

PROCESS INFORMATICS—A NEW PARADIGM FOR BUILDING COMPLEX CHEMICAL REACTION MODELS

Investigators: David M. Golden, Consulting Professor, and Craig T. Bowman, Professor, Department of Mechanical Engineering

Sponsor: Global Climate and Energy Program, Stanford University (Completed 12/31/05)

Scope of Process Informatics: *Process Informatics* is a data-centric approach to developing predictive models for complex chemical reaction systems. It deals with all aspects of integration of pertinent data of complex systems (industrial processes and natural phenomena) whose complexity originates from *chemical reaction networks*. The primary goal of process informatics will be information gathering, validation, and transformation into useable form. The latter will include development of *predictive* (numerical/computer) models with quantified degrees of reliability. While such a scope may be applicable to any process, the immediate focus of Process Informatics will be on *chemically-based* processes.

Motivation: Chemical reaction models will never be complete. A problem is that the data on which models are based are scattered over different sources and are not properly evaluated. Most importantly, these data cannot be applied directly to practical problems—they have to be “transformed” into useful models. Such models, however, cannot be created by simple “compilation” of the data. Chemical reaction model building is a time-consuming activity that requires expert knowledge. The goal is to convert such model building into science, automate the methodology, and make the results available in a prompt and convenient form for the user.

Immediate needs for predictive reaction models presently exist in combustion engineering, the petrochemical industry, atmospheric chemistry, materials processing, biological systems and pharmaceuticals. As computers become more powerful and more readily accessible to industry, the industrial interest in process simulation is continuously growing.

Vision of Process Informatics: The Process Informatics infrastructure will have two principal components: a *Database* and a collection of *Tools*.

The Process Informatics *Database* will represent the most currently complete set of knowledge available in a given field. In the field of combustion, it will contain experimental data, on both combustion systems and on elementary reactions, molecular properties determined from quantum chemical calculations, reaction rates obtained from reaction-rate theories, and similar information. The Process Informatics *Tools* will be of two general kinds, those enabling the collection, transfer, organization, display, and mining of the data —i.e., computer science tools, and those enabling processing and analysis of the data along with assembly of the data into models—i.e., scientific and numerical tools.

The two principal customers of the Process Informatics System are the *data provider* and the *model user*. During the development stage, there will also be a *model builder*, whose role will eventually be automated—and providing the means for this automation will be a prime objective.

A *Data Provider* (Experimenter, Theorist) makes a request to deposit new observations or new computational results. The protocol assures completeness of the data submission. The deposited data are immediately analyzed for consistency with the database and the results are reported both to the data provider and to the *scientific council* (see below). Upon approval of the council, the database is modified. In other words, the database will be fluid and will be continuously modified and these modifications will be documented.

A *Model User* (design engineer, CFD researcher) requests a kinetic model (or a simulation with such a model) and specifies the conditions of interests, the desired level of accuracy, and the mathematic form of the model (detailed, reduced, parameterized, etc). The system checks for the existence of such a model, if none is available one is generated.

Another user might be a project manager (a member of the Scientific Counsel, or a researcher) who wants to know whether a proposed experiment/calculation will improve the current database, and by how much. Various scenarios with the envisioned experiment or calculation can be evaluate. A similar question can be posed differently: what needs to be done to improve the predictability of a given model? Repeat old experiments? Under what conditions? New experiments? Which ones? New Calculations? What level of local error must be maintained to accomplish the stated goal?

The goal is not merely a collection of tools, but a shift in the paradigm of the scientific process: building targeted knowledge by the entire community and providing the wealth of information in its entirety to every user. To attain the vision described above will require the following: a) creation of enabling software infrastructure; b) development and implementation of scientific methods taking the full advantage of the approach; and c) establishment of a new paradigm of scientific collaboration for data collection, evaluation, and utilization.

Road to Process Informatics: Development of the above vision requires a concrete system. Reaction chemistry of combustion will serve as the initial system. Nearly all of the energy currently used in the industrialized world comes from burning fossil fuels and chemistry is the essence of combustion systems, from internal combustion engines to gas turbines. Knowledge of the chemical mechanism is at the center of device design to limit combustion-generated environmental pollution. Societal demands for cleaner and more efficient combustion are rapidly bringing the chemical aspects of combustion processes to the forefront.

A first effort at developing the new paradigm for dynamic models of complex chemical systems has been demonstrated with a quantitative chemical model for natural gas combustion, "GRI-Mech." http://www.me.berkeley.edu/gri_mech/

Organizational Structure: The success of the undertaking will also depend on organization and management. Based on the experience gained from the GRI-Mech project, the following organization is envisaged: To initiate the project, several working teams will be created, organized by subject and each containing representatives of different disciplines as necessary. An underlying goal will be a periodic release of "the best current model" via a dedicated web site. Leaders of all teams will form a management team, with a rotationally assigned leader. The entire team will meet periodically, probably coinciding with or as a part of a professional and contractual meeting

Scientific Council: The concept of the Scientific Council is an important component of the proposal. Its mission is "quality control" of the knowledge buildup in the scientific community, with combustion chemistry as the initial focus. The Council membership will begin with a few experts, with the intent of encompassing and engaging the entire community in time.

In many ways the Scientific Council is similar to a Data Evaluation Panel, an established practice today for database quality control. The difference—and hence the novelty underlying the shift in the scientific paradigm—is that the Council activity will be based on the analysis of the *entire knowledge* available in the field. It will be the goal of Process Informatics to develop tools and infrastructure to enable such operation of the Scientific Council.

Team: As this document is prepared, interest has been expressed in Process Informatics from the several organizations in the United States and abroad.

Status: The GRI-Mech project is complete. This new project will begin with GCEP funding.

Contact: david.golden@stanford.edu

GCEP Final Report Available Online: http://gcep.stanford.edu/research/technical_report.html