# Process Systems Engineering, 6. Cyberinfrastructure and Decision Support Systems

ULLMANN'S ENCYCLOPEDIA OF INDUSTRIAL CHEMISTRY

**Venkat Venkatasubramanian,** Laboratory for Intelligent Process Systems, School of Chemical Engineering, Purdue University, West Lafayette, United States

**Leaelaf Hailemariam,** Engineering Sciences, The Dow Chemical Company, Midland, United States 02811

**Pradeep Suresh,** Engineering Sciences, The Dow Chemical Company, Midland, United States 02811

1. Dawn of the Data-Rich Era . . . . . 1
2. From Data to Knowledge to Decisions . 3
3. Cyberinfrastructure: Current Trends and Future Outlook . . . . . 5
3.1. General Purpose Tools and Environments. . . . . 5
3.2. Recent Progress in some Application Domains of Chemical Engineers . . . 8
4. Decision Support Systems . . . . . 8
5. Summary . . . . . 9

## 1. Dawn of the Data-Rich Era

Among the various recent trends on the application of computational methodologies in chemical engineering, the one that perhaps shows a great deal of novelty and promise is the use of cyberinfrastructure (CI) and informatics concepts, in conjunction with intelligent systems (IS), to support decision-making in all areas of chemical engineering [1]. Recent trends further suggest that there is an informatics revolution underway. One is moving from an era of limited data obtained via time consuming experiments and long duration simulations to one of a deluge enabled by high-throughput experiments and TeraGrid computing environments—it's a dramatic transition from a "data poor" to a "data rich" era. Further, the ever increasing monitoring of equipment, processes, and products at all scales, from individual units to enterprise-level supply chains (e.g., use of radio frequency identifications (RFIDs)), is another source of such data overload.

For instance, in pharmaceutical drug development and manufacturing, the amount and complexity of information of different types, ranging from raw experimental data to lab reports to complex mathematical models that need to be stored, accessed, validated, manipulated, managed, and used for decision-making are staggering [2].

But it is not raw data that is of interest. What is desired are in-depth knowledge and mechanistic, first-principles-based, understanding of the underlying phenomena that can be modeled to aid in rational decision making. However, knowledge extraction and model development from this data deluge are major challenges. Past approaches developed in a "data poor" era do not work well in this new world. The new environment requires imaginative thinking to address these challenges. This is where CI, informatics, and IS will play a crucial role.

Now, CI and informatics may mean different things to different people; therefore a definition, however limited, can be helpful. The Atkins Report [3] uses an analogy to industrial infrastructure such as transportation, communication, or power systems to define CI as "*the infrastructure based upon distributed computer, information, and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy.*" Another definition, also put forward by National Science Foundation (NSF) which may be considered as a


10.1002/14356007.o22_o10

working definition, states that CI is the "*integration of hardware, middleware, software, data bases, sensors, and human resources, all interconnected by a network.*" Obviously, even this is quite broad, but that is the nature of the terrain. Nevertheless, it articulates the different components and stresses the importance of middleware, integration, and networking (see Fig. 1).

In a similar vein, for informatics, a useful working definition is that it is the study of the structure, algorithms, behavior, and interactions of systems that store, process, access, manage, communicate, and use information. While the discipline of informatics has developed its own conceptual and theoretical foundations, it also utilizes foundations developed in other fields. Informatics is generally seen as a key component of CI.

IS, on the other hand, deal with the development and application of artificial intelligence methodologies to problems in chemical engineering. In the 1980s and early 1990s, when the first papers on expert systems, neural networks, machine learning, and genetic algorithm methodologies for process systems engineering (PSE) problems appeared in computers and chemical engineering, they were viewed with skepticism. Since then these methodologies have proven their value by addressing a class of problems of practical importance that were previously hard to solve using the traditional techniques. Artificial intelligence (AI) approaches have become more mainstream now, accepted as a part of the modeling arsenal of PSEs, with several important successes in the domains of process fault detection and diagnosis, molecular products design, process synthesis, process safety analysis, scheduling, and so on. They are now well poised to make further inroads as a natural complement to the CI and informatics approaches.

Much of the past contributions only addressed different slices of the overall problem but not the entire picture — data, information, and knowledge management issues were addressed separately, leading to stand alone systems with limited capabilities and significant integration barriers. In addition, the amount and complexity of data are orders of magnitude greater now, which is changing the scope of the challenge dramatically, resulting in data warehouses that often become data graveyards. The time has come, and the appropriate

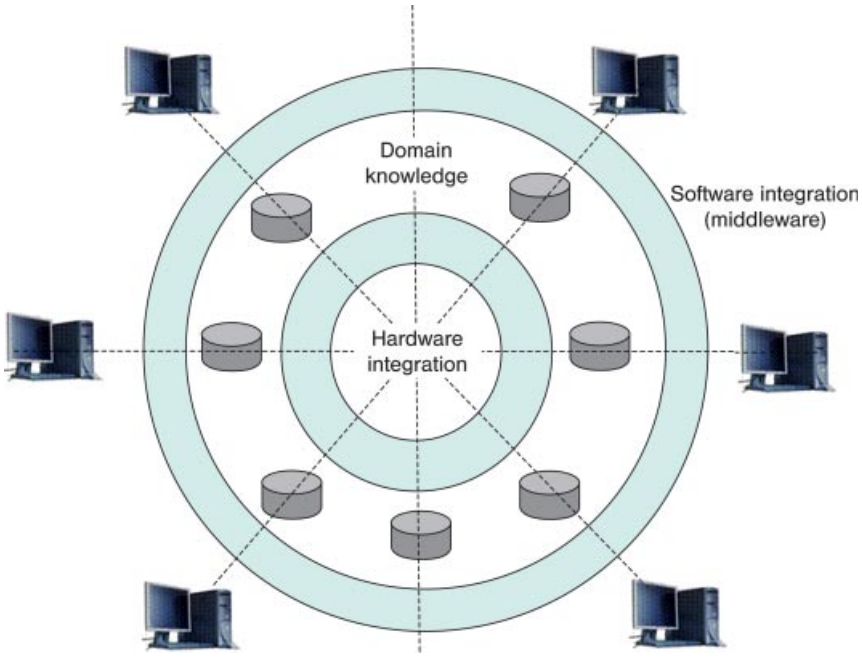


**Figure 1.** Representation of a CI

theoretical frameworks and practical technologies are emerging, to address this problem in all by developing more comprehensive CI and informatics approaches.

However, all this would require some innovative thinking and getting over some misconceptions. For instance, there is a tendency to underestimate informatics as just programming or database management. This would be akin to stating that chemical engineering is nothing but chemistry carried out in large vessels, i.e., "large-scale chemistry". Certainly, there is chemistry being done in those large process units, but, there is much more to it. Similarly, informatics certainly involves programming, but that's not all. Certainly, informatics involves data storage and management, but that's not all.

Another equally important point to note is that these looming challenges can not be addressed by hoping that computer scientists/engineers will somehow address them for us. Once again, invoking another chemical engineering analogy, this would be like thinking that since transport phenomena problems are solved by using differential equations, mathematicians would address these core problems of our discipline for us. Again, they can not as they lack the application domain knowledge.

A similar attitude and approach towards CI and informatics is needed to foster enabled solutions to ones challenges. The use of CI, informatics, and intelligent systems concepts and techniques in chemical engineering is no different from the modeling tools which are historically borrowed, adapted and enhanced from mathematics (e.g., linear algebra and differential equations) or operations research (e.g., mathematical programming).

Exciting opportunities exist in developing modeling environments that can effectively manage the data deluge from high-throughput experiments for the discovery of new materials such as catalysts. Pharmaceutical engineering is another ripe area where research on new ontologies and CI-enabled IS for decision support are already in progress. In the emerging "smart plant" manufacturing environment of the future, opportunities abound for the development of CI–IS solutions. Systems biology and healthcare informatics are also natural domains for the CI–IS framework. Risk identification and management in complex engineered systems is another such prime candidate with problems that would require a CI–IS approach. It is quite obvious that the solutions for our challenges in achieving sustainable energy systems would require a CI-IS approach. Global, multiplant-based, supply chains optimization is another such candidate.

## 2. From Data to Knowledge to Decisions

Addressing the formidable modeling and informatics challenges would require a broader approach to modeling than what chemical engineers are used to. In common parlance, one tends to use the terms data, information, and knowledge more or less synonymously. However, they do represent different concepts and that distinction becomes important as *formal* methodologies are developed necessary for *automating* tasks such as acquisition, representation, storage, manipulation, modification of, and reasoning with, data, information, and knowledge. The formal representation techniques used for data, information, and knowledge are often quite different, the automation protocols for their exchange are also different, and so are the algorithms used for manipulating and reasoning with them for decision-making. In the past, these issues did not seem terribly important as there was a tremendous amount of human intervention in passing data and information from one program to another and people took care of all these barriers. However, given the increased need to automate such computational tasks due to the data deluge, as well as due to the complexity of the problems these issues are addressed.

Generally speaking, when one mentions modeling to chemical engineers, people often think of a system of differential and algebraic equations (DAEs). However, there is a wider variety of knowledge representation concepts leading to other classes of models [4] which will play an important role in this emerging future. While it is not the purpose of this paper to have an extensive discussion on various modeling concepts, it is, nevertheless, useful for the theme to outline and summarize the issues involved.

One may broadly classify models into i) mechanism-driven models based on first-principles and ii) data-driven models. Again, each of these classes may be further categorized into i) quantitative and ii) qualitative models. Combinations of these classes lead to hybrid models.

DAE models are suitable for a certain class of problems that are amenable to such a mathematical description; chemical engineering has abundant examples of this class. However, there are other kinds of knowledge that do not lend themselves to such modeling. For example, reasoning about cause and effect in a process plant is central to fault diagnosis, risk analysis, alarm management, and intelligent supervisory control. Knowledge modeling for this problem class does not typically lend itself to the traditional DAE view of modeling. In some simple cases perhaps one can, but they are incapable of addressing real-life industrial process systems which are often complex and nonlinear with incomplete and/or uncertain data. Further, even for simple systems DAE-based models are not suitable for generating explanations about causal behavior. This problem often requires a hybrid model, such as a combination of a graph theoretical model (e.g., signed digraphs) or a production system model (e.g., rule-based representations) and a data-driven model (e.g., principal component analysis (PCA) or neural networks) [5].

Thus, while one is quite familiar with ODE/PDE, statistical regression, and mathematical programming models, one is less so with other classes such as graph theoretical models (as noted, used extensively to perform causal reasoning in abnormal events identification and diagnosis, risk analysis etc.), Petri nets (used for modeling discrete event systems), rule-based production system models (used in knowledge-based systems for automating higher-order reasoning), semantic network models such as ontologies (used in materials discovery and design that utilize complex relational databases, domain-specific compilers, etc.), object-oriented models such as agents (used in simulating the behavior and decision-making choices of independent, interacting, entities endowed with complex attributes and decision-making powers), and so on. In addition, there are the data-driven quantitative models such as pattern recognition-based models (e.g., neural nets, fuzzy logic), stochastic models (e.g., genetic algorithm, simulated annealing), etc.

Even though, as far back as the 1960s, the need for such an alternative modeling philosophy in the context of process synthesis was recognized [6], not much work on this subject appeared in the chemical engineering literature until the 1980s. This renewed interest was propelled by the progress in knowledge representation and search techniques in AI as well as in computing hardware and software. Outstanding examples from that era are the DESIGN-KIT system for process engineering [7] and the DECADE system for catalyst selection [8]. However, the progress was viewed with skepticism. Nevertheless, recognizing the importance of this alternative philosophy, this need was highlighted in the CAST award acceptance speeches in the hope of stimulating further research [9, 10].

Over the recent decade, much progress has been made as these methodologies proved their value by addressing problems of practical importance, which were previously hard, even impossible, to solve using the traditional modeling techniques. These AI-based modeling approaches have become more mainstream now, accepted as a part of the modeling arsenal of PSE, with several important successes in the domains of abnormal events management, molecular products design, process synthesis, process safety analysis, scheduling, and so on.

Despite all of this progress, the number of academic researchers developing the alternative modeling methodologies in chemical engineering is very small and inadequate considering the emerging challenges. This is mostly due to two factors: (i) the barrier of entry is quite high — requiring a considerable investment in time and effort to achieve a sufficient level of mastery in knowledge representation and search techniques, algorithm engineering, databases, compilers, and so on, which are not part of a typical chemical engineering education program; and (ii) certain misconceptions about what the intellectual challenges are.

But the crucial emerging trend, borne out of several necessitating factors, is the ever increasing automation of higher-order reasoning and decision-making. These activities that were once considered to be the exclusive domain of

humans are slowly, but surely, being taken over by computers. Such automated reasoning and decision-making will be driven by models, but these are not going to be limited to DAE-based models. To be sure, the DAE-based models will play their useful role wherever they are appropriate, but the other kind will play an increasingly important role

In some sense, it would be easier if all our modeling needs could be addressed by just the DAE models. It is a much more mature field of study, with decades of literature on how to define, formulate, and solve such models. The alternative modeling philosophy, on the other hand, are typically used for ill-posed or ill-structured problems, lacking a unique approach, and with a certain amount of heuristic element to them. They often lack the beauty and rigor of the DAE models. But these deficiencies are sufficiently compensated for by their ability to address, though not always, messy, real-life, industrial problems where they often provide very good solutions, even optimal ones.

It is important to recognize that these two approaches, in general, are complementary and not competitive, even though there are problems where they do compete. As noted above, if indeed, there is a DAE-based solution that would be satisfactory for a given problem, one should pursue it instead of the other methodologies. The strength of the alternate modeling philosophy lies is its ability to address a class of problems that are not amenable to the DAE framework. However, one often falls into the trap of using a particular modeling technique as the panacea for all problems — e.g., to a PCA specialist, every problem may seem to require that particular approach. What is needed is not such a 'tool driven' outlook, but a 'tool box' oriented modeling philosophy, where one is comfortable in using a wide variety of modeling tools. It is helpful to remind of what Prof. George Box, the statistician from the University of Wisconsin, once famously said: "All models are wrong, some are useful."

## 3. Cyberinfrastructure: Current Trends and Future Outlook

The emerging "data rich" networked environment will eventually impact all aspects of CI and, in fact, the effects are being felt in some areas already. These effects are spawning new application domains and opportunities for modeling skills. The following summary is not meant to be a comprehensive review but only a representative survey of some recent developments that could serve as useful starting points for interested readers to explore further. First progress in general purpose environments is summarized and then domain or application specific results are outlined.

CI developments have been along the lines of concepts, methodologies, and tools required for two broad categories of needs:

1. Data modeling and management
2. Knowledge modeling and management

CI typically involves:

- Hardware: Progress in hardware generally has been in developing high-performance and parallel computing environments such as the TeraGrid
- Software: Software integration includes middleware between different software or programming languages
- Middleware: Middleware are systems that connect multiple applications through the use of a common data model. Use of a common information model makes it possible for greater flexibility and efficiency.
- Networking: Networking usually exploits the internet and wireless communication technologies
- Domain knowledge components in an integrated environment

### 3.1. General Purpose Tools and Environments

***Modeling and Simulation.*** While efforts such as the computer-aided process engineering-open (CAPE-OPEN) standard, which define interfaces between physical property packages, numerical solvers, and unit operation libraries, have been underway for a decade, issues of interoperability have been gaining a great deal of importance. In an extension of CAPE–OPEN a framework (COGents) for combining and assembling CAPE-OPEN compliant

components using software agents was composed by Braunschweig [11]. Similarly, component-based hierarchical explorative open process simulator (CHEOPS) [12] is a conceptual framework for process model integration. CHEOPS includes a neutral model representation (conceptual object model), a tool wrapper, a modeling tool, a formulation bridge (to convert between two model formulations), and solution algorithms. More recent advances in this project are described in [13].

***LIMS and Data Warehouses.*** Vast amounts of data and information are frequently entered into laboratory information management systems (LIMS) (→ Laboratory Information and Management Systems (LIMS)) and electronic laboratory notebooks (ELN) [14], which are repositories of raw data. Important recent progress towards storing and managing such large data sets include the virtual laboratory (VL) project at the University of Amsterdam [15] and XSIGMA [16]. However, most current LIMS and ELN solutions utilize database schemas, which often limit the capacity for the description of complex relations between information entities. This is addressed by developing domains-specific ontologies as discussed below. Data warehouses [17] use specialized database schemas to abstract and store a copy of data from several sources, and enable those data to be queried through a single query.

***Ontologies, Languages, and Compilers.*** Addressing the challenges in modeling and informatics requires a departure from the *application-centric* approach of the past to an *information-centric* framework. In this new paradigm, the underlying data, information, and knowledge are modeled explicitly, independent of the tools that use these. Instead of encoding such information in objects in a particular programming language or tool specific constructs, the information is explicitly described. However, to describe the information explicitly, the syntax (i.e., structure) as well as semantics (i.e., meaning) of the information must be defined. The explicit description of domain concepts and relationships between these concepts is known as an *ontology* [18].

Recent developments in the field of ontology have created new software capabilities that facilitate the implementation of the information-centric infrastructure. Compared to a database schema which targets physical data independence, and an XML schema which targets document structure, an ontology targets agreed upon and explicit semantics of information. As a result, while the functionalities of this infrastructure can be implemented in a traditional client-server framework, the main benefits of this ontology-driven architecture are its openness and semantic richness.

One recent development is the Purdue ontology for pharmaceutical engineering (POPE) to support automated decision-making by intelligent systems for pharmaceutical product and process development and manufacturing [19]. POPE is composed of several ontological subsystems that formally model data, information, and knowledge regarding experiments, materials, chemical species and reactions, expert knowledge, unit operations, and mathematical models using the web ontology language (OWL) [20]. A particularly innovative contribution in POPE is the creation of a mathematical models ontology that represents mathematical models as well as their underlying assumptions. This framework separates the declarative and procedural components of mathematical models creation, manipulation, and solution. Figure 2 shows a view of some of the ontological relationships in POPE.

The utility of such an environment might be illustrated with an example. A typical problem scenario in pharmaceutical manufacturing is considered, one where the recent batch of tablets have failed the dissolution test–a critical test where one evaluates how well a sample tablet dissolves in 250 mL of buffer. Current approaches to diagnosing and resolving this important product quality problem requires a detailed and thorough investigation, comparing current batch records with those of previous batches, proposing failure hypotheses, analyzing and checking hundreds to thousands of pages of documentation on raw material properties, process operating conditions, design specifications, modeling assumptions (made a few years ago when the models were originally developed for this manufacturing process and the model development engineer is no longer with the company), and even the original lab data on dissolution (experiments conducted

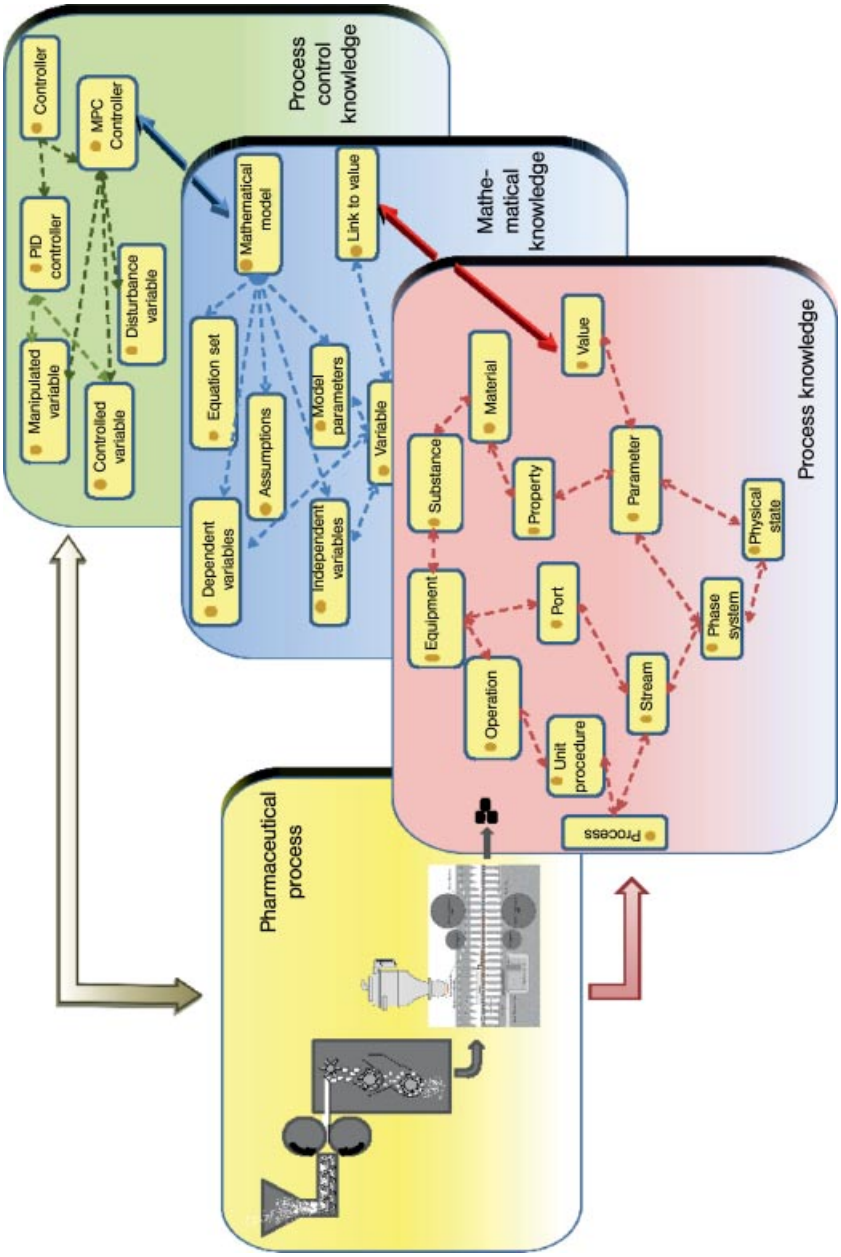


**Figure 2.** View of some of the ontological relationships in POPE

several years ago and data were recorded as entries in hard to read laboratory notebooks). Using a POPE-like ontological informatics system, such data, information, and models would be readily accessible through the ontological relationships for an operator or a design engineer to make the right decisions at the right time. Such a system would greatly empower human decision-making in real-time by screening vast quantities of data and information to find better solutions, thereby reducing the time to complete a complex reasoning task.

### 3.2. Recent Progress in some Application Domains of Chemical Engineers

***Computational Chemistry.*** GridChem [21] is a Java desktop application that includes a client, a grid middleware server, and a set of distributed, high-end computational resources. Computational chemistry tools are connected, through the GridChem client portal that undertakes job management, to a middleware server that handles data management. The middleware server is connected to computational grids like GLOBUS, Condor, and MyProxy. Both the middleware server and GridChem client portal are connected to long-term storage.

The collaboratory for chemical kinetics [22] provides a web environment which includes a database of chemistry documents, tools for search like Tableseer and chemistry entity search as well as access to a globus grid environment for grid computation. NorthWest Chemistry (NWChem) [23] is another important software system for computational chemistry on a grid.

***Molecular Products Design and Discovery.*** Molecular products design deals with the important and difficult problem of discovering and designing new materials and formulations with desired properties. This encompasses a wide variety of products such as fuel and oil additives, polymeric composites, rubber compounds, paints and varnishes, catalysts, etc. Recent research efforts are beginning to address the informatics and multiscale modeling challenges [24] in this domain. One such attempt is the multiscale model-based informatics framework called *Discovery Informatics* [25]. The discovery informatics framework has led to the successful development of automated, rational, materials design systems in several industrial applications, such as gasoline additives, formulated rubbers, and catalyst design [26–28]. As noted above, the richness and complexity of the underlying chemistries in this domain has led to important advancements in knowledge representation, languages, compilers, molecular structure search engines, chemical entities extraction systems, and so on. A recent contribution by [29] shows the value of the informatics-based approach in catalytic chemistry.

There is considerable literature in domains such as drug discovery or bioinformatics [30, 31]. The modeling and informatics challenges in other application domains such as smart manufacturing plants [32, 33], enterprise-wide optimization [34], sustainable energy and environmental systems, pharmaceutical engineering, nanoscale engineering, engineering virtual organizations [35] etc., require CI-based solutions.

## 4. Decision Support Systems

A broad definition of a decision support system (DSS) states that a DSS is a computer-based system that aids the process of decision making [36]. There have been several other attempts at providing an all inclusive definition while covering specific characteristics and behavior of these systems [37–42]. DSS have been classified into communications-driven, data-driven, document-driven, knowledge-driven, and model-driven DSSs [43]. The last two categories are usually synonymous with the term ‘expert systems’ which is a classical branch of AI [44].

Within the domain of PSE, decision support systems have been developed and used in areas including process modeling, optimization, simulation, design, and fault diagnosis. Most of these efforts consisted of reasoning over a knowledge base encoded in software programming languages as rules of thumb [45–48]. Some of them use the term DSS loosely to represent techniques (e.g., spreadsheet-based, simple data manipulation and visualization) that aid decision-making at a low level but do not have the ’intelligence’ to propose or suggest actions in a given situation [49]. Here, the discussion will be restricted to those that support decision-making at a higher level and reason over the domain knowledge that is part of a CI.

An example of the use of a CI for decision support is PHASuite [50, 51]. In this effort, hazard and operability analysis (HAZOP) was automated to aid and guide human experts. The domain knowledge required for this (equipment models, process models, safety related knowledge), was organized based on ontologies [18] and represented in Microsoft Access databases. The knowledge thus stored could be accessed by a variety of tools for a variety of applications.

Information sharing was ensured by the creation of translators or dictionaries that converted input information (Batch Plus simulation files, XML files, CAD files, etc.) into ontologies and output information from the ontologies into external documentation tools (PHAPro). Procedural knowledge was written in Visual C++ which carried the knowhow for using the declarative knowledge in the ontologies. This knowledge was managed using case-based techniques that associated new scenarios to ones already present in the knowledgebase using the manner in which cases were represented and organized. An automated functionality to add and modify the knowledgebase was also made part of this commercial tool. The reasoning over this ontological knowledgebase was carried out using the concept of colored Petri nets which are complex bipartite graphs used to represent discrete event systems. A two level, two layer reasoning methodology was proposed, namely, an operation level and an equipment level with a Petri net layer and a safety model layer as part of each level. The implementation of this commercial decision support tool for hazard and operability analysis was further tested with a typical batch process in a pharmaceutical company. Figure 3 shows the major components of PHASuite.

Another DSS that uses declarative knowledge representation has been developed in the area of process monitoring and utilities optimization [52]. In this work, an agent-based approach was developed for decision support in performance prediction, monitoring, scheduling, and resource optimization. A task ontology and agent communication ontology defined the required concepts and relations which the heuristic rules and inferencing were embedded in Java agent development framework (JADE). Ontologies were represented in a resource description framework (RDF) file using OntoEdit. Various agents reasoned over the ontological knowledgebase based on the heuristic rules, process data, and models in heterogeneous formats. Figure 4 shows a schematic of the proposed system.

Two other notable contributions in the application of CI towards decision support are those by [53, 54]. Although these efforts have shown case studies in pharmaceutical engineering, one can easily see the extension of the proposed ideas and technologies to PSE in general. In the work by Zhao et al. [53], the domain knowledge pertaining to pharmaceutical solid dosage formulation, namely, the process or unit operation related knowledge and the material related knowledge was represented in OWL. OWL unlike other forms for representing ontologies is semantically rich and inherently supports constraint enforcement and rule-based reasoning. The guidelines for drug product formulation were represented in guideline interchange format (GLIF), a platform independent guideline representation and execution format within the ontological framework. The end result of this effort was a prototype automated decision support tool that aided human experts in the selection of ingredients for a solid dosage form [19].

To perform reasoning and data consistency checks [54] focus on the underlying semantics of an ontological CI. OWL supported axioms were used to check the consistency of information input to a process ontology. Semantic web rule language (SWRL) rules were written to enforce certain rules that provided inferencing capabilities. With the combination of these technologies, a compliance management framework in the pharmaceutical processing domain was created, namely OntoREG. The resulting framework ensured information integrity, reasoned over ontological process and regulatory knowledge, and alerted users about missing information and regulatory violations in the process knowledge.

## 5. Summary

Driven by powerful convergent forces of technologies, the coming data deluge poses unprecedented challenges and opportunities in modeling and informatics fronts. While computer scientists and information technologists can help, the demands imposed by the chemical engineering domain-specific knowledge and constraints, which are unlikely to be understood and appreciated by outsiders, make it clear that only chemical engineers can address these challenges — in particular, these need to be addressed by the PSE community as it is the one best positioned to do so. However, addressing these would require discarding some traditional

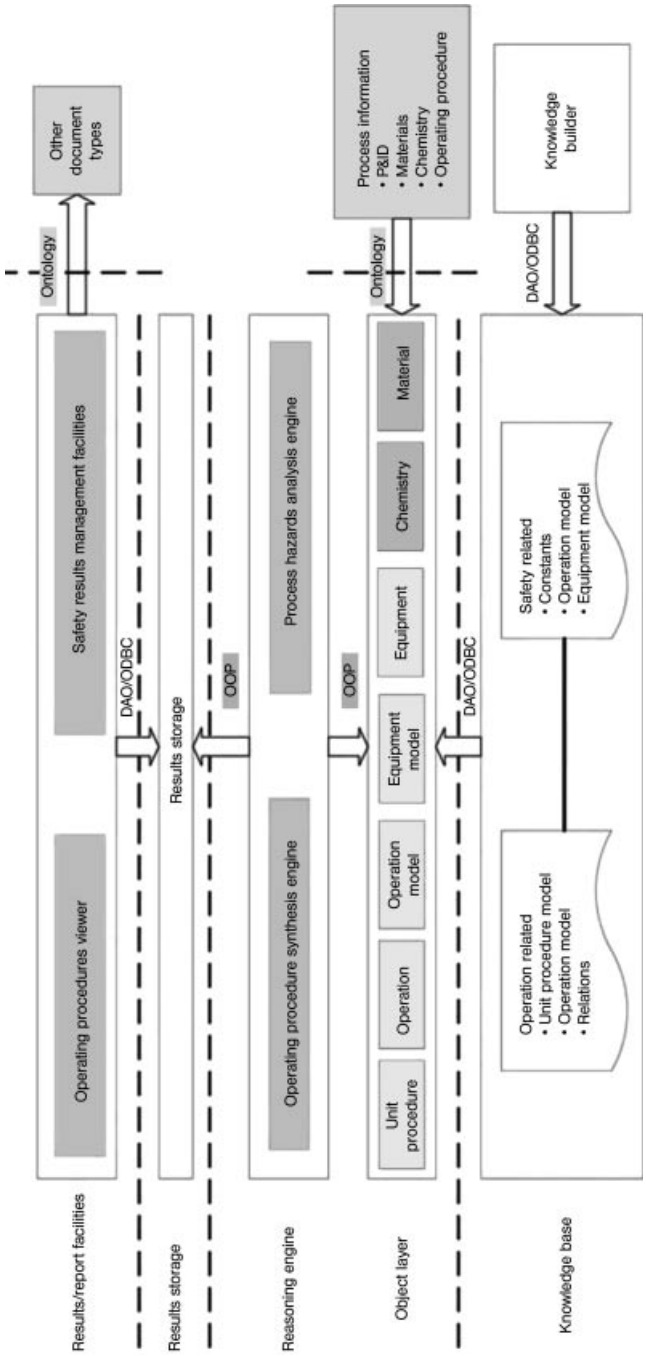


**Figure 3.** Major components of PHASuite

misconceptions about informatics and non-DAE-based modeling methodologies, and fostering innovative approaches towards a wider class of knowledge modeling. Great opportunities for making field defining intellectual contributions await one in inventing chemical engineering domain-specific CI components such as ontologies, compilers, molecular

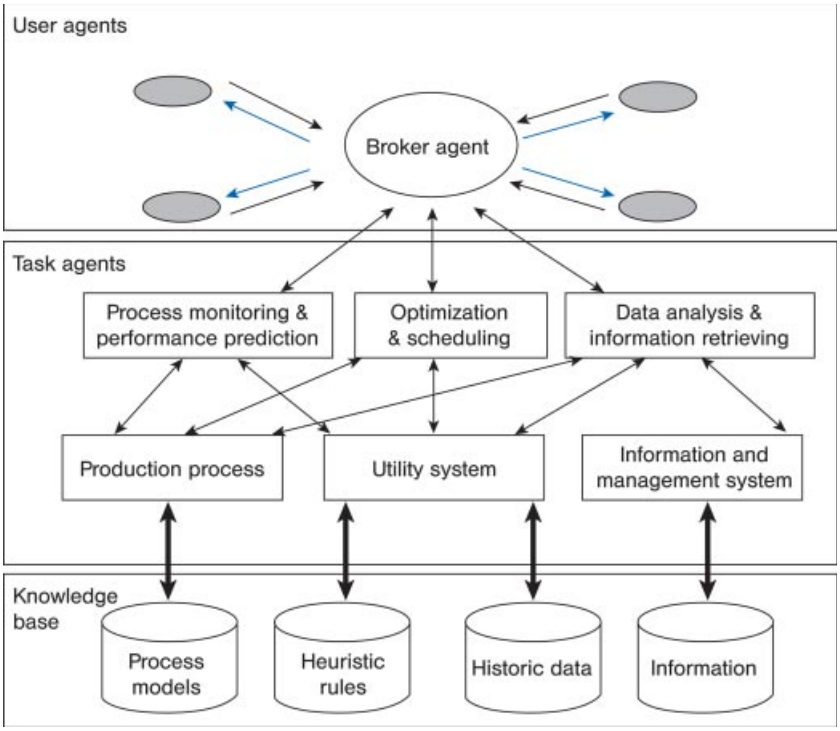


**Figure 4.** Multiagent system framework for decision support in PSE

structure and semantic search engines, chemical entities extraction systems, languages for modeling chemical reaction pathways, modeling and knowledge management environments, visualization, virtual organizations, cybersecurity systems, and so on. Naturally, all these also provide one with opportunities for new business ventures in high-end modeling and informatics products and services. This is a long, adventurous, and intellectually exciting journey that have only barely begun, but progress in this will revolutionize all aspects of chemical engineering for years to come.

## References


1 V. Venkatasubramanian: "Drowning in Data: Informatics and Modeling Challenges in a Data Rich Networked World", *AIChE J.* **55** (2009) no. 1, 2–8.

2 P. McKenzie et al: "Can Pharmaceutical Process Development Become High Tech?", *AIChE J.* **52** (2006) no. 12, 2006.

3 Atkins Report: Revolutionizing Science and Engineering Through Cyberinfrastructure, Report of the Blue-Ribbon Advisory Panel on Cyberinfrastructure, National Science Foundation, 2003.

4 L. Ungar, V. Venkatasubramanian: "Advanced Knowledge Representation", *CACHE Monographs Series*, CACHE, Austin, Texas, 1988.

5 V. Venkatasubramanian et al.: "A Review of Process Fault Detection and Diagnosis–Part I: Quantitative Model-Based Methods", *Comput. Chem. Eng.* **27** (2003) no. 3, 293–311.

6 D.F. Rudd, G.J. Powers, J.J. Siirola: *Process Synthesis*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

7 G. Stephanopoulos et al.: "DESIGN-KIT: An Object-Oriented Environment for Process Engineering", *Comput. Chem. Eng.* **11** (1987) no. 6, 655–674.

8 R. Bañares-Alcántara et al.: "DECADE: A Hybrid Expert System for Catalyst Selection. I: Expert Systems Considerations", *Comput. Chem. Eng.* **11** (1987) no. 3, 265–277.

9 W. Seider: "The Quantitative-Qualitative Dichotomy in Process Engineering", *CAST Communications* **16** (1993) no. 1, 9–15.

10 G. Stephanopoulos: "Knowledge, Computers, and Chemical Engineering: A Critical Synthesis", *CAST Communications* **17** (1994) no. 1, 7–11.

11 B. Braunschweig: "CAPE Web Services: The COGents Way", *Computer Aided Process Engineering* **18** (2004) issue C, 1021–1026.

12 G. Schopfer et al.: "CHEOPS: A tool-integration platform for chemical process modeling and simulation", *International Journal of Software Tools and Technology Transfer* **6** (2004) 186–202.

13 M. Nagl, W. Marquardt: *Collaborative and Distributed Chemical Engineering*, Springer Verlag, Berlin 2008.

14 C. Paszko, C. Pugsley: "Considerations in Selecting a Laboratory Information Management System (LIMS)", *American Laboratory* **9** (2000) 38–42.

15 A. Frenkel et al.: "Information management for material science applications in a virtual laboratory", *Lecture Notes in Computer Science* **2113** (2001) 165–174.

16 D. Kim et al.: "X-SIGMA: XML based Simple data Integration system for Gathering, Managing, and Accessing Scientific Experimental Data in Grid Environments", *Second IEEE International Conference on e-Science and Grid Computing*, Amsterdam, The Netherlands, 2006.

17 S.P. Gardner: "Ontologies and semantic data integration", *Drug Discovery Today,* **10** (2005) no. 14, 1001–1008.

18 T.R. Gruber: "A translation approach to portable ontology specification", *Knowledge Acquisition* **5** (1993) no. 2, 199–220.

19 V. Venkatasubramanian et al.: "Ontological Informatics Infrastructure for Chemical Product Design and Process Development", *Comput. Chem. Eng.* **30** (2006) no. 10–12, 1482–1496.

20 OWL: Web ontology language overview, http://www.w3.org/TR/owl-features/ (accessed 11 January 2012).

21 GridChem, Computational Chemistry Grid, https://www.gridchem.org/ (accessed 11 January 2012).

22 ChemSeer, http://dirac.chem.psu.edu/index.htm (accessed 11 January 2012).

23 R.A. Kendall et al.: "High-performance computational chemistry: an overview of NWChem a distributed parallel application", *Comput. Phys. Commun.* **128** (2000) 260–283.

24 J.J. de Pablo: "Molecular and Multiscale Modeling in Chemical Engineering–Current View and Future Perspectives", *AIChE J.* **51** (2005) no. 9.

25 J.M. Caruthers et al.: "Catalyst Design: Knowledge Extraction from High Throughput Experimentation", in "Understanding Catalysis from a Fundamental Perspective: Past, Present, and Future", *J. Catal.* **216** (2003) no. 1–2, 98–109.

26 A. Sundaram et al.: "Design of Fuel Additives Using Neural Networks and Evolutionary Algorithms", *AICHE J.* **47** (2001) no. 6, 1387–1406.

27 P. Ghosh et al.: "Sulfur Vulcanization of Natural Rubber for Benzothiazole Accelerated Formulations: From Reaction Mechanisms to a Rational Kinetic Model", *Rubber Chem. Technol.* **76** (2003) no. 3, 592–693.

28 S. Katare et al.: "An Intelligent System for Reaction Kinetic Modeling and Catalyst Design", *Ind. Eng. Chem. Res. and Dev.* **43** (2004) no. 14, 3484–3512.

29 V. Prasad, D. Vlachos: "Multiscale Model and Informatics-Based Optimal Design of Experiments: Application to the Catalytic Decomposition of Ammonia on Ruthenium", *Ind. Eng. Chem. Res.* **47** (2008) no. 17, 6555–6567.

30 C.A. Floudas: "Computational Methods in Protein Structure Prediction", *Biotechnol. Bioeng.* **97** (2007) no. 2, 2007.

31 G.L. Moore, C.D. Maranas: "Computational Challenges in Combinatorial Library Design for Protein Engineering", *AIChE J.* **50** (2004) no. 2.

32 J.F. Davis: *Smart Process Manufacturing Workshop Report*, NSF Roadmap Development Workshop, April, Arlington, VA, 2008.

33 T.F. Edgar, J.F. Davis: "Smart Process Manufacturing–A Vision of the Future", *Ind. Eng. Chem. Res. and Dev.* Centennial Issue, 2008.

34 I.E. Grossmann: "Enterprise-wide Optimization: A New Frontier in Process Systems Engineering", *AIChE J.* **51** (2005) no. 7, 1846–1857.

35 pharmaHUB, www.pharmahub.org (accessed 11 January 2012).

36 P.N. Finlay: "Introducing decision support systems", Blackwell Scientific Publ., Oxford 1994.

37 E. Turban: "Decision support and expert systems: management support systems", Prentice Hall, Englewood Cliffs, NY 1995.

38 P.G.W. Keen, M.S. Scott Morton: "Decision support systems: an organizational perspective", Addison-Wesley Publ., Reading, MA 1978.

39 R.H. Sprague, E.D. Carlson: "Building effective decision support systems", Prentice-Hall, Englewood Cliffs, NY, 1982.

40 A. Schroff: "An approach to user oriented decision support systems", Ph.D. thesis, University of Fribourg, Fribourg 1998.

41 P.G.W. Keen: "Decision support systems: a research perspective", in G. Fick, R.H. Sprague (eds.): *Decision support systems: issues and challenges*, Pergamon Press, Oxford 1980.

42 D.J. Power: "What is a DSS?" *The On-Line Executive Journal for Data-Intensive Decision Support* **1** (1997) no. 3.

43 D.J. Power: "Decision support systems: concepts and resources for managers", Quorum Books, Westport 2002.

44 J.C. Giarratano, G.D. Riley: *Expert Systems: Principles and Programming*, 3rd. ed., PWS Publ., Boston 1998.

45 K. Rajaram, J. Ramchandran: "An interactive decision support system for on-line process control", *European Journal of Operational Research* **138** (2002) no. 3, 554–568.

46 O.C. Pires et al.: "A fuzzy-logic based expert system for diagnosis and control of an integrated wastewater treatment", *2nd Mercosur Congress on Chemical Engineering–Enpromer*, Rio de Janeiro, Brasilia, 2005.

47 R. Srinivasan, V. Venkatasubramanian: "Automating HAZOP analysis of batch chemical plants: Part I. The knowledge representation framework", *Comput. Chem. Eng.* **22** (1998) no. 9, 1345–1355.

48 R. Srinivasan, V. Venkatasubramanian: "Automating HAZOP Analysis of Batch Chemical Plants: Part II. Algorithms and Application", *Comput. Chem. Eng.* **22** (1998) no. 9, 1357–1370.

49 F.J. Radermacher: "Decision support systems: Scope and potential", *Decision Support Systems* **12** (1994) no. 4, 257–266.

50 C. Zhao, M. Bhushan, V. Ventkatasubramanian: "PHASuite: an automated HAZOP analysis tool for chemical processes Part II: Implementation and case study", *Process Saf. Environ. Prot.* **83** (2005) no. B6, 533–548.

51 C. Zhao, M. Bhushan, V. Ventkatasubramanian: "PHASuite: an automated HAZOP analysis tool for chemical processes Part I: Knowledge engineering framework", *Process Saf. Environ. Prot.* **83** (2005) no. B6, 509–532.

52 Y. Gao, Z. Shang, A. Kokossis: "Agent-based intelligent system development for decision support in chemical process industry", *Expert Systems with Applications* **36** (2009) no. 8, 11099–11107.

53 C. Zhao et al.: "Towards Intelligent Decision Support for Pharmaceutical Product Development", *Journal of Pharmaceutical Innovation* **1** (2006) 23–35.

54 M.B. Sesen et al.: "An ontological framework for automated regulatory compliance in pharmaceutical manufacturing", *Comput. Chem. Eng.* **34** (2010) no. 7, 1155–1169.