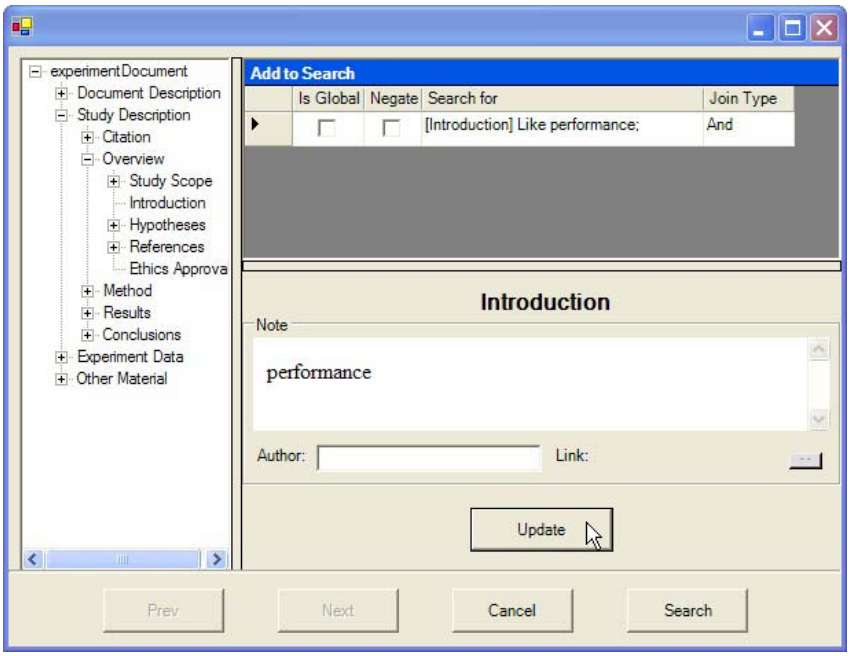

Figure 28. *Choosing a Data Editor for Searching.*

In Global: when this is checked, this term will be searched for in all headings with the same name as this one, throughout the entire *PI Assist* document. If left unchecked (as in Figure 29), the term will be searched for only in under *this* particular heading. Thus, in Figure 29, the search so far is for the word "performance" occurring anywhere in the Overview Introduction, but *not* in any other "Introduction" sections.

Negate: when this item is checked, the search term is negated (you are searching for *PI Assist* documents in which the search string does *not* appear).

Search For: this is not an option you can set, but an indicator of what this search term looks like so far. In Figure 28, it says "[Introduction] Like performance;" to indicate that so far we are searching for the string "performance" in "Introduction".

Join Type: this indicates whether this term is necessary for a match ("And"), or merely sufficient for a match ("Or"). "And" is the default, but can be changed to "Or" with the drop-down menu (see Figure 29).

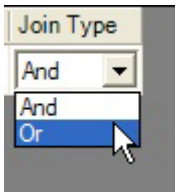

Figure 29. *Specifying AND and OR.*

Note that the Search Facility searches for whole words, so a search for "optim" will not find the word "optimal". A "wildcard" character, the percent sign (%), can be used to match part of a word. It matches any string of characters of length 1 or greater, so "optimal%" matches "optimally" and "optimistic", but not "optimal" or "suboptimal".

Figure 30 shows the result of adding a few more terms to our search, which now looks for all *PI Assist* documents that have the word "performance" in the Overview Introduction, *or* a word beginning with "optim" in *any* Introduction, *and* do *not* have any References anywhere with the word "Doe".

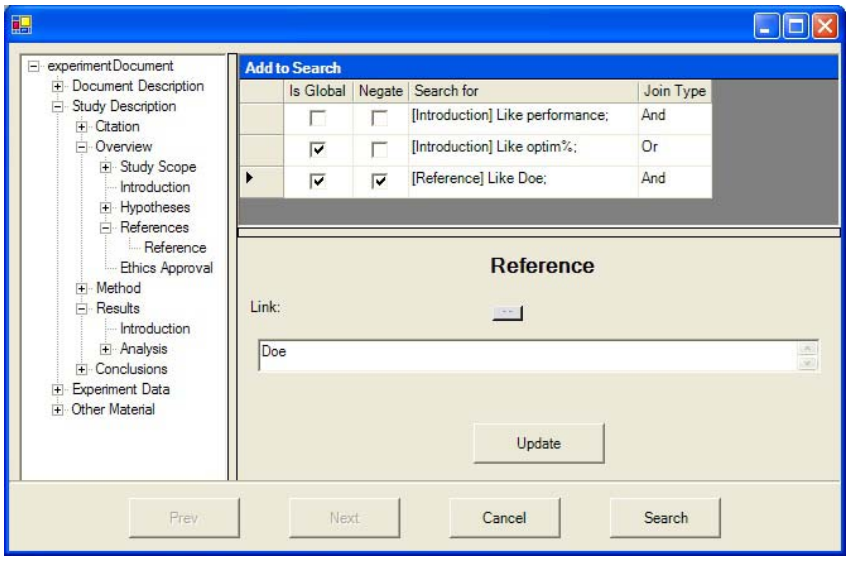

Figure 30. *A Compound Search.*

Click the "Search" button to do the search. When finished, the system will display a list of all documents in the database that fit your requirements. The search may take some time to complete; an hourglass is displayed while you wait.

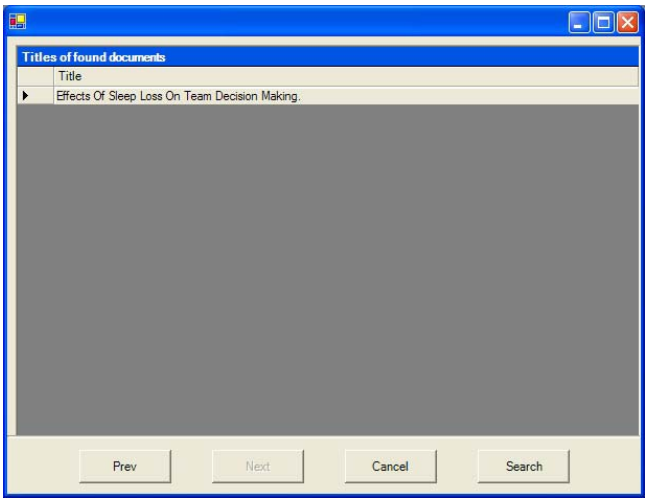

Figure 31. *Search Results.*

The window now displays a list of documents that match your search. In Figure 31, one document is found. If you wish to change your search, just use the "Prev" button to move backwards through the steps. Click "Cancel" at any time to leave the Search Facility.

To view one of the found documents, double-click it and it opens up in *PI Assist.* Note in the title bar, however, that it is a temporary file in a temporary location.

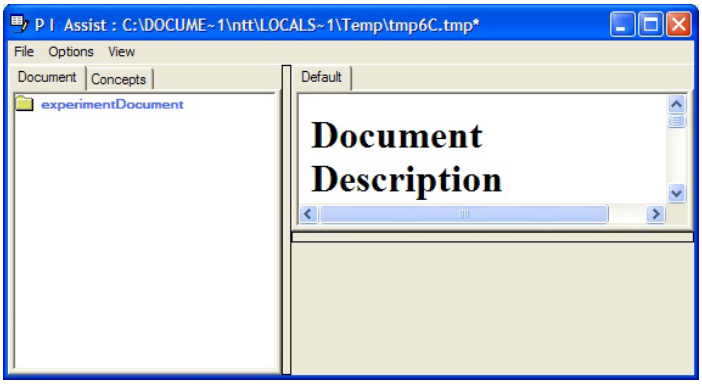

Figure 32. *Opening a Temporary Copy of a Database Document.*

When you exit *PI Assist,* you will be prompted to save the file. If you reply "Yes", the file will be saved in the location specified in the title bar (see Figure 32). If you decline to save, the file will be deleted.