

Participatory Workflow Analysis: Unveiling Scientific Research Processes with Physical Scientists

George Chin, Karen Schuchardt, James Myers, and Debbie Gracio

Pacific Northwest National Laboratory

P.O. Box 999

Richland, WA 99352 USA

+1 509 375-2663

george.chin@pnl.gov; karen.schuchardt@pnl.gov; james.myers@pnl.gov; deborah.gracio@pnl.gov

ABSTRACT

We are working with physical scientists from a diversity of domains to design and develop problem-solving environments for scientific research. This paper describes our current efforts in conducting domain analysis with different classes of physical scientists. We have employed a method we call *participatory workflow analysis*, which allows scientists to easily construct, evolve, and decompose workflow diagrams. Using this analysis approach, the scientists have been able to elucidate and document their various research and problem-solving functions and processes.

The motivation of this paper is to share our general experiences and findings from our application of participatory workflow analysis with physical scientists. We will describe the participatory workflow analysis approach, show examples from its usage, and present some of our findings from the analysis.

Keywords

Scientific workflow, workflow management systems, workflow analysis, problem-solving environments, participatory design, domain analysis, data flow diagrams.

INTRODUCTION

At Pacific Northwest National Laboratory (PNNL), a group of computer scientists have assembled to investigate and develop problem-solving environments (PSEs) for scientific research. As a Department of Energy national laboratory, PNNL possesses world-class, physical scientists with expertise in a wide range of scientific disciplines. The goal of the PSE team is to work closely with PNNL scientists to understand their scientific problem-solving practices and needs and to develop PSEs that support and enhance the scientists' abilities to address and solve more challenging and complex problems.

To many, the general notion of a problem-solving environment is unrestricted and all embracing. Gallopoulos, Houstic, and Rice [4] offer the following definition for a PSE:

In *PDC 2000 Proceedings of the Participatory Design Conference*. T. Cherkasky, J. Greenbaum, P. Mambrey, J. K. Pors (Eds.) New York, NY, USA, 28 November - 1 December 2000. CPSR, P.O. Box 717, Palo Alto, CA 94302 cpsr@cpsr.org ISBN 0-9667818-1-3

A PSE is a computer system that provides all the computational facilities necessary to solve a target class of problems... PSEs can track extended problem-solving tasks and allow users to review them easily. Overall, they create a framework that is all things to all people: they solve simple or complex problems, support rapid prototyping or detailed analysis, and can be used in introductory education or at the frontiers of science.

This basic view of PSEs encompasses nearly everything a scientist might do in his work. The challenge for our PSE team is to comprehend both the problem-solving processes of physical scientists and the current PSE technology far enough to identify those areas or aspects of PSEs that are most important to scientists and in most need of development.

Over the course of 12 weeks, we held a series of five analysis sessions with five different groups of physical scientists. During these sessions, scientists described and analyzed their research processes using an informal participatory analysis method we call *participatory workflow analysis*. The five groups of scientists were the following:

Computational Chemists – Theoretical and computational chemistry is the study of chemical phenomena through the use of mathematical models and computer calculations. Computational chemists apply high-performance systems and tools to conduct molecular modeling and simulation activities.

Nuclear Magnetic Resonance Experimentalists – Researchers apply nuclear magnetic resonance (NMR) spectrometers to characterize and investigate molecular structures. NMR experimentalists place molecules within magnetic fields to identify unknown substances, characterize specific arrangements of atoms within molecules, and to study the dynamics of interactions among molecules in solution.

Automotive Design Engineers – These engineers focus on the design of automobiles and their parts. In their design work, they employ commercial modeling and simulation packages that support various manufacturing and design functions such as forming the shape of a part, welding parts and seams, and simulating crash situations.

Fluid Dynamics Application Experts - Application users and developers apply fluid dynamics models to solve problems involving aqueous solutions. These researchers maintain and apply complex fluid dynamics software that has been developed in-house.

Regional Climate Modelers - Regional climate change models are used to project future environmental and climatic conditions based on current conditions and trends. Atmospheric scientists at PNNL are developing a national modeling center for the study of regional climate change.

PROBLEM-SOLVING ENVIRONMENTS AND WORKFLOW ANALYSIS

Today, most PSE research and development efforts are either focused on a specific domain or application [1, 8] or on the construction of general architectures for problem-solving [2, 9]. Domain-specific PSEs are developed to address the specific needs of a particular group of scientists and are generally not motivated nor designed to extend to other domains. In contrast, general PSE architectures often subscribe to the least common denominator approach where they attempt to provide capabilities that are applicable across many domains. For such architectures, the least common denominator always seems to point towards the computer as the most basic element. Subsequently, most general PSE architectures support scientific problem-solving by integrating and providing access to computers, applications, databases, and other computing resources.

Our view of PSEs is different. We submit that scientists may in fact exhibit similar behaviors and follow similar processes in their scientific research endeavors and believe that a broad-based PSE may be developed to address the real scientific problem-solving functions of different classes of researchers. Rather than supporting the mechanical steps of running applications and operating computers, we wish to provide higher-level tools that allow scientists to effectively manage their research and experiments - not just their computing resources.

Our disposition requires us to investigate PSEs from a different direction. Rather than looking outward from a specific PSE architecture to find applications and users, we want to look inwards from the different scientific domains to identify prototypical scientific problem-solving functions and needs. Our current efforts using participatory workflow analysis are aimed towards gathering scientific domain and work requirements as well as identifying strategic areas for PSE research and development.

Scientific Workflow

Scientific workflow is emerging as a popular concept in the research and development of scientific problem-solving environments [14, 15]. It focuses on those processes that are conducted by scientists during the course of their normal research. In addition to process, scientific workflow is also concerned with the data that are operated upon during the process. As emphasized by Wainer et al. [15], "collecting,

generating, and analyzing large amounts of heterogeneous data is the essence of such work, or at least of the components of scientific work that are more naturally the target of workflow management systems (WFMS)."

Workflow analysis is a natural domain analysis technique because its general aim is to describe the work that users conduct and describe it in a way that makes sense to users. Its use is particularly relevant to our project, given our specific interest in the development of PSEs and scientific WFMSs. If we are able to develop effective workflow representations and scientists are able to construct and apply these representations in their day-to-day research, then we would expect that scientists would also be able to apply such representations in the analysis of their own practices.

In workflow analysis, we seek to capture and comprehend two kinds of processes that scientists carry out. First, we want to capture the mechanical or procedural processes and tasks that scientists conduct in their research. The procedural process is the one most often depicted in current WFMSs. This is also the process that designers typically want to automate and optimize. Second, we want to capture the cognitive processes and tasks that scientists carry out. We wish to elucidate the cognitive problem-solving process at the scientist's level of thinking.

Workflow diagrams have taken on various graphical representations including entity-relationship diagrams [3], state charts [6], and Petri net models [12]. Looking for a simple representation that scientists could easily apply, we employ a representation resembling traditional data flow diagrams (DFDs) [16]. DFDs represent a conventional structured systems analysis approach that has been applied by software developers for decades since the advent of traditional waterfall software engineering methods. The use of DFDs in domain and participatory analysis, however, has largely been unexplored and untested, although Kensing and Munk-Madsen [7] have alluded to the potential adaptability of DFD methods to support user-developer communication.

Participatory Design

Our goal in applying workflow analysis is to gain a basic understanding of the scientific research processes that different classes of physical scientists carry out. Given that scientists specify, execute, and generally own these processes, the use of participatory design (PD) techniques also seems appropriate for our needs. PD advocates the direct participation of users in the analysis and design of systems that are to be eventually deployed in the user's work [5]. Ideally, users and developers would mutually apply analysis and design methods to create analysis and design representations.

In our study, the scientists' participation is confined to domain analysis because we are not seeking specific design solutions but rather seeking a broad view of scientific problem-solving and opportunities for PSE development. Over the past two decades, a variety of PD techniques

have emerged to support participatory analysis. Some well-known participatory analysis techniques include collaborative prototyping [13], storyboard prototyping [10], and CARD (Collaborative Analysis of Requirements and Design) [11]. These methods generally allow users to elucidate the steps of their work processes, the tools and artifacts of their work, and the overall environment from which the work is performed.

CONSTRUCTING WORKFLOW DIAGRAMS

To attune scientists to the construction of workflow diagrams, we provided them a simple, informal example of how a meteorologist might diagram his work in collecting and reporting weather conditions. In the example, we drew and labeled circles to represent specific activities such as “collect weather conditions,” “compute statistics,” and “report conditions and statistics.” We connected the circles using labeled arrows that identified the logical order of activities and the kinds of information that were passed between activities. We then decomposed the “collect weather conditions” activity to unveil its underlying steps such as “operating measurement devices,” “collecting results in a spreadsheet,” and “disseminating the results to local weather bureaus.”

The circles and arrows are basic representations of processes and flow (control and data) as found in data flow diagrams and flow charts [16]. Although we used circles and arrows in our example, we did not impose any specific symbology or rules on the scientists’ construction of workflow diagrams. In fact, the scientists used random combinations of shapes (e.g., circles, squares, ovals) in their diagrams. The purpose of the meteorology example was to illustrate the kinds of processes we were interested in capturing and studying and to provide scientists very basic mechanisms to describe these processes.

The construction of workflow diagrams seemed natural and easy to scientists. At times, the scientists did struggle in developing some diagrams, but the labor was mostly centered on the elucidation of research processes rather than the mechanics of diagramming. Figure 1 presents a workflow diagram developed by computational chemists during an analysis session. The diagram describes what the chemists called “results communications,” which is the process they carry out as they organize and analyze their experimental results in preparation to publication writing. The process consists of various steps such as conducting a literature search, gathering theoretical and experimental data, creating tables and figures, examining relevant comparisons and discrepancies in the data, validating the experimental results against theoretical findings, hypothesizing future results, and submitting to refereed publications. Figure 1 also illustrates the evolution and refinement of experimental results throughout the “results communications” process, as accumulated experimental and theoretical data are successively distilled first into tables and figures, then into general statistics and trends, and finally into validated

estimates. In the diagram, the chemists also marked those steps that typically involved collaborative effort among computational chemists.

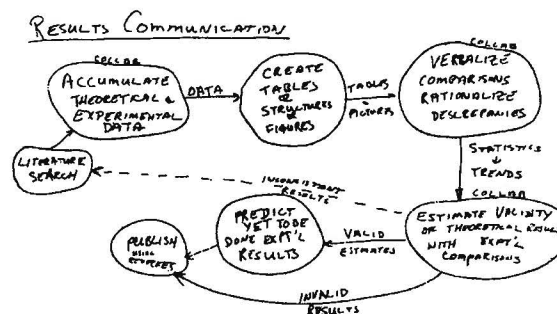


Figure 1. Workflow diagram representing the “results communications” process as drawn by computational chemists.

For the most part, the chemists were able to describe the “results communications” process in the context of their scientific research activities. One exception is the “create tables or structures or figures” step of Figure 1, which identifies the mechanics of the table creation activity rather than its scientific purpose. This activity may be rephrased in a scientific context as “identifying the salient, crucial portions of the theoretical and experimental data.” During workflow analysis, we found that scientists would sometimes describe portions of their workflow in terms of the manual operations they perform. In such cases, we encouraged the scientists to elaborate on the scientific purposes and meanings behind those operations as well.

The primary usefulness of the workflow diagrams was to provide a concrete artifact for analysis and discussion. The diagrams allowed scientists to quickly create and share representations of their work. Explicitly, the workflow diagrams described research activities and the control and data flow among these activities. Additionally, we asked scientists to identify

- points in the workflow where critical decisions were made
- stages in the workflow that require the collaborative effort of multiple scientists
- specific scientific and computing resources that are applied during each stage
- specific applications that are applied during each stage
- external resources, databases, and people that are involved in each stage.

As scientists further elaborated the details of their scientific workflows, they would embellish their workflow diagrams with additional symbols, shapes, and labels to convey additional information.

After an analysis session, we would sharpen and formalize the workflow diagrams that scientists had created. We

example, research data may reside on loose paper, in a notebook, in an electronic file, or in a database. Yet computational chemists continually reference and evolve all these research data over the course of their experiments. The accumulation and longevity of the research data emphasize the potential utility of integrated records management capabilities to help manage experimental data.

Another notable characteristic of the workflow diagram is the presence of a *calculation table*. The chemists in our study use a calculation table to manage their calculations. The calculation table specifies different combinations of molecules, computational methods, or other chemical or computational attributes. When the chemists initially formulate the problem they wish to attack, they develop a calculation table to specify the set of calculations to be conducted. If the current problem involves extending or completing research initiated by others, portions of the calculation table may initially be filled with results from previous research as gathered from publications, chemistry databases, and interactions with peers.

In addition to managing the set of calculations, the calculation table is also used to assign and manage tasks among collaborators. A particular chemist might be responsible for all calculations associated with a particular molecule, computational method, or some other attribute. In such an instance, the subset of calculations assigned to the chemist would represent a row or column in the table. Typically, the calculation table is maintained on paper and each chemist keeps and manages her own personal copy of the table. The chemists may synchronize their copies of the calculation table at key points during the research.

The calculation table represents a unique concept and capability that may be supported by a PSE. An electronic version of the calculation table would provide a centrally managed tool that chemists may share. Through such a tool, chemists could verify the current research status of their collaborators. Furthermore, if a specific entry in the calculation table was actually linked to the actual research data and log of activity of the calculation it represents, chemists may gain immediate access to the calculation's historical record to verify the calculation's correctness or to track down anomalies or errors in its results.

Depicting Procedural Activities

In Figure 2, the workflow diagram for "problem identification" describes the authentic practice of scientists during a particular stage of their research. One aim of workflow analysis is to unveil the scientific practice and the problem-solving behind that practice. The normal work functions of scientists, however, also involve many activities that are auxiliary to their scientific research. Many functions centered on the operation of computers fall under this classification. Tasks such as creating input files, translating between data formats, and executing programs on remote machines are all specific side effects of using the computer as a tool in scientific

work.

Among most of the scientists participating in the workflow analyses, the employment of high-performance computers is critical to their research because it allows them to solve bigger and more complex problems. Nevertheless, the truth remains that the less time scientists devote to attending to the details of operating computers, the more time they will have to devote to their scientific pursuits. This view leads to yet another aim of our workflow analysis, which is to identify processes that may be optimized or automated. Conspicuous candidates for optimization are those processes that are mechanical or procedural in nature such as submitting a computer job to run on a remote machine.

The workflow diagram in Figure 3 conveys the procedural execution of a computer simulation as described by regional climate modelers. The regional climate model in the diagram simulates local effects caused by changes in the regional climate. The model is a sophisticated, resource-demanding application that is typically executed on high-performance computers. The workflow diagram identifies the main parameters fed into the model as spatial resolution, scenario, temporal range, temporal frequency, and other model parameters. The complete set of model parameters is collected in an input file. The modeler then selects a computer on which to execute the regional climate model with the created input file. The computer job is then executed on the selected computer. As the job is executing, the modeler monitors both the status of the computer job and the generation of the regional climate data set. Regional climate data sets produced from successful runs of the model are often archived for future use and analysis.

In addition to monitoring the progress of a computer job, the ability to monitor the progress of the application or model is also important to the modeler. The regional climate modeler typically reviews the intermediate data produced by the climate model as it is executing. If discrepancies or unexpected results are found in the intermediate data, the modeler will try to trace the error back to incorrect or questionable input parameters.

A useful function of PSEs would be to allow users to select specific properties or fields of the intermediate results produced by the model and have these properties be monitored and tested against certain conditions. If the selected properties fail to meet the specified conditions, the PSE could alert the user and facilitate the modification of the input parameters and the restart of the application.

As shown in Figure 3, the modeler may iteratively modify the input parameters and rerun the regional climate model. In fact, modelers would often run their models with different input parameters to explore and experiment with the functionality and boundary conditions of the application or model. In other cases, the modeler may partition their target region into smaller areas and run multiple regional climate computations in parallel. A large number of computations may be difficult

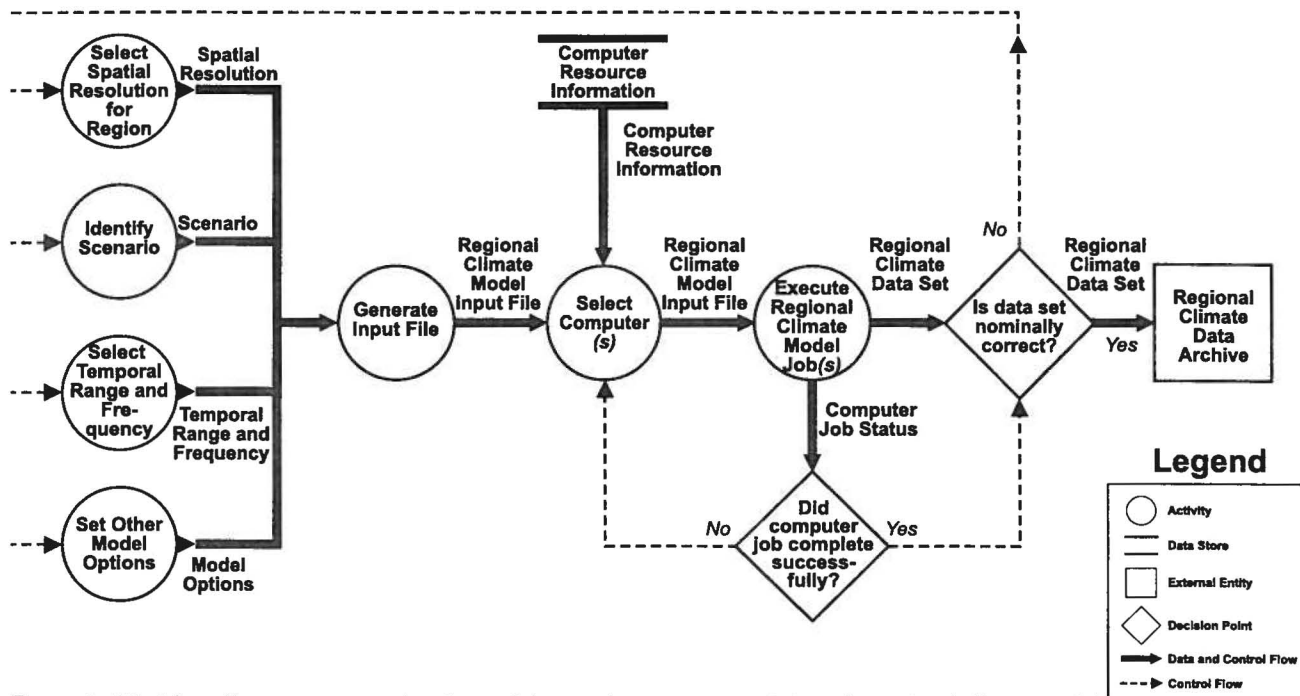


Figure 3. Workflow diagram representing the model execution process carried out by regional climate modelers.

for modelers to maintain and manage. Records management capabilities would be useful to assist the modeler in keeping track of his computations as well as to provide a historical record of activities that may be reviewed and verified

STAGES OF PROBLEM-SOLVING

Different scientists have different problem-solving requirements and foci depending on their specific research functions and roles. One way to better understand a scientist's research functions in comparison to those of other domains is to examine and contrast the individual stages of their research processes.

For each domain in our study, we developed high-level workflow diagrams to describe the domain's general research process. In Figure 4, we have collapsed this collection of high-level workflow diagrams into a single generalized workflow representation. The generalized workflow consists of the following six specific stages.

1. Define Problem – The first step in the problem-solving process is to identify and specify the problem to be addressed. For research problems, scientists may conduct literature searches and consult with peers to target specific scientific issues. For more applied problems, the issues are manifested in real-world conditions that need resolution or rectification. In these cases, the problems are not actively sought but rather part of the everyday function and work of applied scientists.
2. Design Experiment – Once the problem has been specified, the scientist develops an appropriate experimental strategy or system to attack the problem. The scientist maps out the experimentation process along with the models and resources she will apply in

finding a solution. Furthermore, the scientist may also begin to accumulate background or base research data that support the problem under investigation. If the experiment is a collaborative effort, individual tasks may also be negotiated and assigned during this stage.

3. Prepare Experiment – Prior to conducting the experiment, the scientist prepares the preliminary sample or data set to be processed. In addition, the initial and boundary conditions of the experiment are also defined as well as other input and control parameters that are fed into the experimental model.
4. Conduct Experiment – The experiment is carried out on computers or physical experimental apparatus. Raw data are generated from the experiment.
5. Collect and Analyze Results – Data from the experiment are compiled and analyzed. Scientists employ various kinds of analysis tools such as visualization, statistical, and domain-specific software packages to make sense of the data.
6. Gather and Distribute Findings – Scientists gather salient results and findings from the experiment in a publication or technical report.

The workflow diagram in Figure 4 also identifies key decision points in the problem-solving process. The first decision point occurs after the experiment is designed. If the experiment's specification is not complete at this point, the scientist must either further define the problem or abandon the problem as one that cannot be resolved given the experiment's current design. Upon conducting the experiment, a second decision point is reached. Scientists look for specific patterns, trends, and behaviors in the experimental data to verify its

reasonableness. If the specific patterns do not exist, the scientist must redesign the experiment to correct any flaws in its design.

The final decision point comes after the results of the experiment have been collected and analyzed. At this point, the scientist verifies the completeness and conclusiveness

of her findings. If the findings are found to be incomplete or inconclusive, the scientist may redefine the problem or redesign the experiment and conduct additional runs to validate or bolster the findings. Alternatively, the scientist may conclude that the problem is unsolvable or intractable and abandon it altogether.

Figure 4 presents an expanded view of scientific problem-solving that extends beyond the scope of most modern-day problem-solving environments. General PSE architectures focus on the “conduct experiment” stage of our scientific research model. They generally provide tools to integrate, execute, and monitor computer-based simulations and calculations in distributed computing environments. In comparison, domain-specific PSEs have greater breadth. Such PSEs may provide tools that allow scientists to build preliminary data sets, identify specific scientific properties, construct input files, execute simulations or calculations, statistically analyze data, and visualize experimental results. These are all functions of the “prepare experiment”, “conduct experiment”, and “collect and analyze results” stages of our scientific research model. Both general and domain-specific PSEs are typically geared to solve computational problems.

The high-level workflow diagram in Figure 4 illustrates the other problem-solving stages that are generally neglected by the PSE development community. In the three problem-solving stages supported by domain-specific PSEs, these three stages capture the mechanical process that scientist carry out as they conduct experiments. What existing PSEs fail to capture is the very process in which the mechanical process is designed. An experiment has specific scientific goals. It has constraints (e.g., time, space, money, manpower) and access to specific resources. The ability to fashion an experiment to optimally make use of resources, alleviate constraints, address goals, and identify solutions is a critical part of scientific problem-solving. We call this function of scientists *experimental design*, which is encapsulated in the “design experiment” stage of our scientific research model.

Furthermore, experimental design fits within an even larger context. Scientists design experiments to investigate and validate specific hypotheses. When scientists identify problems, they also formulate the hypotheses they wish to prove. We call this more basic function of scientists *research design*, which is encapsulated in the “define problem” stage of our scientific research model. The problem and the hypotheses are the drivers of the research.

The high-level workflow diagram provides a context from which we can review and analyze the problem-solving characteristics of different classes of scientific problem-solvers. The research processes that scientists elaborated in their workflow analyses emphasize different stages of our high-level model. In Table 1, we examine the research activities of each class of problem-solvers with respect to the individual stages of our scientific research model.

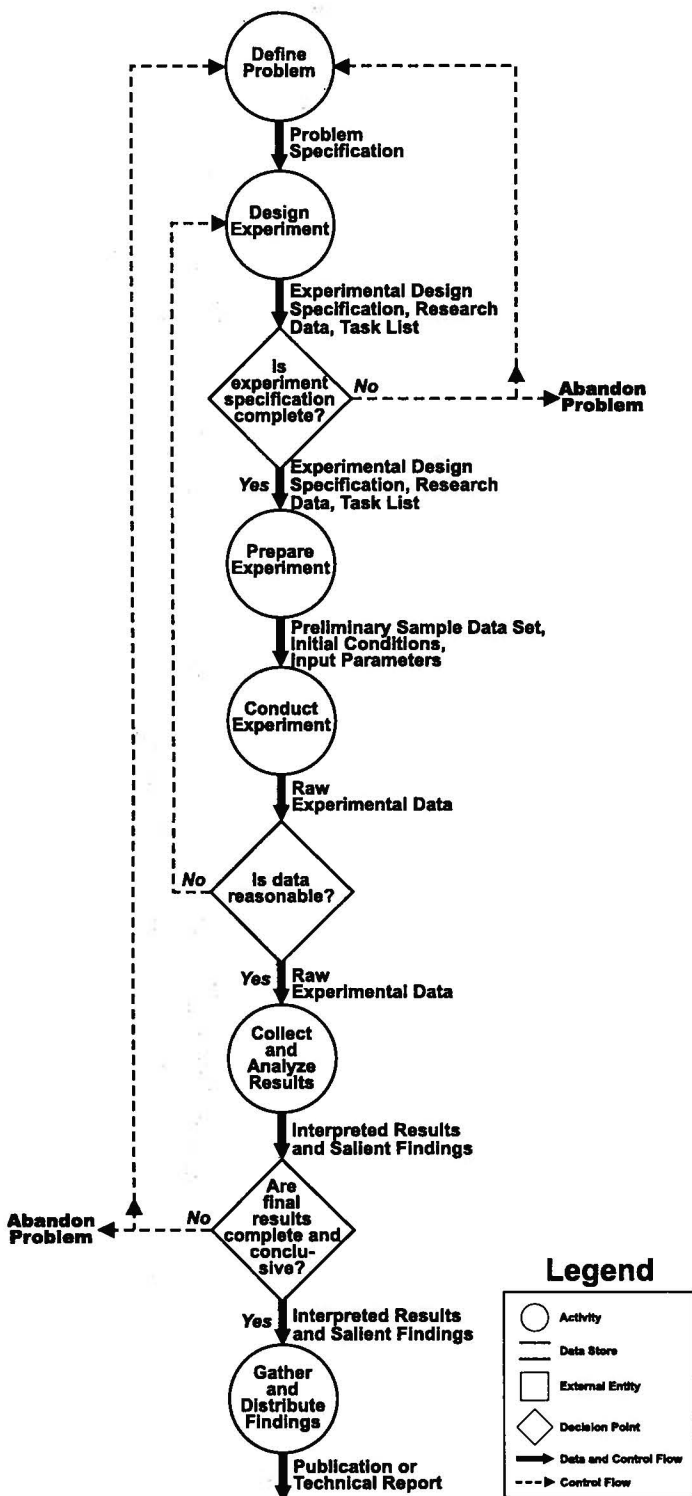


Figure 4. High-level workflow diagram of the scientific research process.

Defining the Problem

During the problem definition stage, the computational chemist and the NMR experimentalist examine existing research to identify areas that have yet to be investigated and to which they may contribute. They may browse through reference databases and research bibliographies to find relevant articles, as well as draw in the knowledge and expertise of peers. Thus, theoretical and experimental scientists would benefit from facilities that allow them to access, gather, and manage published research work as well as to identify and establish communications with peers.

In the case of the automotive design engineer and the regional climate modeler, these problem-solvers do not play a significant role in defining the problem. The engineer attacks design problems that are assigned by his manager, while the modeler leaves it to the external client to specify and solve her own particular problems. As a result, engineers and modelers may require little PSE support for problem definition. In contrast, the fluid dynamics application expert works with his client to specify the problem. The client originates the problem, but the application expert helps the client elaborate the problem in the context of the application. Thus, the application expert may make use of capabilities that allow her to share and brainstorm problems and issues with clients.

For the computational chemist, NMR experimentalist, and fluid dynamics application expert, these scientists first identify the problem and then fashion a research design or strategy to reach a solution. To support the activity of research design, scientists may desire high-level capabilities to develop, manage, and track research hypotheses and to tie these hypotheses more closely to the actual experiment.

Designing the Experiment

In next stage of our scientific research model, scientists design experiments. Except for the automotive design engineer, all scientists in our study actively planned out the experimental process. In their experimental designs, the scientists identified the specific steps of their experiments and the key decision points in the process. Furthermore, the experimental design may also identify the computing and scientific resources to be applied. Typically, the experimental design is informally specified in loose notes or in the laboratory notebooks of individual scientists. Capabilities to assist scientists in identifying resources and in specifying, maintaining, and sharing the experimental design would be useful to scientists. The capability to manage and navigate the process that the experimental design represents would also be useful. These kinds of capabilities have largely been associated with WFMSs.

For the automotive design engineers, the design of their work processes is somewhat fixed. The engineer follows a particular sequence of steps (i.e., forming, welding, crash simulation) and employs specific applications and computing resources that are mandated by his organization. To support engineers, one might envision capabilities that allow managers to define

and make available specific design models for engineers to follow.

For the computational chemist and the NMR experimentalist, these scientists often conduct experiments that involve completing or extending existing research work. In such cases, the scientists derive their problems and data from research that may be gathered from published work, molecular or chemistry databases, or directly from collaborating researchers. A useful capability for these scientists would be some facility to draw and accumulate theoretical and experimental data from a variety of sources.

For the fluid dynamics application experts and regional climate modelers, their models are typically executed in conjunction with other domain-specific models. For example, a hydrologist may execute a groundwater model on a data set that was originally produced by a regional climate model. Thus, the ability of scientists to incorporate new domain-specific models and advisor tools that may assist them in selecting and applying the correct set of domain-specific models would be a useful capability.

For all scientists, the negotiation and assignment of tasks were typically accomplished during this stage. Task scheduling and monitoring would be another useful capability for scientists – particularly if scientists were able to define and manage tasks along meaningful mechanisms such as the computational chemists' calculation table.

Preparing the Experiment

Except for the NMR experimentalist, all scientists in our analyses run computational models as the primary activity in their experiments. To prepare for the calculations, these scientists collect required input information to define their models. The input information represents specific scientific attributes of the model, and thus, is domain and model specific. To collect model attributes, scientists may employ general computational tools such as molecule builders, data analysis packages, and CAD/CAM systems. In other cases, more specialized, domain-specific programs may also be applied. The diversity of computational tools emphasizes the need for an integration platform where different programs may be integrated into a common environment.

For all scientists, a data management facility for storing and managing model attributes would allow scientists to reproduce, review, validate, and share model information. In addition, the regional climate modeler may have to conduct field studies to collect salient information to be fed into the regional climate model. Thus, data acquisition and management facilities for collecting and managing external data would also be useful to the climate modeler.

In some cases, the computational chemist, fluid dynamics application expert, and the regional climate modeler may need to develop custom functions to be integrated into their scientific models. The main applications that computational chemists and regional climate modelers apply are normally

Stage of Scientific Process	Computational Chemist	NMR Experimentalist	Automotive Design Engineer	Fluid Dynamics Application Expert	Regional Climate Modeler
Define Problem	Research existing work; consult peers; research design	Research existing work; consult peers; research design	Problem handed down by management	Problem initially defined by client, but elaborated with application expert; research design	Problem defined by external clients
Design Experiment	Experimental design, accumulate theoretical and experimental data; define calculation table; select molecules and computational methods; assign tasks	Experimental design, accumulate theoretical and experimental data; assign tasks	Experimental design is established by organization; assign tasks	Experimental design; collect preliminary data set, select physics models; predict model behavior; assign tasks	Experimental design; collect preliminary data set, select environment models; assign tasks
Prepare Experiment	Identify chemical structures and basis sets; program custom functions	Generate physical sample; prepare experiment apparatus; identify proper control settings	Define base geometry	Define base geometry; define initial boundary and field conditions; program custom functions	Collect preliminary data from global climate models and field studies; preprocess preliminary data; program custom functions
Conduct Experiment	Construct computational chemistry model input file; execute computational chemistry model	Physically place sample in NMR; set controls; run experiment	Construct input files for and execute forming, welding, and crash simulation models; translate among data formats of models	Construct fluid dynamics model input file; execute fluid dynamics model	Construct regional climate model input file; execute regional climate model; apply additional environmental models
Collect & Analyze Results	Identify comparisons and discrepancies in experimental results; validate and visualize results; extrapolate results to other scenarios	Examine spectra produced from experiment; validate and visualize results	Validate and visualize results after each stage	Compare experimental results against predicted behavior; validate and visualize results	Validate and visualize results
Gather & Distribute Findings	Generate publication	Generate publication	Generate specification and technical report	Generate technical report	Generate publication or technical report

Table 1. Research activities of different classes of scientists throughout specific stages of the scientific research process

designed to be extensible – allowing end-users to incorporate and access their own functions from within the application. The chemists and modelers did occasionally incorporate new codes into their main applications. The fluid dynamics application experts, however, did often modify or extend the code of their main fluid dynamics application. Their main application was developed in-house – giving them the freedom to modify the actual code to solve specific problems.

To accommodate end-user development of code, another useful feature would allow the end-user to integrate and configure his own code modules and allow them to be used with other applications and models. A more sophisticated solution would be to provide actual software development support through a programming environment.

Conducting the Experiment

Except in the case of the NMR experimentalist, all scientists in our analysis execute one or more computational models in their experiments. Scientists may seek a variety of capabilities to assist in executing computational models such as intelligent user interfaces, job launching, resource management, records management, data archiving, application monitoring, and interactive steering.

Furthermore, the automotive design engineer and the regional climate modeler typically work with a number of different models. In their cases, additional capabilities to manage

and integrate different computational models and to provide translations between different accompanying data formats would also be valuable.

Collecting and Analyzing Results

In the next stage, scientists collect and analyze results from the experiment. Unlike other kinds of scientists, the NMR experimentalist collects raw data directly from a physical device. He uses a specific hardware and software system to acquire the data and to store the data on a computer. For the NMR experimentalist, data management facilities need to support the collection, management, and integration of data originating from physical devices.

During data analysis, the precise reasoning that a scientist follows is bound to vary among different domains and individual scientists. Regardless, all scientists generally evaluate their experimental results in relation to the scientific hypotheses they established during problem identification. The capability to track hypotheses and manage the research design process is relevant to this stage as scientists attempt to validate their experimental results and conclusions against the problem and hypotheses that they originally defined.

Furthermore, scientists employ a variety of tools during analysis such as visualization and statistics packages. The use of such tools again emphasizes a general need for an integration platform from which different applications may be accessed and linked.

Gathering and Distributing Findings

In the final stage of our scientific research model, all scientists conclude with the writing of a publication or technical report. While commercial word processing and desktop publishing tools are in abundant supply, specific capabilities to gather and organize experimental hypotheses, models, data, and findings would be particularly useful and relevant to scientists and their writing functions. In addition, scientists may also wish to have collaborative writing and publishing tools that allow them to collaboratively develop publications and reports.

NEXT STEP

We have applied participatory workflow analysis with physical scientists from five different domains. The collective results of the analyses underscore the similarities and differences among the problem-solving functions and processes of different scientists and domains. These results are important and useful because they

- describe the overall scope of problem-solving activities that may be supported by PSEs
- serve as a point of comparison for the problem-solving requirements of different domains
- highlight novel or attractive problem-solving areas where we might target initial development efforts
- assess the range of usefulness of problem-solving tools and capabilities as they are developed and deployed.

The workflow analyses were designed to give us a general overview of the problem-solving functions and activities of different domains. Our intentions were not to collect a complete set of requirements for the purpose of designing a specific system. Rather, we sought to survey the current scientific environment to understand the different kinds of problems that physical scientists attack in their research and how they go about solving them.

For our next step, we plan to work with atmospheric scientists to develop a PSE for regional climate research and applications. Working from our current domain analysis of regional climate modeling, we will conduct more detailed requirements analysis and design, using additional participatory design methods such as collaborative and storyboard prototyping.

REFERENCES

1. Akers, R., Elaine, K., Curtis, R., Stanley, S., and Young, R. SciNapse: A problem-solving environment for partial differential equations. *IEEE Computational Science and Engineering* 4, 3, 32-42.
2. Catlin, A., Chui, C., Crabill, C., Houstis, E.N., Markus, S., Rice, J.R., and Weerawarana, S. PDELab: An object-oriented framework for building problem solving environments for PDE based applications, in *Proceedings of the Second Annual Object-Oriented Numerics Conference*, (Sun River, OR, April, 1994), RogueWare Software, 79-92.

3. Chen, P. *Entity-Relationship Approach to Information Modeling and Analysis*. North Holland, Amsterdam, 1983.
4. Gallopoulos, E., Houstis, E.N., and Rice, J.R. Computer as thinker/doer: Problem solving environments for computational science. *IEEE Computational Science and Engineering* 1, 2, 11-23.
5. Greenbaum, J., and Kyng M. Introduction: Situated design. In Greenbaum, J., and Kyng M. (eds.), *Design at Work: Cooperative Design of Computer Systems*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1-24.
6. Harel, D. Statecharts: A visual formalism for complex systems. *Science of Computer Programming* 8, 1987, 231-274.
7. Kensing, F., and Madsen, K.H. PD: Structure in the Toolbox. *Communications of the ACM* 36, 4, 78-85.
8. Knox, R.G., Kalb, V.L., Kendig, D.J., and Levine, E.R. A workbench for interactive simulation of ecosystems (WISE). *IEEE Computational Science and Engineering* 4, 3, 52-60.
9. Lakshman, Y.N., Char, B.W., Johnson, J. Software components using symbolic computation for problem solving environments. In *Proceedings of the 1998 International Symposium on Symbolic and Algebraic Computation*, (Rostock, Germany, August, 1998), ACM Press, 46-53.
10. Madsen, K.H., and Aiken, P.H. Experiences using cooperative interactive storyboard prototyping. *Communications of the ACM* 36, 4, 57-66.
11. Muller, M.J., Tudor, L.G., Wildman, D.M., White, E.A., Root, R.A., and Deikmann, E. Bifocal tool for scenarios and representations in particular activities with users. In Carroll, J.M. (ed.), *Scenario-Based Design: Envisioning Work and Technology in System Development*, J. Wiley, New York, 1995, 135-163.
12. Peterson, J. *Petri Net Theory and the Modeling of Systems*. Prentice Hall, Englewood Cliffs, NJ, 1981.
13. Schuler, D., and Namioka (eds.). *Participatory Design: Principles and Practices*. Erlbaum Associates, Hillsdale, NJ, 1993.
14. Singh, M. and Vouk, M. Scientific workflows: scientific computing meets transactional workflows. In *Proceedings of the NSF Workshop on Workflow and Process Automation: State-of-the-Art and Future Directions*, (Athens, GA, May 1996).
15. Wainer, J., Weske, M., Vossen, G., and Medeiros, C.B. Scientific workflow systems. In *Proceedings of the NSF Workshop on Workflow and Process Automation: State-of-the-Art and Future Directions*, (Athens, GA, May 1996).
16. Yourdon, E. *Modern Structured Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1988.